

ISSN (Print): 2077-7973  
ISSN (Online): 2077-8767  
DOI: 10.6977/IJoSI.202412\_8(4)

# International Journal of Systematic Innovation



VOL. 08 NO. 04  
December, 2024

Published by the Society of Systematic Innovation

---

***Opportunity Identification  
&  
Problem Solving***

# The International Journal of Systematic Innovation

---

## **Publisher:**

The Society of Systematic Innovation

## **Editorial Team:**

### Editor-in-Chief:

Sheu, Dongliang Daniel (National Tsing Hua University, Taiwan)

### Executive Editor:

Deng, Jyhjeng (DaYeh University, Taiwan)

Yeh, W. C. (National Tsing Hua University, Taiwan)

### Editorial Team Members (in alphabetical order):

- Cavallucci, Denis (INSA Strasbourg University, France)
- De Guio, Roland (INSA Strasbourg University, France)
- Feygenson, Oleg (Algorithm Technology Research Center, Russian Federation)
- Filmore, Paul (University of Plymouth, UK)
- Kusiak, Andrew (University of Iowa, USA)
- Lee, Jay (University of Cincinnati, USA)
- Litvin, Simon (GEN TRIZ, USA)
- Lu, Stephen (University of Southern California, USA)
- Mann, Darrell (Ideal Final Result, Inc., UK)
- Sawaguch, Manabu (Waseda University, Japan)

- Shouchkov, Valeri (ICG Training& Consulting, Netherlands)
- Song, Yong-Won (Korea Polytechnic University, Korea)
- Yu, Oliver (San Jose State University, USA)
- Zhang, Zhinan (Shanghai Jiao Tong University)

### Managing Editor:

- Jimmy Li
- Maggie Y. Sheu

### Assistant Editors:

- Ming Chan
- Fish Shih
- Chiaoling Ni

## **Editorial Office:**

- The International Journal of Systematic Innovation
- 6F, # 352, Sec. 2, Guanfu Rd, Hsinchu, Taiwan, R.O.C., 30071
- e-mail: editor@systematic-innovation.org  
ijosi@systematic-innovation.org
- web site: <http://www.IJoSI.org>
- Tel: +886-3572-3200

# INTERNATIONAL JOURNAL OF SYSTEMATIC INNOVATION

## CONTENTS

DECEMBER 2024 VOLUME 8 ISSUE 4

---

### FULL PAPERS

- Examining the structural attributes of TRIZ contradiction matrix using exploratory data analysis  
.....W-S. Shin, Y. Choi, J. S. Hyun. 1-17
- Performance evaluation of various optimizers on Alzheimer’s disease classification using deep neural network  
.....T.S. Sindhu, N. Kumaratharan, P. Anandan, P. Durga. 18-26
- Revolutionizing age and gender recognition: an enhanced CNN architecture  
.....R. S. Tumpa, Md. Khaliluzzaman, MD Jiabul Hoque, Roshni Tasnim. 27-45
- Analysis of deep actor-critic methods for classifying cancer subtypes through gene expression  
.....Jayakrishnan R, S. Meera. 46-66
- Efficient task scheduling in the cloud with queuing and multi-tactic harris hawks optimization  
..... S. Antony, Sujatha. S. R. 67-84
- Brain tumor detection using MRI images- a comparative study based on different classifiers  
.....S. R. Puligurti, P. Chitra, A. V. Bharadwaja. 85-102
- Code switching: exploring perplexity and coherence metrics for optimizing topic models of historical documents  
..... M. A. Yusof, S. Saeed. 103-118
- Iris liveness detection for biometric access control system in smart home security using deep convolutional neural network.  
..... Y. G. Waghmare, S. D. Thepade. 119-130
- Exploring maritime movement information: an explainable ai approach using HI-DBSCAN and SHAP analysis  
..... N. Newaliya, V. Siwach, H. Sehrawat, Y. Singh. 131-145
- Research on the path of coordinated development of innovation model and supply chain capability building of new energy vehicle enterprises in China  
.....Y. Zhuang, L. Bin, K. Guihua, Z. Wenxin. 146-156

# Examining the structural attributes of TRIZ contradiction Matrix using exploratory data analysis

Won-Shik Shin<sup>1</sup>, Youngjoon Choi\*, Jung Suk Hyun<sup>2</sup>

<sup>1</sup> CEO of SIXWEZ, sixwez@naver.com

<sup>2</sup> Professor, Department of Management Information Systems, Jeju National University, Korea, jshyun@jejunu.ac.kr

\* Associate Professor, Department of International Office Administration, Ewha Womans University,  
young.choi@ewha.ac.kr

(Received 18 January 2024; Final version received 22 May 2024; Accepted 29 July 2024)

## Abstract

While the efficacy of TRIZ Contradiction Matrix (CM) is still being disputed, CM has been widely used due to its relative ease of use compared to other TRIZ methodologies. However, research on CM has been dominated by case studies of its use, the fundamental structure and components of CM have not been thoroughly explored. This study aims to enhance our understanding on CM by analyzing the relationships between its structural elements and constituents. To do so, Exploratory Data Analysis (EDA) was utilized to explore the correlations between the parameters of improving and worsening features and inventive principles within the intersection boxes. Frequency analysis conjectured a considerable similarity between the same attribute parameters that possessed opposing features. Association rules analysis (ARA) was conducted to identify the structural features of CM. On average, 56.67% of similarity was observed in the inventive principles located within the symmetrical intersection boxes around the matrix's main diagonal. Remarkably, 93.62% of the intersection boxes shared at least one common inventive principle. Propositional logic was adopted as a conceptual tool to interpret and understand the observed probabilities of inventive principles within CM's symmetrical intersection boxes. The findings showed that both improving and worsening parameters tend to converge in function enhancement due to the inventive principles in the intersection boxes. Given that parameters symmetric with the CM's main diagonal represented physical contradiction relations, this study suggests that the intersection box's inventive principles could potentially offer solutions to these physical contradictions. By examining the correlations between 39 Parameters of CM and 40 inventive principles within the intersection boxes, this study provides meaningful insights to understand the complex mechanisms of CM.

*Keywords: Exploratory Data Analysis, TRIZ Contradiction Matrix, 40 Inventive Principles*

## 1. Introduction

Today, the significance of creativity is emphasized at personal, corporate, and even national levels (DiPietro, 2004; Puccio, 2017). In this regard, TRIZ (Theory of Inventive Problem Solving: Teoriya Resheniya Izobretatelskih Zadach) has gained substantial attention as a creative and systematic approach to problem-solving (Ogot & Okudan, 2006). Given that various methodologies of TRIZ have been widely applied and continuously evolved, Ilevbare etc. (2013) found that the 40 Inventive Principles (40IPs) has been most frequently utilized among various problem-solving methodologies within TRIZ, followed by Ideality and Ideal Final Result (IFR), Function Analysis (FA), and Contradiction Matrix

(CM). Although 40IPs and CM are practical tools for addressing problems, the main objective of IFR focuses on problem-solving (Rantanen etc., 2017; Terninko etc., 1998) and FA is a tool that focuses on understanding the functions and interactions of system components (Rantanen etc., 2017). Among these four TRIZ methodologies, this study focuses on CM.

Altshuller (1984, 2002) suggested utilizing 40IPs with the assistance of CM due to its strength of resolving Technical Contradiction (TC). Later, however, CM's effectiveness had been a subject of debate and Altshuller declared the discontinuation of CM due to this controversial issue on effectiveness (Cascini, 2012). Nevertheless, owing to its intuitive and convenient application, CM continues to be widely used (Cempel, 2014; Ma etc., 2015; Pokhrel

etc., 2015; Su & Lin, 2008). To apply CM for various study fields and domains, it has been continuously refined and persistently enhanced (Bogatyrev & Bogatyreva, 2009; Craig etc., 2008; Lim etc., 2018; Mann, 2002, 2004, 2009, 2021; Mann & Dewulf, 2003; Pokhrel etc., 2015). However, analytical studies on CM still remain relatively scarce.

The main purpose of this study is to deepen the fundamental understanding of CM through Exploratory Data Analysis (EDA). More specifically, this study aims to establish a basis for CM research and improvement for the effective utilization of CM by analyzing the relationship between CM and 40IPs. To do so, this study conducted a simultaneous literature review and EDA to facilitate structural understanding of CM. Initially, an in-depth analysis of CM was conducted by reviewing the relevant books and articles. This study then investigated the theoretical background of the reasons for selecting specific data and methodology for analysis. In this study, CM was analyzed by conducting Association Rule Analysis (ARA) using Python after frequency analysis using IBM SPSS version 21 and visualizing structural characteristics.

## 2. Literature Review

### 2.1 TRIZ and contradiction

Altshuller, who researched invention methodology since 1946, presented a problem-solving approach

using TRIZ with CM, consisting of 39 engineering parameters and 40 IPs, in his 1969 book entitled *The Innovation Algorithm* (Altshuller, 1999). Contradiction is an essential concept within TRIZ, which refers to the conflict between objects or properties, or the clash between solutions (Rantanen etc., 2017). Resolving these contradictions is the core of innovative invention (Altshuller, 1984).

Contradictions are categorized into three types: Administrative Contradiction (AC), Technical Contradiction (TC), and Physical Contradiction (PC). AC refers to a situation in which the problem is recognized but no solution is available. It involves eliminating obstacles to obtain essential benefits (Orloff, 2017). TC arises when improving one aspect of a technical system leads to the deterioration of another aspect (Altshuller, 1984). To address this, 40IPs and CM are primarily used (Rousselot etc., 2012). PC occurs when two conditions must be satisfied simultaneously. To solve PC, Altshuller proposed the principles of separation (Altshuller, 1999).

### 2.2 40IPs and CM

Altshuller articulated 40 IPs by conducting patent analysis (Table 1). Since then, 40IPs have been considered the most representative invention tools within TRIZ. It is the most commonly used TRIZ tool for surveys, targeting TRIZ practitioners (Ilevbare etc., 2013).

**Table 1.** 40 Inventive Principles (40IPs)

1. Segmentation	2. Extraction	3. Local Quality	4. Asymmetry	5. Consolidation
6. Universality	7. Nesting	8. Counterweight	9. Prior Counteraction	10. Prior Action
11. Cushion in Advance	12. Equipotentiality	13. Do It in Reverse	14. Spheroidality	15. Dynamicity
16. Partial or Excessive Action	17. Transition into a New Dimension	18. Mechanical Vibration	19. Periodic Action	20. Continuity of Useful Action
21. Rushing Through	22. Convert Harm into Benefit	23. Feedback	24. Mediator	25. Self-service
26. Copying	27. Dispose	28. Replacement of Mechanical System	29. Pneumatic or Hydraulic Constructions	30. Flexible Membranes or Thin Films
31. Porous Material	32. Changing the Color	33. Homogeneity	34. Rejecting and Regenerating Parts	35. Transformation of Properties
36. Phase Transition	37. Thermal Expansion	38. Accelerated Oxidation	39. Inert Environment	40. Composite Materials

CM categorizes the characteristics of TC into improving and worsening features related to 39 parameters (Fig 1). It is structured as a grid with 39 rows and columns and serves as a tool for analyzing the interactions between these features (Altshuller,

1999; Hipple, 2012). Within the intersection boxes where the parameters of two features meet, IPs for solving the given problem are arranged in order of frequency (Haines-Gadd, 2016; Shin & Hyun, 2022; Terninko etc., 1998). Although CM has continuously

improved since the early stages of TRIZ development, its fundamental structure has remained intact (Altshuller, 1984; Pala & Srikant, 2005).

CHARACTERISTICS		Characteristic that is getting Worse								
		1	2	3	4	5	6	7	8	9
Characteristics to be improved	1 Weight of a mobile object			15, 8, 29,34		29, 17, 38, 34		29, 2, 40, 28		2, 8, 15, 38
	2 Weight of a stationary object				10, 1, 29, 35		35, 30, 13, 2		5, 35, 14, 2	
	3 Length of a mobile object	8, 15, 29, 34				15, 17, 4		7, 17, 4, 35		13, 4, 8
	4 Length of a stationary object		35, 28, 40, 29				17, 7, 10, 40		35, 8, 2, 14	
	5 Area of a mobile object	2, 17, 29, 4		14, 15, 18, 4				7, 14, 17, 4		29, 30, 4, 34
	6 Area of a stationary object		30, 2, 14, 18		26, 7, 9, 39					
	7 Volume of a mobile object	2, 26, 29, 40		1, 7, 4, 35		1, 7, 4, 17				29, 4, 38, 34
	8 Volume of a stationary object		35, 10, 19, 14	19, 14	35, 8, 2, 14					
	9 Speed	2, 28, 13, 38		13, 14, 8		29, 30, 34		7, 29, 34		

Fig 1. Snapshot of CM

In order to select parameters for problem-solving, a precise understanding of these parameters is necessary (Domb, E., Miller, J., MacGran, E., & Slocum, 1998; Haines-Gadd, 2016). The following are the 39 parameters: 1) weight of moving object, 2) weight of stationary object, 3) length of moving object, 4) length of stationary object, 5) area of moving object, 6) area of stationary object, 7) volume of moving object, 8) volume of stationary object, 9) speed, 10) force, 11) stress or pressure, 12) shape, 13) stability of the object's composition, 14) strength, 15) duration of moving object, 16) duration of stationary object, 17) temperature, 18) illumination intensity, 19) use of energy by moving object, 20) use of energy by stationary object, 21) power, 22) loss of energy, 23) loss of substance, 24) loss of information, 25) loss of time, 26) quantity of substance/the matter, 27) reliability, 28) measurement accuracy, 29) manufacturing precision, 30) object-affected harmful factors, 31) object-generated harmful factors, 32) ease of manufacture, 33) ease of operation, 34) ease of repair, 35) adaptability or versatility, 36) complexity of device, 37) difficulty of detecting and measuring, 38) extent of automation, and 39) productivity.

The selection of a version of CM is crucial due to the potential discrepancies in IPs listed in the intersection boxes, which may vary based on the source. In this study, the materials from Tools of Classical TRIZ (Altshuller etc., 1999), 40 Principles (Altshuller, 2002), and The Innovation Algorithm (Altshuller, 1999) were reviewed. CM presented in Tools of Classical TRIZ includes some blank intersection boxes, differing from the other two books. In 40 Principles, some intersection boxes

contain IPs, deviating from the principles where 0 to 4 IPs are typically listed. Considering these factors, the CM from The Innovation Algorithm, which provides the most stable and consistent presentation of IPs, was chosen as the research model. In particular, CM from The Innovation Algorithm consists of 1,521 intersection boxes. Among these, 234 intersection boxes are excluded because they are located around the main diagonal or have no IP. This leaves a total of 1,248 intersection boxes with 4,200 IPs listed in them.

### 2.3 Exploratory Data Analysis (EDA)

Exploratory Data Analysis (EDA) is an analytical approach that involves examining data from various angles without preconceived notions, aiming to understand the distribution, anomalies, patterns, etc. (Asian etc., 2016; Komorowski etc., 2016; Martinez etc., 2017; Morgenthaler, 2009). Through visualization and pattern discovery, EDA helps uncover insights, formulate hypotheses, identify patterns in complex phenomena, and gain insights necessary for decision-making in big data analysis (Karageorgiou, 2011; Martinez etc., 2017; Morgenthaler, 2009). EDA plays a significant role in understanding the structure and characteristics of the data.

In this study, EDA was conducted to examine the distribution of 40IPs within CM. This can be seen as an extension of the research related to the frequency or ranking of TRIZ principles. Previous studies have focused on analyzing the frequency order and comparative analysis of IPs (Dave, 2017; Hwang etc., 2018; Mann, 2004; Sen etc., 2021; Terninko etc., 1996). However, this study expands on this by analyzing the statistical characteristics of 40IPs for 39 parameters. Based on the insights obtained, visualization tasks and ARA were carried out.

Association Rule Analysis (ARA), commonly known as Market Basket Analysis, is a data mining technique used to identify relationships between items in large databases. In essence, it examines the associations between the purchase of specific items X and the purchase of item Y within customer buying patterns (Chen etc., 2005; Lu etc., 1998). In this analysis, measurement tools such as support, confidence, and lift are utilized (Agrawal etc., 1996; Hornik etc., 2005; Lu etc., 1998). It is used in conjunction with the Apriori algorithm to eliminate

infrequent item sets. Support represents the proportion of cases in which items X and Y occur together, as shown in Equation (1).

$$\text{Support}(X \rightarrow Y) = P(X \cap Y) = \frac{P(X \cap Y)}{N(\text{total transactions})} \quad (1)$$

Confidence is a measure that indicates the proportion of cases where items containing X also include Y, as in Equation (2). It serves as a tool to evaluate the uncertainty of a rule and is based on the conditional probability.

$$\text{Confidence}(X \rightarrow Y) = P(Y|X) = \frac{P(X \cap Y)}{P(X)} \quad (2)$$

Lift is a measure that assesses the increase in the occurrence probability of Y given the presence of item X, as shown in Equation (3). It is used to validate the significance of the discovered rules by comparing them with randomly set rules. Lift serves as an essential tool for determining the accuracy of correlations

$$\text{Lift}(X \rightarrow Y) = \frac{P(Y|X)}{P(Y)} = \frac{P(X \cap Y)}{P(X)P(Y)} \quad (3)$$

By using an unsupervised learning approach without specific hypotheses or predictive models, ARA examines the co-occurrence frequency between

items to uncover data sparsity and unique patterns. ARA confirms correlations but does not determine causality. In the field of TRIZ, ARA is employed for tasks such as text mining in patent information classification (He & Loh, 2010). In this study, the support calculation method was applied to analyze the similarity of IPs within the intersection boxes around the main diagonal. This approach was also extended to explore identical components between intersection boxes within CM. The analysis tools utilized included Pandas and NumPy, data analysis packages in Python, and Mlxtend, which provide an effective Apriori algorithm for ARA (Raschka, 2018).

### 3. Method and Results

#### 3.1 Frequency analysis of 40IPs for CM

This study compared the results of the CM analysis from The Innovation Algorithm and the study by Dave (2017), which used the triz40.com dataset. In The Innovation Algorithm (N(Total number of IPs in CM)=4,200) and triz40.com (N=4,202), it was found that while there were frequency differences for certain IPs (10, 28, 18, 15, 19, 13, 3, 16, 40, 25, 23), the overall results were largely similar (Table 2). Due to space constraints, only a portion of the results is presented here.

**Table 2.** Frequency by IP (Due to space constraints, only a portion of the results is presented here)

R	Algorithm			triz40			R	Algorithm			triz40		
	IP	F	%	IP	F	%		IP	F	%	IP	F	%
1	35	413	9.8	35	413	9.8	4	28	231	5.5	28	229	4
2	10	272	6.5	10	274	6.5	5	2	222	5.3	2	222	5
3	1	232	5.5	1	232	5.5	6	18	162	3.9	18	163	6
.....													
35	5	35	0.8	23	35	0.8	38	33	31	0.7	33	31	0.7
36	23	34	0.8	5	35	0.8	39	9	26	0.6	9	26	0.6
37	12	32	0.8	12	32	0.8	40	20	19	0.5	20	19	0.5
R=Frequency Ranking, IP=Inventive Principle, F=Frequency count							Sum	(N=4,200)		100	(N=4,202)		100

### 3.2 40IPs of 39 parameters

Previous research in the field of TRIZ primarily focused on the frequency of IPs, investigating 40 IPs within CM. However, since CM is structured based on 39 parameters to determine inventive principles within intersection boxes, analyzing the frequency of

40 IPs within these parameters allows us to grasp the correlations between each parameter and IPs, as well as their detailed rankings. In this study, this approach was adopted to analyze the frequency of IPs for each parameter. The improvement and worsening features of parameters with similar attributes were observed to note similarities in the frequency of IPs (See Table 3, due to space constraints, only a portion of the results are presented here.).

Table 3. 40 IPs Frequency by Parameter

Weight of moving object							Length of moving object						
R	Improving			Worsing			R	Improving			Worsing		
	IP	F (N=126)	%	IP	F (N=126)	%		IP	F (N=104)	%	IP	F (N=120)	%
1	35	13	10.3	35	9	7.1	1	1	11	10.6	1	11	9.2
2	28	8	6.3	28	8	6.3	2	29	11	10.6	15	8	6.7
3	18	7	5.6	8	7	5.6	3	15	9	8.7	29	8	6.7
.....							.....						
34	27	6	4.8	23	1	.8	28				25	1	0.8
35				25	1	.8	29				32	1	0.8
36				30	1	.8	30				38	1	0.8
Total		126	100.0		126	100.0	Total		104	100.0		120	100.0

Ease of repair							Productivity						
R	Improving			Worsing			R	Improving			Worsing		
	IP	F <sub>y</sub> (N=111)	%	IP	F (N=103)	%		IP	F (N=136)	%	IP	F (N=132)	%
1	1	16	14.4	1	16	15.5	1	35	18	13.2	35	20	15.2
2	10	12	10.8	10	12	11.7	2	10	17	12.5	10	14	10.6
3	2	11	9.9	2	11	10.7	3	28	12	8.8	28	13	9.8
.....							.....						
21	26	1	.9	19	1	1.0	31	31	1	.7			
22	29	1	.9	29	1	1.0	36	36	1	.7			
23	31	1	.9				40	40	1	.7			
Total		111	100.0	Total	103	100.0	Total		136	100.0	Total	132	100.0

(R=Frequency Ranking, IP=Inventive Principle, F=Frequency count, N=Total number of IP)

Shulyak presented a method of utilizing CM to discover IPs or applying them sequentially (Altshuller, 2002). However, Dave (2017) proposed an approach that prioritizes the usage of the top 20 most frequent IPs, aiming for efficient problem-solving. In this study, it was suggested to apply frequency rankings for each parameter. The frequency order for each parameter allows for convenient application of IPs, even when focusing on only one parameter at a time.

### 3.3 Visualization of CM

TRIZ has undergone processes of learning, training, application, and continuous refinement and development. However, there remains a lack of structured and statistical analyses of TRIZ methodology. Just as John Tukey defined EDA as an attitude of seeking both what is believed to exist and what is believed not to exist, with flexibility and

spontaneity (Martinez etc., 2017; O’Neil & Schutt, 2013), this study aims to advance the analysis of TRIZ.

For EDA, tasks such as pattern recognition and visualization are essential. In this study, the different characteristics of the 39 parameters (improvement features and worsening features) were used to confirm the similarity of IP frequencies. This formed the basis for graphing the distribution of IPs within the

intersection boxes of CM. This approach is akin to TRIZ deriving patterns and laws of invention by analyzing numerous patents. Through this study, the aim was to gain a clear understanding of the structure of CM and the distribution of IPs within each intersection box. Throughout this process, the distribution of IPs from 1 (Segmentation) to 40 (Composite Materials) was comprehensively visualized as shown in Fig 2.

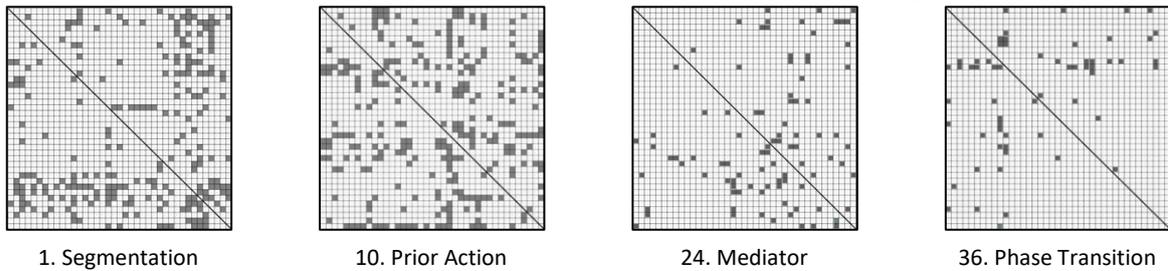


Fig 2. Visualization graphs of the distribution of IPs within CM

Interestingly, a symmetrical pattern was observed around the main diagonal. Therefore, the IPs within

the intersection boxes symmetrically positioned across the diagonal were compared in Fig 3.

CHARACTERISTICS		Characteristic that is getting Worse								
		1	2	3	4	5	6	7	8	9
Characteristics to be improved	1 Weight of a mobile object		-	15, 8, 29,34	-	29, 17, 38, 34	-	29, 2, 40, 28	-	2, 8, 15, 38
	2 Weight of a stationary object	-		-	10, 1, 29, 35	-	35, 30, 13, 2	-	5, 35, 14, 2	-
	3 Length of a mobile object	8, 15, 29, 34	-		-	15, 17, 4	-	7, 17, 4, 35	-	13, 4, 8
	4 Length of a stationary object		35, 28, 40, 29	-		-	17, 7, 10, 40	-	35, 8, 2, 14	-
	5 Area of a mobile object	2, 17, 29, 4	-	14, 15, 18, 4	-		-	7, 14, 17, 4	-	29, 30, 4, 34
	6 Area of a stationary object	-	30, 2, 14, 18	-	26, 7, 9, 39	-		-	-	-
	7 Volume of a mobile object	2, 26, 29, 40	-	1, 7, 4, 35	-	1, 7, 4, 17	-		-	29, 4, 38, 34
	8 Volume of a stationary object	-	35, 10, 19, 14	19, 14	35, 8, 2, 14	-		-		-
	9 Speed	2, 28, 13, 38	-	13, 14, 8	-	29, 30, 34	-	7, 29, 34	-	

Fig 3. Comparison between intersection boxes symmetrically positioned around the main diagonal of CM

The IPs within the intersection boxes showed symmetry around the main diagonal. For instance, in the intersection box between the improvement feature 1, Weight of a mobile object, and the worsening feature 3, Length of a mobile object, the IPs are 15, 8, 29, and 34. Similarly, in the symmetrically positioned

intersection box around the main diagonal, between the improvement feature 3, Length of a mobile object, and the worsening feature 1, Weight of a mobile object, the IPs are 8, 15, 29, and 34. This symmetry allows for a more detailed comparison when arranging the parameters of the improvement and worsening features both vertically and horizontally (Table 4)

**Table 4.** Arrangement of identical parameters horizontally in CM

	1	2	3
[Improving] Weight of a mobile object		-	15, 8, 29,34
[Worsing] Weight of a mobile object		-	8, 15, 29, 34
[Improving] Weight of a stationary object	-		-
[Worsing] Weight of a stationary object	-		-
[Improving] Length of a mobile object	8, 15, 29, 34	-	
[Worsing] Length of a mobile object	15, 8, 29,34	-	

### 3.4 Association Rule Analysis (ARA)

(1) Analysis of similarity between symmetrical intersection boxes around the main diagonal

Out of the 3,042 intersection boxes generated by separating and rearranging the rows and columns of CM, 2,496 contain one or more IPs, while the remaining 546 are empty. Among these, 78 empty intersection boxes are located on the main diagonal, and there are 416 intersection boxes where both sides are symmetrically empty across the main diagonal. Additionally, there are 52 intersection boxes in which only one side is empty of IPs.

To measure the inferred similarities from the process so far, the support measure from set theory's tool, ARA, was utilized. Intersection boxes without IPs listed were excluded from the analysis. A total of 1,222 pairs of intersection boxes containing one or more IPs were analyzed using Python. To calculate

the similarity between the improvement and worsening features, the number of common IPs was divided by the total number of distinct features. Here, the total number of distinct features refers to the union of the IPs corresponding to the 'improvement feature' and the 'worsening feature' within the set of intersection boxes under analysis, as in Equation (4).

$$\text{Support (rate of similarity)} = \frac{n(X \cap Y)}{N(\text{total transactions})} \quad (4)$$

For the analysis, 2,444 intersection boxes were organized into a set data structure. In this process, 'i' was designated for the improving feature, and 'w' was used for the worsening feature. For instance, an intersection box with the characteristics of wanting to improve the weight of a moving object and worsening the length of a moving object was labeled as 'i1\_w3' (Fig 4, left). Then, the following function was created (Fig 4, right).

```

1 i1_w3={15,8,29,34}
2 w1_i3={8,15,29,34}
3 i1_w5={29,17,38,34}
4 w1_i5={2,17,29,4}
5 i1_w7={29,2,40,28}
6 w1_i7={2,26,29,40}
...
2439 i39_w36={12,17,28,24}
2440 w39_i36={12,17,28}
2441 i39_w37={35,18,27,2}
2442 w39_i37={35,18}
2443 i39_w38={5,12,35,26}
2444 w39_i38={5,12,35,26}

```

```

1 total_prob = []
2
3 def two_prob(x, y):
4     twointer = x.intersection(y)
5     print(x, y, 'twointer:', twointer)
6     twounion = set.union(x, y)
7     print(x, y, 'twounion:', twounion)
8     twoprob = (len(twointer)/len(twounion))
9     print(x, y, 'twoprob:', twoprob)
10    total_prob.append(twoprob)

```

**Fig 4.** Treatment for similarity analysis

Note. Left: Set Data Structure; Right: Function for similarity verification

Subsequently, the arguments for executing the function were generated (Fig 5, left). The results of

significance probability execution range from 1.0 to 0.0 are shown in Fig 5, right.

1	two_prob(i1_w3,w1_i3)	{8, 34, 29, 15} {8, 34, 29, 15} twointer : {8, 34, 29, 15}
2	two_prob(i1_w5,w1_i5)	{8, 34, 29, 15} {8, 34, 29, 15} twounion : {34, 8, 15, 29}
3	two_prob(i1_w7,w1_i7)	{8, 34, 29, 15} {8, 34, 29, 15} twoprob : 1.0
4	two_prob(i1_w9,w1_i9)	{17, 34, 29, 38} {17, 2, 4, 29} twointer : {17, 29}
5	two_prob(i1_w10,w1_i10)	{17, 34, 29, 38} {17, 2, 4, 29} twounion : {34, 2, 4, 38, 17, 29}
6	two_prob(i1_w11,w1_i11)	{17, 34, 29, 38} {17, 2, 4, 29} twoprob : 0.3333333333333333
...		
1217	two_prob(i39_w33,w39_i33)	{18, 35, 2, 27} {18, 35} twointer : {18, 35}
1218	two_prob(i39_w34,w39_i34)	{18, 35, 2, 27} {18, 35} twounion : {18, 35, 2, 27}
1219	two_prob(i39_w35,w39_i35)	{18, 35, 2, 27} {18, 35} twoprob : 0.5
1220	two_prob(i39_w36,w39_i36)	{26, 35, 12, 5} {26, 35, 12, 5} twointer : {26, 35, 12, 5}
1221	two_prob(i39_w37,w39_i37)	{26, 35, 12, 5} {26, 35, 12, 5} twounion : {35, 5, 12, 26}
1222	two_prob(i39_w38,w39_i38)	{26, 35, 12, 5} {26, 35, 12, 5} twoprob : 1.0

**Fig 5.** Function argument generation and Execution result

Note. Left: Function argument generation; Right: Execution result

By analyzing a total of 2,444 intersection boxes (1,222 pairs), the average similarity was calculated to be 56.67%. For a specific case, in the instance of  $i2\_w35$  {19, 15, 29} and  $w2\_i35$  {19, 15, 29, 16}, where the common IPs were the same, the probability

of agreement based on set theory was 75%. Considering this, the average similarity of 56.67% can be interpreted as a significant Fig representing the correlation between the improvement and worsening features (Fig 6).

1	len(total_prob)	1222
1	percent_sum = sum(total_prob)	
2	percent_sum	692.4761904761921
1	total_mean_round = round(percent_sum/len(total_prob) * 100, 2)	
2		
3	print(total_mean_round)	56.67

**Fig 6.** Average similarity

From the analysis, it was found that out of the 1,222 pairs of intersection boxes, approximately 94% had at least one common IP. Among them, 350 sets (28.64%) had the same IPs, and only 78 sets (6.38%)

had no common IPs. This confirms that there is a meaningful relationship between the intersection boxes located around the main diagonal (Table 5)

**Table 5.** Calculation of significance group ratios

Equal ratio	Number of intersection box sets	Overall Ratio
100%	350	0.286416
75%	88	0.072013
.....		
14.29%	64	0.052373
0%	78	0.06383
Total	1222	1

As evident from Table 5, intersection boxes with 100% agreement account for approximately 29% (350

sets) of the total sets, whereas sets with no common IPs constitute around 6% (78 sets), making the former

about 4.5 times more prevalent than the latter. This further confirms the significant relationship between the intersection boxes located around the main diagonal.

(2) Analysis of common components among intersection boxes of parameters

Intersection boxes contain IPs based on the improvement and worsening relationships between features, and identifying intersection boxes sharing the same IPs allows insights into the associations among parameters. To analyze such relationships, a set of 726 intersection boxes, each containing all IPs, was examined to investigate the distribution of the same IPs (Fig 7).

1	i1_w3={15, 8, 29, 34}
2	i1_w5={29, 17, 38, 34}
3	i1_w7={29, 2, 40, 28}
...	...
724	i39_w36={12, 17, 28, 24}
725	i39_w37={35, 18, 27, 2}
726	i39_w38={5, 12, 35, 26}

Fig 7. Set data structure for the analysis of common components

Through the analysis of common components, the similarities between each intersection box and the other 725 intersection boxes were compared. For instance, Fig 8 demonstrates that the verification of the similarity between i39\_w2 and other intersection boxes was conFigd to output cases where the IPs of

the intersection boxes were identical using Bool coding. As shown in Fig 8, it was observed that i39\_w2 is identical to i37\_w14.

```

1  if i39_w2==i1_w3:
2     print("i39_w2==i1_w3:", i39_w2==i1_w3)
3  if i39_w2==i1_w5:
4     print("i39_w2==i1_w5:", i39_w2==i1_w5)
5  if i39_w2==i1_w7:
6     print("i39_w2==i1_w7:", i39_w2==i1_w7)
7  if i39_w2==i1_w9:
8     print("i39_w2==i1_w9:", i39_w2==i1_w9)
...
1447 if i39_w2==i39_w36:
1448    print("i39_w2==i39_w36:", i39_w2==i39_w36)
1449 if i39_w2==i39_w37:
1450    print("i39_w2==i39_w37:", i39_w2==i39_w37)
1451 if i39_w2==i39_w38:
1452    print("i39_w2==i39_w38:", i39_w2==i39_w38)
i39_w2==i37_w14: True

```

Fig 8. Bool coding for the analysis of common components

The analysis revealed that there were 107 cases in which 2 intersection boxes were identical, 10 cases which 3 intersection boxes were identical, and 3 cases which 4 intersection boxes were identical. While some instances showed the same IP in intersection boxes with different parameters, the similarity was mainly observed among intersection boxes symmetrically positioned around the diagonal. Among these symmetric intersection boxes, there were a total of 93 cases in which the same IPs were present (Table 6).

Table 6. Symmetric Intersection Boxes with the Same IPs

No.	cell_a	cell_b	No.	cell_a	cell_b	No.	cell_a	cell_b
1	i1_w3	i3_w1	32	i11_w31	i31_w11	63	i21_w25	i25_w21
2	i1_w11	i11_w1	33	i12_w17	i17_w12	64	i21_w30	i30_w21
3	i1_w36	i36_w1	34	i12_w23	i23_w12	65	i21_w36	i36_w21
.....								
29	i9_w32	i32_w9	60	i19_w36	i36_w19	91	i35_w36	i36_w35
30	i11_w17	i17_w11	61	i20_w30	i30_w20	92	i36_w37	i37_w36
31	i11_w23	i23_w11	62	i21_w23	i23_w21	93	i38_w39	i39_w38

Furthermore, the results identified a total of 14 cases in which intersection boxes were diagonally asymmetric while still sharing the same IP (Table 4).

In most cases, these were either identical or similar parameters (same parameters with different features). However, instances such as i5\_w39 and i36\_w28,

i6\_w16 and i15\_w7, i14\_w21 and i38\_w2, and i39\_w2 and i37\_w14 did not overlap in terms of the 39 parameters (Table 7).

**Table 7.** Asymmetry intersection boxes with identical IPs

No.	cell_c	cell_d	IPs			
1	i1_w29	i38_w1	18	26	28	35
2	i1_w34	i2_w34	2	11	27	28
3	i3_w38	i37_w3	16	17	24	26
.....						
6	i5_w39	i36_w28	2	10	26	34
7	i6_w16	i15_w7	2	10	19	30
8	i13_w34	i34_w30	2	10	16	35
9	i14_w21	i38_w2	10	26	28	35
.....						
14	i39_w2	i37_w14	3	15	27	28

Among the 10 cases where 3 intersection boxes are identical, 9 of them exhibited a diagonal symmetry relationship between two of them. However, [i5\_w33,

i18\_w34, i19\_w35] did not display such symmetry (Table 8).

**Table 8.** Three intersection boxes with identical IPs

No.	cell_e	cell_f	cell_g	IPs			
1	i2_w27	i27_w2	i27_w10	3	8	10	28
2	i5_w11	i11_w5	i39_w10	10	15	28	36
3	i5_w33	i18_w34	i19_w35	13	15	16	17
.....							
9	i25_w31	i31_w39	i39_w31	18	22	35	39
10	i30_w36	i30_w37	i36_w30	19	22	29	40

There were a total of 3 cases where 4 intersection boxes matched, and in all cases, the intersection boxes

were symmetrically related across the main diagonal (Table 9)

**Table 9.** Four intersection boxes with identical IPs

No.	cell_h	cell_i	cell_j	cell_k	IPs			
1	i5_w28	i6_w28	i28_w5	i28_w6	3	26	28	32
2	i22_w30	i22_w31	i30_w22	i31_w22	2	21	22	35
3	i24_w25	i25_w24	i28_w37	i37_w28	24	26	28	32

Out of the 107 cases, 93 were part of the 350 intersection box sets identified in the previous similarity analysis. The remaining 14 cases shared the same IPs while having different parameters.

Interestingly, some intersection boxes shared the same IPs despite being composed of entirely different parameters. These findings suggest the need for further research and investigation.

Choi (2015) and Hyun (2018) utilized propositions for clarity in contradiction resolution within TRIZ. Similarly, this study also employed propositional logic for inference to find evidence of the significance of the probabilities associated with the IPs listed in the intersection boxes around the main diagonal symmetry. A proposition refers to a declarative statement or expression that can be judged as true or false. The primary proposition is an indivisible basic unit. A compound proposition is formed by combining two or more primary propositions using logical connections. Among these, a conditional proposition takes the form if  $p$ , then  $q$ , is expressed as  $p \rightarrow q$ . According to the contraposition law, if  $p \rightarrow q$  is true, then  $\sim q$  (the negation of  $q$ )  $\rightarrow \sim p$  (the negation of  $p$ ) is also true.

In this study, the parameters associated with the features to be improved and those associated with the features to be worsened are denoted as 'i' and 'w' respectively. According to this notation, CM can be explained as follows: If a certain parameter is to be improved (i), then another parameter will worsen (w). Altshuller described CM as follows: "(If) to resolve a parameter that requires improvement, (then) using conventional methods known for this purpose, (but) a parameter that worsens (is worsened) is written (in the matrix)." (Altshuller, 1984). For example, 'i1\_w3' can be interpreted as follows: if (i1) the weight of a moving object is to be improved, then (w3) the length of the moving object will worsen ( $i1 \rightarrow w3$ ). This problem can be solved using the IPs associated with the cross-point box {8, 15, 29, 34}. In other words, the IPs {8, 15, 29, 34} provide a solution where i1 (weight of a moving object) is improved, while w3 (length of a moving object) does not worsen ( $\sim w3$ ). The contraposition of 'i1\_w3' would be: if ( $\sim w3$ ), the length of the moving object does not worsen, then ( $\sim i1$ ) the weight of the moving object will not improve ( $\sim w3 \rightarrow \sim i1$ ). Therefore, the IPs that can solve this contraposition problem are {8, 15, 29, 34}. In other words, {8, 15, 29, 34} are the IPs that can satisfy all conditions [i1, w3,  $\sim w3$ ,  $\sim i1$ ] of the problem.

The cross-point box i3\_w1, located symmetrically around the diagonal, can be understood as follows. If (i3), the length of a moving object is to be improved, then (w1) the weight of the moving object will worsen ( $i3 \rightarrow w1$ ). To resolve this, the IPs {8, 15, 29, 34} are used. In other words, the IPs {8, 15, 29, 34} offer a solution where i3 (length of a moving object) is improved, while w1 (weight of a moving object) does not worsen ( $\sim w1$ ). The contraposition of

'i3\_w1' would be: if ( $\sim w1$ ), the weight of the moving object does not worsen, then ( $\sim i3$ ) the length of the moving object will not improve ( $\sim w1 \rightarrow \sim i3$ ). The IPs that can solve this contraposition problem are also {8, 15, 29, 34}. In other words, {8, 15, 29, 34} are the IPs that can address the conflicting conditions [i3, w1,  $\sim w1$ ,  $\sim i3$ ].

'Improving' signifies the intention to enhance the current situation, while 'not worsening' represents the desire to maintain or restore the present state without deterioration. Although improvement and worsening are opposing concepts, 'improving' and 'not worsening' both aim at enhancing performance in the current context. Thus, the relationships expressed in i1\_w3 and i3\_w1 may differ in the current state, but both are aimed at enhancing performance. Therefore, the associations between i1 and  $\sim w1$ , as well as i3 and  $\sim w3$ , in the context of CM's diagonal symmetry are closely related. In essence, the symmetry around the diagonal of the CM can often be seen as contraposition relationships. Hence, the similarity in IPs between cross-point boxes symmetrically located around the diagonal of the CM can be understood from this perspective.

The IPs {8, 15, 29, 34} for both i1\_w3 and i3\_w1 encompass solutions for the conditions [i1,  $\sim i1$ , i3,  $\sim i3$ , w1,  $\sim w1$ , w3,  $\sim w3$ ]. Notably, i1 and  $\sim i1$ , i3 and  $\sim i3$ , w1 and  $\sim w1$ , and w3 and  $\sim w3$  are opposing parameters that represent PC relationships. This illustrates the deep correlation between TC and PC and the cyclical solution pattern, as Altshuller suggested algorithms for eliminating TC: transposing it to PC, using S-Field transformations to remove the PC, and applying the system of operators such as CM at ARIZ 71 and the 76 Standard Solution at ARIZ 85-C (Altshuller, 1984; 1999).

Some TRIZ researchers explain that when a box in the CM is empty, it represents a situation in which the occurrence of TC is unlikely. However, Altshuller and his colleagues acknowledge the presence of empty cells, but have not yet found solutions for them (Altshuller, 1984). Therefore, by utilizing the propositional logic as described above, it might become more feasible to fill in these empty cells within the CM.

### (3) Analysis of the overall associative rules in CM

This study conducted an analysis of associative rules targeting 1,248 out of 1,521 intersections in the CM that included IPs. The setting for the analysis measurement tools generally included a minimum

support of 1% and a minimum confidence of 50%. Additionally, associations were considered if the lift was greater than 1. However, since the required number of rules might not emerge or an excessive number of associative rules might be generated depending on the evaluation criteria, it was necessary to iteratively adjust the evaluation threshold to obtain an appropriate number of rules. In the analysis of associative rules, the observed values of transactions were not considered. Thus, in this research, only the IPs within the intersection boxes were used for analysis, excluding the names of the intersection boxes themselves. Initially, the analysis of associative

rules was performed with a minimum support of 1%, without considering confidence and lift (Fig 9).

```
1 freq_itemsets = apriori(cont_df,
2                       min_support=0.01,
3                       use_colnames=True)
4 freq_itemsets
```

Fig 9. Minimum setting for support

Table 10 presents the results obtained when setting the minimum support to 1%, sorted in descending order, yielding 173 outcomes.

Table 10. Results of ARA with a minimum support set to 1%  
(Due to space constraints, only a portion of the results is presented here.)

No.	support	item sets	No.	support	item sets
1	0.3309	IP35	88	0.0184	IP35, IP38
2	0.2179	IP10	89	0.0176	IP1, IP19
3	0.1859	IP1	90	0.0176	IP1, IP26
.....					
85	0.0184	IP29, IP15	172	0.0104	IP39, IP19
86	0.0184	IP35, IP16	173	0.0104	IP22, IP39
87	0.0184	IP28, IP24			

In cases where a single IP is mentioned, the results align with previous frequency analysis findings. However, for instances where two IPs are mentioned together, some showed more significant results compared to a single IP. For example, the combination of '10. Preliminary Action' and '35. Change of Attribute' exhibited higher support than '24. Intermediary.'

In general, as the number of items increases, obtaining meaningful ARA results becomes more challenging. Therefore, judicious use of appropriate

variables is recommended. However, in this study, all 40 variables were used to investigate the relationships among 40 IPs. Hence, when using standard measurement tool settings, it becomes difficult to derive suitable association rules. Due to this issue, this study set the confidence threshold to be lower than the typical setting. Setting the confidence threshold above 50% significantly narrowed down ARA results, resulting in only two rules being generated (Table 11).

Table 11. Result of ARA with a confidence level of 50% or higher

No.	Antecedents	Consequents	Support	Confidence	Lift
1	IP20	IP10	0.0104	0.6842	3.1393
2	IP21	IP35	0.0152	0.5135	1.5517

Accordingly, evaluation criteria were adjusted to obtain suitable results. The modified measurement tool settings were as follows: a minimum support of 1%, a minimum confidence of 30%, and a minimum lift of 1 (Fig 10).

```
1 rules[ (rules['antecedent_len'] >= 1) &
2       (rules['support'] >= 0.01) &
3       (rules['confidence'] >= 0.3) &
4       (rules['lift'] >= 1) ]
```

Fig 10. Value amendment for support, confidence, and lift

Through these adjustments, 22 results were generated, and Table 12 below presents the results sorted in descending order based on lift.

**Table 12.** Analysis results with support  $\geq 1\%$ , confidence  $\geq 30\%$ , and lift  $\geq 1$

No.	Antecedents	Consequents	Support	Confidence	Lift
1	IP36	IP37	0.0160	0.3333	6.9333
2	IP37	IP36	0.0160	0.3333	6.9333
3	IP20	IP10	0.0104	0.6842	3.1393
.....					
20	IP10	IP35	0.0745	0.3419	1.0332
21	IP2	IP35	0.0601	0.3378	1.0209
22	IP18	IP35	0.0433	0.3333	1.0073

'36. Transition' and '37. Thermal Expansion' have the highest lift value (6.9333), indicating a very strong relationship between these two IPs. Additionally, principles such as '35. Change of Attributes', '10. Preliminary Action', and '1.

Segmentation', which showed high frequencies in the frequency analysis, are frequently appearing as Consequents. ARA results among IPs are in Table 13, and the sorting criterion in the table is based on decreasing lift values

**Table 13.** ARA results for 40 IPs

**1. Segmentation**

No	A	C	S	C	L
1	IP1	IP11	0.0136	0.0733	1.8663
2		IP13	0.0345	0.1853	1.6522
3		IP16	0.0168	0.0905	1.1527
4		IP17	0.0144	0.0776	1.1130
5		IP32	0.0240	0.1293	1.0978

**13. Do It in Reverse**

No	A	C	S	C	L
1	IP13	IP1	0.0345	0.3071	1.6522
2		IP17	0.0128	0.1143	1.6394
3		IP15	0.0184	0.1643	1.2814
4		IP32	0.0144	0.1286	1.0915
5		IP2	0.0208	0.1857	1.0440

**26. Copying**

No	A	C	S	C	L
1	IP26	IP24	0.0160	0.1418	1.9241
2		IP28	0.0296	0.2624	1.4177
3		IP17	0.0104	0.0922	1.3226
4		IP27	0.0120	0.1064	1.0882
5		IP32	0.0144	0.1277	1.0838

**39. Inert Environment**

No	A	C	S	C	L
1	IP39	IP22	0.0104	0.1688	2.5083
2		IP18	0.0128	0.2078	1.6008
3		IP19	0.0104	0.1688	1.3169
4		IP35	0.0264	0.4286	1.2951
5		IP2	0.0136	0.2208	1.2411

(A : Antecedents, C : Consequents, S : Support, C : Confidence, L : Lift)

In the ARA of 40 IPs, '7. Nesting', '9. Preliminary Anticipatory Action', '12. Equilibrium', '30. Thin Film', and '33. Homogeneity' had lift values of 1 or lower, indicating random relationships, and were therefore excluded from the analysis. Additionally, '5. Merge', '8. Floating in the Air', '11. Preliminary Prevention', '20. Continuity of Useful Action', '21. Hurry Through', '23. Feedback', '25. Self-Service', '31. Porous Materials', '38. Oxidizing Agents', and others

had only one Consequent when they were antecedents and showed low associations with other IPs. This result could be understood in relation to the frequency analysis. In the frequency analysis, '35. Change of Attributes' was the most frequent with 413 occurrences, while '20. Continuity of Useful Action' was the lowest with 19 occurrences. Due to the dominance of these top 20 IPs, which accounted for

75% of all IPs, some principles were excluded from ARA.

Considering this, the study concluded that conducting separate association analyses for each of 40 IPs would be more meaningful than analyzing the overall associations across all 40 principles. This approach allowed for understanding how each IP is connected to others, and this information will serve as valuable data for building a network of relationships among the IPs in the future.

#### 4. Discussion and Conclusion

Despite being declared discontinued by Altshuller, CM continues to maintain popularity as one of the essential tools in TRIZ. However, CM remains at a fundamental analysis level, which can be a significant hindrance to its development. For these reasons, this study aimed to gain a clear understanding of the underlying structure of CM and verify the relationships among its elements. To achieve this, frequency analysis and EDA were conducted. The findings of this study provide meaningful theoretical implications as follows:

(1) For users who find it challenging to generalize CM's parameters, a method was proposed that applies 40 IPs according to the frequency order of each parameter. In addition, it was discovered that parameters with contrasting characteristics often share similar IPs.

(2) Through EDA, the overall structure of CM and the similarity of IPs within cross-reference boxes were verified. This process revealed the potential for cluster analysis to be employed.

(3) ARA revealed that the relationships between parameters and cross-reference boxes were identified. This process led to a clearer understanding of the relationships between IPs and provided insights into comprehending the intricate mechanisms of CM.

(4) The findings suggest that further research is necessary concerning parameters that share the same components.

(5) The analysis of the overall associations among all IPs revealed that certain principles exhibit a high degree of correlation with other principles.

(6) The significance probability of IPs within diagonally symmetrical intersection boxes of CM was analyzed using propositional logic around the main diagonal.

These research findings will contribute to both the theoretical understanding and practical utilization of CM. They offer more efficient and objective problem-solving methods and provide valuable guidance to researchers and practitioners using CM. This study is anticipated to serve as a catalyst for expanding the depth and breadth of research and applications in the field of CM.

However, this study has certain limitations due to the intricate theoretical background of CM and the complexity of its exploration and validation processes. One limitation is that different versions of CM exist, which can lead to variations in the results depending on the version used for analysis. This could impose constraints on the universal application of CM, highlighting the need for a detailed analysis of these variations. Another limitation is that this study employed EDA methods, rather than traditional hypothesis testing, to extract insights from various perspectives. However, these findings need to be validated through subsequent research. This study utilized only a subset of various analytical methods. Therefore, it is anticipated that research employing a broader range of analytical methods could offer a deeper understanding of the CM.

For the full results of the analysis, see the Google Drive shared file:

[https://docs.google.com/spreadsheets/d/109pcZYn-roD9BA\\_naIdJ7Fsc3qhRBI1jl/edit?usp=drive\\_link&oid=109092790199240646038&rtfpof=true&sd=true](https://docs.google.com/spreadsheets/d/109pcZYn-roD9BA_naIdJ7Fsc3qhRBI1jl/edit?usp=drive_link&oid=109092790199240646038&rtfpof=true&sd=true)

#### References

- Agrawal, R., Mannila, H., Srikant, R., Toivonen, H., & Verkamo, A. I. (1996). Fast discovery of association rules. *Advances in Knowledge Discovery and Data Mining*, 12(1), 307–328.
- Altshuller, G. S. (1984). *Creativity as an exact science: the theory of the solution of inventive problems*. Gordon and Breach Science Publishers.
- Altshuller, G. S. (1999). *The innovation algorithm: TRIZ, systematic innovation and technical creativity*. Technical innovation center, Inc.

- Altshuller, G. S. (2002). 40 principles: TRIZ keys to innovation (Vol.1). Technical Innovation Center, Inc. Altshuller, G. S., Zlotin, B., Zusman, A., & Philatov, V. (1999). Tools of Classical TRIZ. Ideation International Inc.
- Asian, S., Ertek, G., Haksoz, C., Pakter, S., & Ulun, S. (2016). Wind turbine accidents: A data mining study. *IEEE Systems Journal*, 11(3), 1567–1578.
- Bogatyrev, N. R., & Bogatyreva, O. A. (2009). TRIZ evolution trends in biological and technological design strategies. *Proceedings of the 19th CIRP Design Conference—Competitive Design*, 30–31.
- Cascini, G. (2012). TRIZ-based Anticipatory Design of Future Products and Processes. *Journal of Integrated Design & Process Science*, 16(3), 29–63.
- Cempel, C. (2014). The ideal final result and contradiction matrix for machine condition monitoring with TRIZ. *Key Engineering Materials*, 588, 276–280.
- Chen, Y.-L., Tang, K., Shen, R.-J., & Hu, Y.-H. (2005). Market basket analysis in a multiple store environment. *Decision Support Systems*, 40(2), 339–354.
- Choi, S. W. (2015). Review and Application of Creative Problem-Solving Processes for Technical and Physical Contradictions Using Cause-And-Effect Contradiction Tree and Integrated Principles of TRIZ. *Journal of Korea Safety Management & Science*, 17(2), 215-228.
- Craig, S., Harrison, D., Cripps, A., & Knott, D. (2008). BioTRIZ suggests radiative cooling of buildings can be done passively by changing the structure of roof insulation to let longwave infrared pass. *Journal of Bionic Engineering*, 5(1), 55–66.
- Dave, H. p. (2017). TRIZ: 40 principles and their ranking by contradiction matrix. 2017 2nd International Conference for Convergence in Technology (I2CT), 1258–1261. IEEE.
- DiPietro, W. R. (2004). Country creativity and IQ. *The Journal of Social Political and Economic Studies*, 29(3), 345.
- Domb, E., Miller, J., MacGran, E., & Slocum, M. (1998). The 39 Features of Altshuller's Contradiction Matrix. *The TRIZ Journal*, 11(11), 10.
- Haines-Gadd, L. (2016). *TRIZ for Dummies*. John Wiley and Sons Inc.
- He, C., & Loh, H. T. (2010). Pattern-oriented associative rule-based patent classification. *Expert Systems with Applications*, 37(3), 2395–2404.
- Hipple, J., & Hipple, J. (2012). The TRIZ contradiction table. *The Ideal Result: What It Is and How to Achieve It*, 105-130.
- Hornik, K., Grün, B., & Hahsler, M. (2005). arules-A computational environment for mining association rules and frequent item sets. *Journal of Statistical Software*, 14(15), 1–25.
- Hwang, D., Nam, H., Kim, A. J., & Kweon, O. S. (2018). A Study on the Correlation Between TRIZ Principle and Universal Design : Focused on the Case of Universal Design Project. *Archives of Design Research*, 31(3), 121–131.
- Hyun, J. S. (2018). *Principles and Practice of Creative Contradiction Solving*. Cheongram.
- Ilevbare, I. M., Probert, D., & Phaal, R. (2013). A review of TRIZ, and its benefits and challenges in practice. *Technovation*, 33(2–3), 30–37.
- Karageorgiou, E. (2011). The logic of exploratory and confirmatory data analysis. *Cognitive Critique*, 35–48.
- Data, M. C., Komorowski, M., Marshall, D. C., Salciccioli, J. D., & Crutain, Y. (2016). Exploratory data analysis. *Secondary analysis of electronic health records*, 185-203.
- Lim, C., Yun, D., Park, I., & Yoon, B. (2018). A systematic approach for new technology development by using a biomimicry-based TRIZ contradiction matrix. *Creativity and Innovation Management*, 27(4), 414–430.
- Lu, H., Han, J., & Feng, L. (1998). Stock movement prediction and n-dimensional inter-transaction association rules. In 1998 ACM SIGMOD

- Workshop on Research Issues on Data Mining and Knowledge Discovery, Seattle, WA, USA, ACM, New York, USA.
- Ma, S. S., Yang, S. Y., Xu, Y. X., & Zhang, C. W. (2015). Application of TRIZ Contradiction Matrix in Equipment's combat resilience Design. *Applied Mechanics and Materials*, 722, 149–153.
- Mann, D. (2002). Assessing the accuracy of the contradiction matrix for recent mechanical inventions. *The TRIZ Journal*, 4(1), 1–9.
- Mann, D. (2004). Comparing the classical and new contradiction matrix part 1-zooming out. *Change*, 1(3), 7.
- Mann, D. (2009). *Matrix 2010: Re-updating the TRIZ contradiction matrix*. Ifr Press.
- Mann, D. (2021). Matrix 2022: Re-imagining the Contradiction Matrix. In *Creative Solutions for a Sustainable Development: 21st International TRIZ Future Conference, TFC 2021, Bolzano, Italy, September 22–24, 2021, Proceedings 21* (pp. 197–208). Springer International Publishing.
- Mann, D., & Dewulf, S. (2003). Updating the contradiction matrix. *TRIZCON12003*, Philadelphia, March.
- Martinez, W. L., Martinez, A. R., & Solka, J. L. (2017). *Exploratory data analysis with MATLAB®*. Chapman and Hall/CRC.
- Morgenthaler, S. (2009). Exploratory data analysis. *Wiley Interdisciplinary Reviews: Computational Statistics*, 1(1), 33–44.
- O'Neil, C., & Schutt, R. (2013). *Doing data science: Straight talk from the frontline*. O'Reilly Media, Inc.
- Ogot, M., & Okudan, G. E. (2006). Integrating systematic creativity into first-year engineering design curriculum.
- Orloff, M. A. (2017). *ABC-TRIZ*. Springer International Publishing.
- Pala, S., & Srikant, A. (2005). *TRIZ: A New Framework for Innovation - Concepts and Cases*. ICFAI University Press.
- Pokhrel, C., Cruz, C., Ramirez, Y., & Kraslawski, A. (2015). Adaptation of TRIZ contradiction matrix for solving problems in process engineering. *Chemical Engineering Research and Design*, 103, 3–10.
- Puccio, G. J. (2017). From the Dawn of Humanity to the 21st Century: Creativity as an Enduring Survival Skill. *The Journal of Creative Behavior*, 51(4), 330–334.
- Rantanen, K., Conley, D. W., & Domb, E. R. (2017). *Simplified TRIZ: New problem solving applications for technical and business professionals*. Taylor & Francis.
- Raschka, S. (2018). MLxtend: Providing machine learning and data science utilities and extensions to Python's scientific computing stack. *Journal of Open Source Software*, 3(24), 638.
- Rousselot, F., Zanni-Merk, C., & Cavallucci, D. (2012). Towards a formal definition of contradiction in inventive design. *Computers in Industry*, 63(3), 231–242.
- Sen, N., Baykal, Y., & Civek, T. (2021). Prevention of Tearing Failure during Forming of Lower Control Arm via TRIZ Methodology. *Journal of Materials Engineering and Performance*, 30(10), 7721–7729.
- Shin, W.S., & Hyun, J. S. (2022). The Study on the TRIZ Contradiction Matrix through the Literature. *Journal of Tourism & Industry Research*, 42(1), 45–56.
- Su, C. T., & Lin, C. Sen. (2008). A case study on the application of fuzzy QFD in TRIZ for service quality improvement. *Quality and Quantity*, 42(5), 563–578.
- Terninko, J., Zusman, A., & Zlotin, B. (1996). *Step-by-step TRIZ: Creating innovative solution concepts*. Responsible Management Incorporated.
- Terninko, J., Zusman, A., & Zlotin, B. (1998). *Systematic innovation: an introduction to TRIZ (theory of inventive problem solving)*. CRC press.

**AUTHOR BIOGRAPHIES**

**Won-Shik Shin** is a Ph.D. in the Faculty of Data Science for Sustainable Growth, majoring in Management Information Systems at Jeju National University. He is the CEO of SIXWEZ, a company focused on the development of logical creativity, and serves as a Director of the Korea TRIZ Association and The Korean Society of Creative Application. His work has been recognized with multiple awards, including First Place and Honorable Mention at the 14th ICSI & 13th GCSI in 2023, as well as the Best Paper Award at the 2022 Spring Conference of The Korean Society of Intellectual Property Education & Research. His current research interests lie in effective problem-solving and efficient capacity building through the convergence and collaboration between TRIZ and generative artificial intelligence.



**Dr. Jung Suk Hyun** is a Marketing professor at Jeju National University, recognized with multiple awards including "Excellence in Research," "Professor Who Illuminated Jeju University," and "Best Professor Award." He received the "Patent Office Commissioner's Commendation," spoke at KTF 2015, and was Editor-in-Chief of Korea Business Review. His course, "Resolving Contradictions for Innovation," is in K-MOOC. He is the President of the Korean Strategic Marketing Society and won the "2023 Best Paper Award" from the Korean Society of Computer Education.



**Youngjoon Choi** is an Associate Professor in the Department of International Office Administration at the College of Science and Industry Convergence at Ewha Womans University since September 2021. From August 2015 to July 2021, he worked as an Assistant Professor in the School of Hotel and Tourism Management at the Hong Kong Polytechnic University. He received his BS in Tourism Management from Kyung Hee University. He then obtained his MSc and PhD in Recreation, Park & Tourism Management from The Pennsylvania State University, USA in 2011 and 2014 respectively. His main research interests include MICE (meeting, incentive, convention, and exhibition) management, event tourism, technology innovation in hospitality, and destination marketing.

# Performance evaluation of various optimizers on Alzheimer's disease classification using deep neural network

T.S. Sindhu<sup>1\*</sup>, N. Kumaratharan<sup>2</sup>, P. Anandan<sup>3</sup>, P.Durga<sup>4</sup>

<sup>1,4</sup> Department of Electronics and Communication Engineering, C. Abdul Hakeem College of Engineering and Technology, Ranipet, Tamilnadu, India.

<sup>2</sup>Department of ECE, Sri Venkateswara College of Engineering, Sriperumbudur, Chennai, Tamilnadu, India.

<sup>3</sup>Department of ECE, Saveetha Institute of Medical and Technical Sciences, Chennai, Tamilnadu, India.

\* Corresponding author E-mail: sindhuts1992@gmail.com

(Received 19 July 2023; Final version received 17 June 2024; Accepted 29 June 2024)

## Abstract

The proposed work focuses on using transfer learning and CNN models for the classification of Alzheimer's Disease (AD) based on different classes of datasets. The goal is to improve the early diagnosis and classification of AD, which can contribute to better patient recovery and management. The study compares the performance of four different CNN models: AlexNet, GoogLeNet, SqueezeNet, and MobileNet V2. These models have been widely used in various computer vision tasks and have proven to be effective in image analysis. Additionally, three different optimizers are evaluated: Stochastic Gradient Descent with Momentum (SGDM), RMSProp, and ADAM. Optimizers play a crucial role in training deep neural networks, as they determine how the model updates its weights during the learning process. According to the results of the study, the MobileNet V2 model with the SGDM optimizer achieved the highest classification accuracy of 91% among all the tested classifiers. Here, datasets are taken from Kaggle and Mobilenet classifies the output into four classes namely Very Mild Demented, Mild Demented, Moderately Demented, Non-Demented. This suggests that this combination is particularly effective for AD diagnosis and classification based on the given datasets. The automated Alzheimer's disease classification system developed in this work has the potential to identify early signs and symptoms of the disease. Early detection is crucial because it allows medical professionals to intervene at an earlier stage, providing timely treatment and management strategies. By leveraging medical image analysis and transfer learning techniques, this system can contribute to more effective and efficient AD diagnosis, leading to improved patient outcomes.

*Keywords: Alexnet, Googlenet, Squeezenet, Mobilenet2, ADAM, RMSProp and SGDM, AD Classification*

## 1. Introduction

Alzheimer's Disease (AD) is a neurodegenerative brain disorder that affects a significant number of people worldwide. As you mentioned, approximately 50 million individuals are affected by AD globally (Raza etc. 2023), with over six million in the United States alone. Additionally, there are currently 9.44 lakh (944,000) individuals living with dementia (Arafa etc.2023).AD primarily affects older individuals, typically those aged 65 and above. The disease is characterized by progressive memory loss and the deterioration of thinking and learning abilities.

It also leads to a reduction in the size of the hippocampus and cerebral cortex, which are important

brain regions involved in memory and cognitive processes. The shrinkage of these areas contributes to the irreversible damage caused by AD and can ultimately result in death. Although there is currently no complete recovery from AD, early diagnosis and treatment can significantly improve the health condition of patients. The stages of AD classification are based on the patient's brain and health condition and can be categorized as Mild Cognitive Impairment, Mild Alzheimer's, Moderate Alzheimer's, and Severe Impairment.

“Rallabandi etc. (2020)” proposed a machine learning classifier based on the Radian Basis Function (RBF) kernel and non-linear Support Vector Machine (SVM) using the Auto WEKA 2.6 tool for accurate

classification of different stages of AD. Amir “Ghahnavieh etc. (2019)” suggested a combination of Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN) for AD classification. Features extracted from CNN were fed as input to RNN to improve classification accuracy. “Hon & Khan (2017)” utilized the Inception v3 transfer learning network for AD classification. Transfer learning involves leveraging pre-trained models on large datasets to improve performance on smaller datasets. “Fu'adah etc. (2021)” employed the AlexNet architecture for AD classification, using a 75% training and 25% validation split of the datasets.

“Helaly etc. (2022)” proposed a transfer learning model architecture in CNN for the multi-class and binary classification of 2D and 3D medical images in AD. “Wen etc. (2020)” explored four different approaches for AD classification: 3D subject-level, ROI-based, patch-level, and slice-level CNN models. “Oh etc. (2019)” suggested an unsupervised learning approach using a convolutional autoencoder for feature extraction and supervised transfer learning for the classification of AD, normal controls (NC), and mild cognitive impairment (MCI). “Acharya etc. (2021)” employed VGG16, ResNet, and AlexNet transfer learning models to classify AD. “Zhang etc. (2016)” proposed a landmark-based framework to extract features for AD classification. These features were then fed into an SVM classifier. “Rallabandi etc. (2020)” used the FMRIB's Software Library (FSL) for skull stripping, FSL-FAST4 for segmentation, and FreeSurfer for feature extraction from gray matter segmented images. AD classification was performed using the Auto WEKA2.6 ML tool. These studies demonstrate the application of various machine learning techniques and architectures for the classification of AD, aiming to improve early detection and diagnosis of the disease.

Our research contributions are summarized below:

This paper explores the impact of various CNN models and optimizers on AD classification accuracy and suggests effective optimizers for classification tasks.

For classification accuracy, the study compares widely used models (Googlenet, Squeezenet, MobilenetV2) with common optimizers (ADAM, RMSProb, SGDM).

Validation is accomplished by comparing results to those of existing methods. Notably, with a

minimum epoch size of 10 and a learning rate of 0.0003, a greater accuracy of 91.34% is achieved.

The remaining sections of this paper are organized in the following manner. Section 2 explores the proposed methodology, including data collection, preprocessing techniques, data augmentation, the optimizer used, and the classifiers employed. Section 3 delves into the results and offers a comparative discussion of the proposed method with other existing methods.

## 2. Proposed Methodology

The objective of the proposed work is to evaluate the performance of convolutional neural networks (CNNs) using different optimizers for the task of classifying Alzheimer's disease datasets. The study focuses on four pretrained models: AlexNet, SqueezeNet, GoogLeNet, and MobileNetV2. To assess the classification accuracy of these models, three popular optimizers are employed: Root Mean Square Propagation (RMSProp), Adaptive Moment Estimation (ADAM), and Stochastic Gradient Descent (SGD). These optimizers are widely used in deep learning for training neural networks.

The performance evaluation involves training each pretrained model on the Alzheimer's disease dataset and measuring its classification accuracy. Each model is trained using one of the optimizers mentioned above, and their classification accuracy is compared. The classification accuracy metric measures how well the models can correctly classify instances of Alzheimer's disease within the dataset.

By conducting this analysis with different optimizers, the study aims to determine which optimizer yields the highest classification accuracy for each pretrained model. This evaluation helps identify the optimizer that is most suitable for training CNNs on Alzheimer's disease datasets, potentially leading to improved diagnostic capabilities or other applications related to the disease.

### 2.1 Data Collection

In the proposed work, the datasets used for training and evaluating the CNN models are obtained from the Kaggle database. These datasets consist of various modalities of data related to Alzheimer's disease. The dataset is labeled and divided into four different classes, namely Very Mild Demented, Mild

Demented, Moderately Demented, Non Demented. These classes represent different stages or levels of dementia severity. By classifying the input data into these categories, the CNN models aim to perform Alzheimer's disease classification (AD classification). Datasets are collected from Kaggle Alzheimer's dataset 4 class of images.

The classification task involves training the CNN models using the labeled dataset, where the input data includes various modalities, such as images, clinical features, or other relevant information. The models are trained to learn patterns and features from the input data that can distinguish between the different dementia classes. By utilizing the pretrained models (AlexNet, SqueezeNet, GoogLeNet, and MobileNetV2) and fine-tuning them on the Alzheimer's disease dataset, the models can leverage their prelearned knowledge to improve the classification accuracy for AD classification.

The performance evaluation of the CNN models, as mentioned earlier, involves comparing the classification accuracy achieved by different optimizers (RMSProp, ADAM, and SGD). This analysis helps determine which optimizer performs best for each pretrained model in terms of accurately

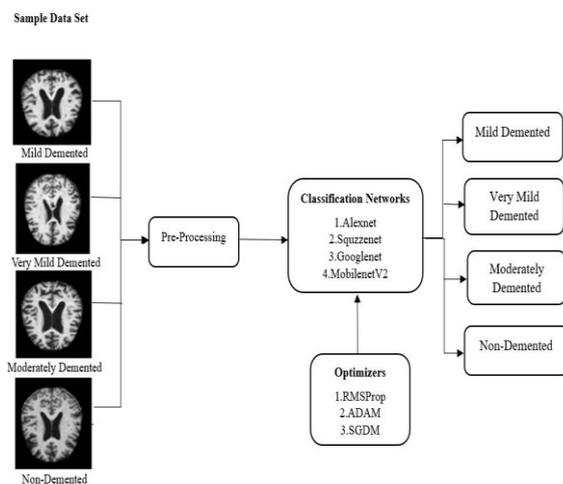


Fig 1. Block Diagram of Proposed Methodology.

classifying the input data into the four dementia classes. Overall, the goal of this proposed work is to leverage CNN models and different optimizers to accurately classify Alzheimer's disease data obtained from the Kaggle database, specifically into the classes of Very Mild Demented, Mild Demented, Moderately Demented, and Non-Demented.

## 2.2 Preprocessing

In the preprocessing step of image datasets for convolutional neural network (CNN) models, resizing the images is a common practice to ensure compatibility with the input requirements of the models. For AlexNet and SqueezeNet, the input image size is typically resized to 227x227x3. These models were popularized before the widespread adoption of the ImageNet dataset, and this input size was chosen to fit the architecture of these models. On the other hand, GoogLeNet and MobileNetV2, which were developed later, typically expect input images to be resized to 224x224x3. This input size has become a standard for many CNN

## 2.3 Data Augmentation

To enhance the size of the original datasets, data augmentation techniques are employed to create various versions of the real dataset. Computer vision and Natural Language Processing (NLP) models employ data augmentation techniques to address data scarcity and a lack of sufficient diversity. For Data augmentation generally increase the number of data samples by using various augmented techniques to train a model effectively. To expand the amount of data samples in our proposed study, data augmentation techniques such as reflection, translation, and scaling are applied.

## 2.4 Classifier

In this proposed work, four pretrained CNN models are used for performance evaluation of AD classification. They are discussed below.

**1. AlexNet:** Alex Krizhevsky unveiled AlexNet, a deep architecture, in 2012. It consists of eight learned layers, including five convolutional layers with max pooling, followed by fully connected layers. The input image size for AlexNet is 227x227x3 RGB. All layers utilise ReLU (Rectified Linear Unit) as their activation function. AlexNet computes approximately 62.3 million parameters, enabling efficient classification.

**2. GoogLeNet:** GoogLeNet, proposed by the Google team in 2014, is an architecture with 22 layers designed to improve computational efficiency. The

input image size for GoogLeNet is 224x224x3 RGB. GoogLeNet incorporates a 1x1 convolutional layer in the middle, along with global average pooling. It utilizes filter sizes ranging from 1x1 to 5x5, ReLU activation for convolutional layers, dropout for regularization, and a SoftMax classifier. GoogLeNet computes around 7 million parameters for effective classification.

**3. SqueezeNet:** SqueezeNet is an architecture proposed by researchers from DeepScale, University of California, Berkeley, and Stanford. It aims to provide a 50 times reduction in the number of parameters compared to AlexNet. The input image size for SqueezeNet is 227x227x3 RGB. The architecture starts with a convolutional layer followed by eight fire modules. ReLU is used as the activation function, and softmax is employed in the classifier. SqueezeNet analyzes approximately 1.24 million parameters, offering better classification with reduced computational complexity.

**4. MobileNetV2:** MobileNetV2 is a neural network architecture proposed by researchers from Google, specifically optimized for mobile devices. It is a 53-layer deep model. The input image size for MobileNetV2 is 224x224x3 RGB. MobileNetV2 employs inverted residual blocks with bottleneck features, distinguishing it from MobileNet. It computes a total of 3.5 million parameters for efficient classification.

These pretrained CNN models provide different architectural variations and parameter counts, enabling researchers to explore their performance in AD classification tasks. The proposed work aims to evaluate and compare the performance of these models using different optimizers, such as SGDM, ADAMT, and RMSProp, to determine the most effective approach for AD classification.

## 2.5 Optimizer

**1. Stochastic Gradient Descent (SGD):** SGD is an optimizer commonly used for training neural networks. Instead of computing the gradients for the entire dataset, SGD randomly selects a subset of samples (known as a batch) at each iteration. This approach speeds up the optimization process and allows the network to update its parameters based on

a smaller set of data. By iteratively updating the parameters, SGD aims to find the global minimum of the objective function. It is a widely used optimizer due to its simplicity and effectiveness (Haji & Abdulazeez, 2021) (Zaheer & Shaziya, 2019).

**2. Adaptive Moment Estimation (ADAM):** ADAM is an optimization algorithm that combines the concepts of both SGD and RMSProp. It maintains separate learning rates for each parameter in the network, adapting them over time based on the historical gradients. ADAM incorporates a bias correction mechanism to provide more accurate estimates of the first and second moments of the gradients. This optimizer is known for its ability to handle sparse gradients and works well in practice for a wide range of deep learning tasks (Haji & Abdulazeez, 2021) (Zaheer & Shaziya, 2019).

**3. Root Mean Square Propagation (RMSProp):** RMSProp is another popular optimizer used in deep learning. It calculates the weight updates based on the running average of squared gradients. By keeping only the current gradient information and eliminating the past gradient information, RMSProp mitigates the influence of older gradients. This helps to address the issue of slow convergence that can occur with SGD. The use of the squared gradients provides a form of adaptive learning rate, allowing the optimizer to adapt the learning rate for each parameter based on the magnitude of the gradients (Haji & Abdulazeez, 2021) (Zaheer & Shaziya, 2019).

These three gradient-based optimizers offer different strategies for updating the network parameters during the training process. By evaluating their performance in the proposed work, can able to determine which optimizer yields the best accuracy for AD classification.

## 3.Results and Discussion

In this proposed work 90% of image dataset has taken for training and remaining 10% is used for validation. Here Accuracy is considered as an evaluation metric and it's calculated by below formula,

$$\text{Accuracy} = \frac{(\text{True positive} + \text{True Negative})}{(\text{True positive} + \text{True Negative} + \text{False Positive} + \text{False Negative})}$$

### 3.1 Results

Fig 2 represents training progress of classifiers with respective optimizers. The fig 2 shows the training accuracy and loss value.

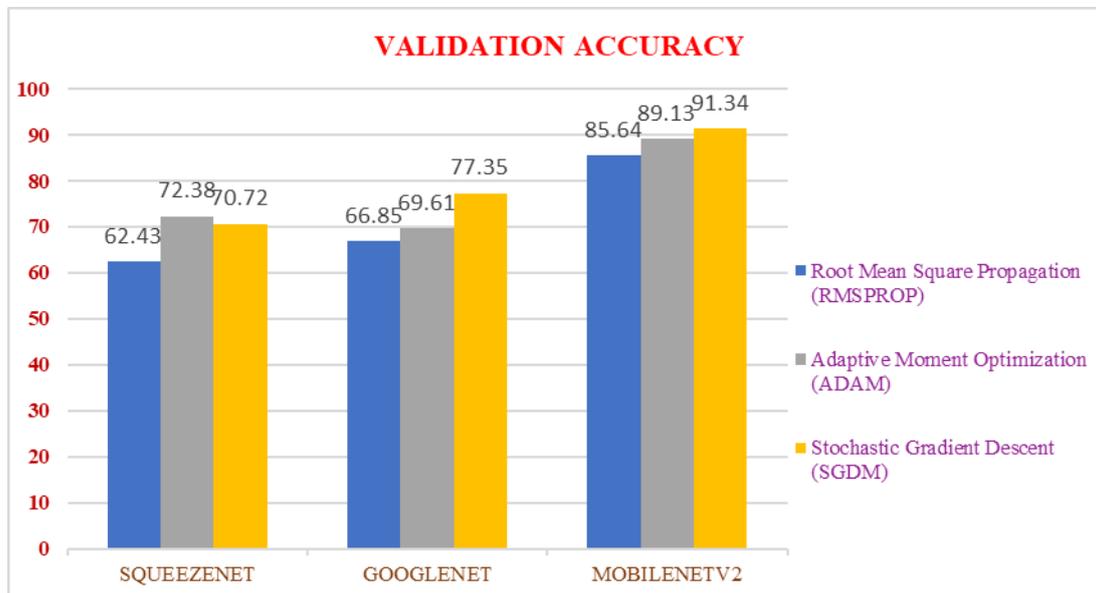


Fig 2. Training progress of CNN models using RMSProp, ADAM AND SGDM optimizers (a) - (l)

### 3.2 Performance Evaluation

The relative evaluation summary in table 1 shows that MobileNetV2 with the SGDM optimizer outperformed other deep CNN transfer learning models on the AD Kaggle dataset, achieving an accuracy of 91%. The evaluation was conducted using specific hyperparameter settings, including an epoch

size of 10, a learning rate of 0.0003, and a batch size of 10. Each epoch consisted of 489 iterations. This indicates that MobileNetV2, a lightweight and efficient CNN architecture, combined with the SGDM optimizer, yielded better results compared to other deep models on the AD Kaggle dataset. Fig 3 represents the graphical sketch of table 1.

Table 1. Accuracy of CNN models using various optimizers

DEEP NEURAL NETWORK	Root Mean Square Propagation (RMSPROP)	Adaptive Moment Optimization (ADAM)	Stochastic Gradient Descent (SGDM)
ALEXNET	57	59	61
SQUEEZENET	62	72	71
GOOGLNET	67	69	77
<b>MOBILENETV2</b>	85	89	<b>91</b>

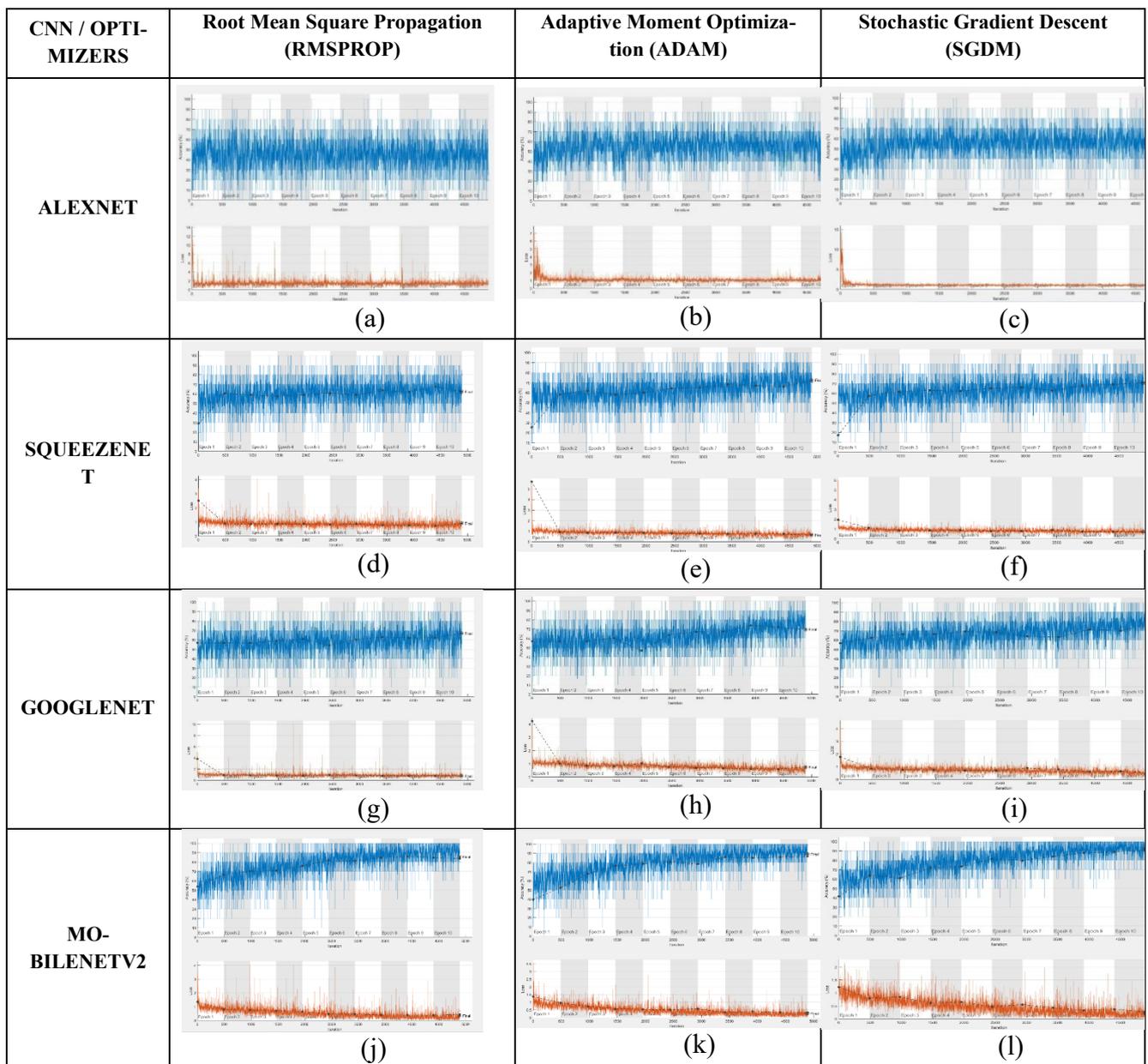
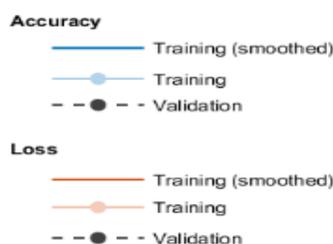


Fig 3. Graphical representation of proposed work validation Accuracies using various optimizer



### 3.3 Discussion

In contrast to the innovative methodologies investigated, our methodology presents a unique

standpoint by utilizing a Mobilenet V2 model with SGDM optimizer for the purpose of Alzheimer's disease (AD) classification. In a previous study, a convolutional neural network (CNN) transfer learning approach was employed on segmented grey matter, resulting in an accuracy of 93.11% (Raza et al, 2023). In contrast, our method achieves a commendable accuracy of 91.31% without the requirement of image segmentation, utilizing an epoch value of 10.

In a further study, a Deep Neural Network (DNN) was utilized with the optimized ADAM optimizer, leading to a notable classification accuracy of 91.29%. The success of this accomplishment was ascribed to the use of Local Binary Pattern (LBP) features- (Ghahnavieh etc.2019), underscoring the efficacy of this methodology.

**Table 2.** Comparison of proposed work with existing method

Research	Model	Accuracy (%)
Suresha et al, 2020	DNN with ADAM	91.29
Antony et al, 2022	VGG-16	81
Al-Aiad et al, 2021	VGG-16	70.3
Shruti Pallawi et al, 2023	Resnet 50 & Resnet 101 (at epoch 10)	65.82 & 58.69
Shruti Pallawi et al, 2023	InceptionV3 (at epoch 10)	79
Proposed Work	MOBILENETV2 with SGDM	91.34

In a similar manner, the VGG16 model was utilized by the author, resulting in an accuracy rate of 81% (Antony et al, 2022). Nevertheless, our approach outperforms the aforementioned outcome by attaining a precision rate of 91.31%. Furthermore, the VGG-16 model, which is a prominent architectural design (Mgddadi et al, 2021), exhibited a classification accuracy of 70.3%. This result further emphasizes the progress achieved by our methodology.

Table 1 represents the comparison of our work with existing state of art techniques. Our methodology distinguishes itself by utilizing a Mobilenet V2 model and SGDM optimizer, resulting in a noteworthy accuracy rate of 91.31% without the need for image segmentation. This statement underscores the strength and reliability of our methodology, as well as its capacity to improve the effectiveness of AD classification. Ultimately, it has the potential to advance the frontiers of accuracy and innovation within this discipline.

#### 4. Conclusion

This paper provides an extensive analysis of four conventional pretrained CNN models in the context of Alzheimer's disease (AD) classification. The focus of

the analysis is on evaluating the performance of different gradient-based optimizers, including SGDM, ADAMT, and RMSProp, to enhance the CNN models. The objective of the study is to demonstrate the effectiveness of various training optimizer techniques in improving the performance of CNN models for AD classification. The comparative evaluation conducted in the paper reveals that the SGDM optimizer achieved the highest accuracy of 91% for AD classification among the optimizers tested. The findings presented in the paper can be valuable for researchers and practitioners in the field of deep learning and medical image processing. By considering the comparative evaluation of deep CNN models and optimizers, researchers can make informed decisions regarding the selection of appropriate models and optimizers for achieving optimum AD classification performance. It's worth noting that the specific performance and accuracy values mentioned in the paper are specific to the experiments conducted by the authors. The performance of deep learning models can vary depending on factors such as the dataset, hyperparameter settings, and specific task requirements. Therefore, it's important for researchers to conduct their own evaluations and consider the specific context of their work when selecting models and optimizers. In future optimization algorithm can be applied to hypertune the deep network parameters also segmentation technique may be used for better classification.

#### References

- Ebrahimi-Ghahnavieh, A., Luo, S., & Chiong, R. (2019), Transfer Learning for Alzheimer's Disease Detection on MRI Images, The 2019 IEEE International Conference on Industry 4.0, Artificial Intelligence, and Communications Technology (IAICT), IEEE 2019.
- Doaa Ahmed Arafa, Hossam El-Din Moustafa, Hesham A. Ali, Amr M. T. Ali-Eldin, & Sabry F. Saraya. (2023). A deep learning framework for early diagnosis of Alzheimer's disease on MRI images, Multimedia Tools and Applications, 2023, <https://doi.org/10.1007/s11042-023-15738-7>
- Antony, F., Anita, H.B., George, J.A. (2023). Classification on Alzheimer's

- Disease MRI Images with VGG-16 and VGG-19. vol 312. Springer,
- Halebeedu Subbaraya Suresha, Srirangapatna Sampathkumaran Parthasarathy. (2020). Alzheimer Disease Detection Based on Deep Neural Network with Rectified Adam Optimization Technique using MRI Analysis, 978-1-7281-9183-6/20/\$31.00 ©2020 IEEE.
- Junhao Wena, Elina Thibeau-Sutre, Mauricio Diaz-Meloe, Jorge Samper-González, Alexandre Routiere, Simona Bottanie, Didier Dormonte, Stanley Durrleman, Ninon Burgos, Olivier Colliot. (2020). Convolutional Neural Networks for Classification of Alzheimer's Disease: Overview and Reproducible Evaluation, *Medical Image Analysis*, volume 63, July 2020,101694.
- Mggdadi E, Al-Aiad A, Al-Ayyad MS, Darabseh A. (2021). Prediction Alzheimer's disease from MRI images using deep learning. 2021 12th International Conference on Information and Communication Systems, ICICS 2021, pp. 120–125.
- Noman Raza, Asma Naseer, Maria Tamoor & Kashif Zafar. (2023), Alzheimer Disease Classification through Transfer Learning Approach, *Diagnostics* 2023, 13, 801. <https://doi.org/10.3390/diagnostics13040801>
- P.S. Rallabandi, K. Tulpule, & M. Gattu. (2020). Automatic classification of cognitively normal, mild cognitive impairment and Alzheimer's disease using structural MRI analysis, *Informatics Med. Unlocked*, vol.18, 2020, DOI:10.1016/j.imu.2020.10030 5.
- Zaheer, Raniah, & Shaziya, Humera. (2019). A Study of the Optimization Algorithms in Deep Learning, *International Conference on Inventive Systems and Control (ICISC 2019)* IEEE Xplore Part Number: CFP19J06-ART; ISBN: 978-1-5386-3950-4. <https://www.kaggle.com/datasets/tourist55/alzheimers-dataset-4-class-of-images>
- Saad Hikmat Haji & Adnan Mohsin Abdulazeez. (2021). Comparison of Optimization Techniques Based On Gradient Descent Algorithm: A Review, *PJAEE*, 18 (4) (2021).
- Shruti Pallawi & Dushyant Kumar Singh. (2023). Review and analysis of deep neural network models for Alzheimer's disease classification using brain medical resonance imaging, *Cognitive Computation, and Systems*, Volume 5, Issue 1, Mar 2023.
- Subramanyam Rallabandi, Ketki Tulpule, Mahanandeeswar Gattu. (2020). Automatic classification of cognitively normal, mild cognitive impairment and Alzheimer's disease using structural MRI analysis, *Informatics in Medicine Unlocked*, Volume 18,2020,100305.
- V.P. Sub Jun Zhang, Mingxia Liu, Le An, Yaozong Gao, & Dinggang Shen. (2016). Alzheimer's Disease Diagnosis using Landmark-based Features from Longitudinal Structural MR Images,2168-2194 (c) 2016 IEEE.
- Fu'adah, YN, Wijayanto1, I., Pratiwi1, NKC, Talingsih, FF, Rizal1, S, & Pramudito. M A. (2020). Automated Classification of Alzheimer's Disease Based on MRI Image Processing using Convolutional Neural Network (CNN) with AlexNet Architecture, *Journal of Physics*, IOP Publishing, 1844(2021)012020.
- Helaly, Hadeer A., Badawy, Mahmoud, & Haikal, Amira Y. (2022). Deep Learning Approach for Early Detection of Alzheimer's Disease, *Springer, Cognitive Computation* (2022) 14:1711–1727.

## AUTHOR BIOGRAPHIES



**Sindhu T.S.**, a dedicated scholar and Assistant Professor at C. Abdul Hakeem College of Engineering and Technology, Melvisharam, Tamil Nadu, holds a B.E. in Electronics and Communication Engineering from Thanthai Periyar Government Institute of Technology, Anna University, Chennai (2014), and an M.E. in Applied Electronics from Kingston Engineering College, Anna University, Chennai (2016). Currently pursuing her Ph.D. at Anna University, Chennai, her research interests span Image Processing, Biomedical Research, Artificial Intelligence, and Machine Learning. With approximately 4 journal publications and 12 international conference proceedings, she showcases a commitment to scholarly contributions, blending expertise with a passion for advancing knowledge in these dynamic fields.



**N. Kumaratharan**, obtained his Ph.D. in the field of Wireless Communication from Pondicherry University in 2010 and is currently working as a Professor in the Department of Electronics and Communication Engineering at Sri Venkateswara College of Engineering, Sriperumbudur, India. His area of research includes Wireless Communication, Wireless Sensor Networks and Biomedical Signal Processing. He has more than 19 years of teaching and research experience. He has guided Twelve Ph.D. scholars of Anna University, Chennai. He has published 36 international journals indexed by SCI, Scopus and WoS. He has 55 conference proceedings indexed by Scopus. He is a fellow member of IETE and a senior member of Association of Computer Electronics and Electrical Engineers, member of International Association of Engineers and Life member of Indian Society for Technical Education.



**Dr. P. Anandan** received his B.E. Degree in Electrical and Electronics Engineering from Vellore Engineering College, University of Madras, Vellore, Tamil Nadu in 2002. M.E., in Applied Electronics from CEG, Anna University, Chennai, Tamil Nadu in 2004 and Ph.D. in VLSI and Nano Electronics in Anna University, Chennai, Tamil Nadu in 2015. His research interest includes VLSI and Nano Electronics. Currently, he is working as a Professor in Saveetha School of Engineering, Saveetha University, Chennai. He has about 28 international journal publications in reputed journals and about 40 international conference proceedings.

# Revolutionizing age and gender recognition: an enhanced CNN architecture

Rashna Sharmin Tumpal, Md. Khaliluzzaman<sup>1\*</sup>, MD Jiabul Hoque<sup>2</sup>, Roshni Tasnim<sup>1</sup>

<sup>1</sup> Department of Computer Science and Engineering, International Islamic University Chittagong, Chattogram-4318, Bangladesh

<sup>2</sup> Department of Computer and Communication Engineering, International Islamic University Chittagong, Chattogram-4318, Bangladesh

\* Corresponding author E-mail: khalil@iiuc.ac.bd

(Received 27 November 2023; Final version received 22 May 2024; Accepted 31 July 2024)

## Abstract

The recognition of age and gender in images has had a significant impact on computer vision, particularly with the increasing application of digital platforms. Deep Convolutional Neural Networks (DCNNs) show promising performance. However, they demand substantial computational resources, limiting their deployment in real systems, especially those with constraints on resources or cost. This study performs a sensitivity analysis in order to show how some changes in the architecture of the network can influence the tradeoff between accuracy and performance. For that, in this work, we have investigated various CNN architectures and introduced an effective convolutional neural network (CNN) model to precisely predict gender and age attributes using the Adience dataset. Amidst unfiltered and diverse image sources from various devices, our model exhibits an impressive 92.24% accuracy across eight distinct age groups and two gender categories. The model's strength lies in its adeptness at handling intricate image data, allowing comprehensive adjustments to age and gender parameters. By employing advanced deep learning techniques and comparing with MiniVGGNet, our model showcases exceptional performance.

*Keywords: Adience dataset, Age and Gender recognition, Computer Vision, Deep Convolutional Neural Network (DCNN), MiniVGGNet.*

## 1. Introduction

Facial attribute recognition, particularly age and gender classification, has garnered considerable attention in recent years due to its multifaceted applications in various domains such as security, marketing, computer vision, and human-computer interaction (Smith et al, 2021; Parkhi etc., 2015). The ability to accurately infer age and gender from facial features has been a longstanding challenge. However, advancements in deep learning, especially with convolutional neural networks (CNNs), have revolutionized this field (Liu and Chen, 2020).

Researchers have thoroughly investigated gender and age identification using a variety of handcrafted features and machine learning techniques (Smith et al, 2021). The pursuit described encompasses a wide range of implications, which include applications in the field of robotic programming that necessitate the use of demographic information. Additionally, it extends to surveillance systems that engage in human-computer interaction, human pose estimation, automated analysis of

gender-related behaviors, and biometrics, as well as access management in secure environments (Parkhi etc., 2015). These various applications highlight the significance and potential impact of this pursuit. The customized Convolutional Neural Network (CNN) model designed for age and gender classification holds substantial practical consequences and has the potential for diverse applications in real-world situations. On digital platforms, it can improve user experiences by offering customized content, recommendations, and advertisements that are specifically designed for the user's demographic profile. The model in surveillance systems assists in the identification of demographic patterns, enhancing crowd management and enabling focused monitoring in public settings (Toshev and Szegedy, 2015). Rapid patient classification in healthcare applications can enhance the development of tailored treatment programs. Businesses can enhance their marketing and advertising efforts by doing a more precise analysis of customer demographics. Smart devices can utilize this framework to provide customized interactions, such as modifying the responses of voice assistants according to

the user's age and gender. In addition, educational platforms have the ability to tailor learning materials to better align with the age and gender of students, hence improving their learning experience. These applications highlight the broader importance of the model, demonstrating its potential influence in other fields by offering accurate age and gender categorizations (Smith et al, 2021).

In recent years, there has been a notable demonstration of the impressive capabilities of deep learning and convolutional neural networks (CNNs) (LeCun et al., 2015). These developments have demonstrated outstanding performance in the field of computer vision, namely in tasks like image classification and recognition. The ability of convolutional neural networks (CNNs) is to iteratively acquire complex information from the first to the final layers. That resulting in the creation of more abstract and advanced representations, which signifies a significant paradigm shift in the domain (Liu and Chen, 2020).

The contemporary state of age and gender recognition pose complex issues that persistently influence research endeavors within the field of computer vision. The domain continues to encounter persistent challenges in the areas of apparent age estimate, accurate age prediction, robust face recognition, and combined identification and verification (Parkhi et al., 2015; Yang et al., 2018). It has wide-ranging societal and technological consequences. Precise identification improves user experiences on digital platforms by providing tailored content and services. It enhances public safety by supporting surveillance systems in doing demographic analysis. Within the healthcare field, the use of patient profiling enables rapid assessment of patients, hence enhancing the development of treatment strategies. Marketing businesses derive advantages by employing focused methods that are tailored to specific demographic profiles. Educational technology modifies content to more effectively align with students' age and gender. Smart devices enhance their intuitiveness by analyzing and comprehending user demographics. This research fulfills the need for sophisticated and dependable techniques of classifying age and gender. It fosters advancements in computer vision and artificial intelligence, facilitating technological growth (Rothe et al., 2016; Sun et al., 2014).

The utilization of convolutional neural networks (CNNs) and deep learning methods has been propelled as transformational instruments in the discipline in response to these issues. These technologies present

intriguing opportunities for improving accuracy, reducing biases, and achieving resilience against fluctuations in illumination, posture, facial expression, and biases within datasets. The convergence of convolutional neural networks (CNNs) and deep learning is positioned at the forefront of technological progress, enabling breakthroughs in addressing the complexities associated with age and gender recognition tasks.

This paper introduces a customized developed convolutional neural network (CNN) model. Its design is customized for the accurate classification of gender and age. The model underwent training and evaluation on the Adience dataset. The model showcased exhibits a high level of precision, achieving a remarkable accuracy level of 92.24%. The analysis was conducted on a varied dataset comprising of eight distinct age groups and two gender divisions. We improved the performance of our model by utilizing advanced deep learning techniques. We conducted rigorous assessments of MiniVGGNet. This approach required making incremental adjustments to parameters in a total of ten models. Consequently, we achieved an exceptional level of accuracy. Our approach focuses performing a comprehensive evaluation utilizing the Adience standard. This highlights the model's capacity to efficiently process a diverse array of raw image data acquired from various sources, such as smartphones. This approach enables significant adjustments to age and gender variables. Therefore, it enhances the complexity of the dataset and establishes a robust framework for evaluation. Furthermore, we have incorporated a distant-view Convolutional Neural Network (CNN) model for the sake of comparison. This further strengthens the exceptional performance of our proprietary approach. The importance of our model in the field of gender and age prediction is reinforced by doing comparative comparisons with other prominent methods. This substantiates the superiority of our model in these areas. The contribution of the paper is provided below:

- Investigates various CNN architectures and introduced an effective convolutional neural network (CNN) model to precisely predict gender and age attributes using the Adience dataset.
- Model's performance on the Adience benchmark dataset was subjected to a rigorous examination.
- Implementation of a baseline Convolutional Neural Network (CNN) model is employed to establish a comparative benchmark for distant view images.

- Model's efficacy is demonstrated by its exceptional achievement of 92.24% accuracy over eight age and two gender groups.
- Our suggested model has exhibited superiority over state-of-the-art approaches, hence confirming its robustness and excellence in predicting gender and age.

The rest of the paper is summarized as follows. The related work is detailed at section 2. The proposed method is presented at section 3. The experimental analysis is explained at section 4. The paper is concluded at section 5.

## 2. Related works

Deep convolutional neural networks (CNNs) have played a key part in advancing age and gender classification, which are critical components of facial recognition technology. This comprehensive literature review examines several approaches, systems, and tactics. Those have been proposed in recent decay with the goal of improving accuracy in recognition tasks. Many authors in various articles examine innovative techniques. The strategies include shallow convolutional neural network (CNN) designs, ensemble learning, and attribute identification algorithms. The primary objective is to improve the precision of age and gender predictions. This thorough analysis investigates the development of these approaches. It emphasizes their noteworthy findings. Furthermore, it examines the diverse applications in the domain of face attribute recognition. Such as, Patil et al. (2021) show that performance can be significantly improved by learning representations while using deep convolutional neural networks (CNN) and Extreme Learning Machines (ELMs). The methodology employed in this study utilizes CNN and ELM to acquire fundamental representations. The CNN module is responsible for extracting features from input images, while the ELM module is responsible for classifying the intermediate results. This research advised to use CNN for giving more accurate results of age and gender classification.

In the study, Abir et al. (2023) put out a comprehensive facial recognition framework that integrates identification, age, and gender recognition. Their research showcases the capacity of machine learning to make complex predictions. Srivastava and colleagues (2023) provide research indicating a notable rise in the

They demonstrated a system for real-time gender recognition on an embedded device that achieves very high accuracy (up to 98.73% in the field).

utilization of deep neural networks, namely convolutional neural networks (CNNs), to effectively demonstrate the accuracy of age and gender categorization. The observed accuracy surpasses that of other previously established claims. The model consists of three convolutional layers, two max pooling layers, and two dense layers. The study revealed that the mean accuracy for age recognition is 82.2%, while the mean accuracy for gender recognition is 94.10%. The authors Balan et al. (2022) have introduced the concept of a Pike neuron based convolutional neural network (SN-CNN). The accuracy of age and gender recognition was assessed by the evaluation of the orthopantomogram dataset, yielding a significant degree of performance. Abood et al. (2023) introduced a fresh methodology in their research, wherein they employed the AlexNet model to categorise age and gender. The performance evaluations were conducted by utilising the UKTFace dataset. The current investigation successfully attained a reasonable degree of precision in the classification of age and gender. The introduction of a novel convolutional neural network (CNN) classification technique was presented by Mamatkulovich et al. (2023) in their paper. This algorithm demonstrates notable advantages over current approaches, as it exhibits much reduced training parameters and training time. In spite of its lower complexity, our model exhibited superior accuracy in classifying age and gender on the UTKFace dataset. In a previous study, Sharma et al. (2022) put out an enhanced convolutional neural network (CNN) model designed for the assessment of age and gender. The authors conducted a comparative analysis of their method with the current state-of-the-art approach using the UKTFace dataset, and achieved a respectable level of accuracy.

The authors Agbo-Ajala et al. (2020) explore techniques for automatically determining a person's age and gender from raw images of their faces. They framed the challenge as one of multiple classes and suggested using a loss function based on categorization to guide model training. They studied the precision with which this model could classify data using the original dataset. Proposed model obtains the state-of-the-art performance in both age group and gender classifications. Shallow deep convolutional neural network architecture for Gender Recognition from face images was proposed by Greco et al. (2020) research. Author also conducted sensitivity analysis to demonstrate how varying the network's architecture can modify the efficiency-accuracy tradeoff.

A deep learning based network called GRA\_Net is presented for estimating a person's age and gender based on their appearance by Garain et al. (2021). The

author views the issue of age prediction as a hybrid of categorization and regression. The authors Nada et al. (2020) propose a new method for verifying a user's gender and age range as determined by their image. Their Deep Learning-based inputs now include a layer validator that checks inputs like user image, gender, and birth date. Using a Convolutional Neural Network, they were able to determine a person's gender and approximate age from just one image. After analysis, it was shown that this approach accurately predicted both gender and age. Hassan et al. (2020) propose a different model for the gender and age classification problem employing multiple sub-CNN is proposed. Here, each sub-CNN was evaluated separately, in addition to the final model evaluation that applied the voting ensemble. The idea behind employing several smaller CNNs and combining their predictions into one larger one is to increase accuracy. They receive a diverse representation for the photographs Feature. Therefore, improving classification accuracy necessitates developing a more accurate model of the age estimation problem. Compared to employing a single CNN model, the error rate in all age group classifications is lower when the voting ensemble method is used.

Benkaddour et al. (2021a) introduce three CNN network models with various architectures were tested using the number of filters and number of convolution layers. Authors showed that CNN networks significantly increase the system's usability and recognition accuracy. Later, Benkaddour et al. (2021b) proposed another CNN method for age and gender recognition to reduce the complexity of the proposed model. The model is evaluated based on the Adience dataset and achieved the accuracy for age is 91.75% and for gender is 95.6% respectively. The authors of Levi et al. (2015) explored and proposed a novel approach by exploiting the neighborhood information among image samples, which showed that accurate attribute detection is possible by utilizing the automatically generated neighborhood graph topology. A DNN regression is shown in Human Posture Estimation through Deep Neural Networks by Toshev et al. (2014). This research achieves high precision pose estimations and has completed their work on real-world images on four academic benchmarks. Author propose a DNN-based pose predictors, which allowed to increased precision of joint localization, where they focuses on the relevant region of the image, which is then cropped, and the pose displacement regression is combined with the

sub-image. During training, simulated predictions are generated.

Ozbulak et al. (2014) introduce a method, where, a collection of Deep hidden Identity features (DeepID) are learned using deep learning and are derived from the final hidden layer neuron activations of deep Convolutional networks (ConvNets). It has a 160-dimensional DeepID system that can accurately predict 10,000 classes at the end of a cascade. DeepID reduced the number of neurons in the top layer and extracted features as it learned to recognize classes in the training set. Additionally, it eliminates a few neurons from the hidden layers. They combined the Bayesian strategy with DeepID-based neural network training for face verification. In the paper Sun et al. (2016), transferring cutting-edge deep Convolutional Neural Network models is investigated for autonomous age and gender prediction with two widely used soft biometric features. Two different face recognition models—one general and one specific to the face recognition domain—were chosen to explore whether these models can be used to classify people by age and gender. For age and gender classification tasks, few findings were made by comparing generic AlexNet-like and domain-specific VGG-Face CNN models against task-specific GilNet CNN models on the difficult Adience benchmark. According to experimental findings, transferred models perform 7% and 4.5% better for age and gender categorization tasks than the most recent GilNet model respectively.

A CNN for Pedestrian Gender Recognition is suggested by Ng et al. (2013) using a convolutional neural network that had been trained to discriminate between both genders of pedestrians. They achieved an accuracy of 80.4% on a dataset consisting of frontal and rear images of pedestrians' whole bodies using a relatively simple architecture and minimal picture pre-processing. A deep CNN was proposed by Raza et al. (2017) to determine the gender of a pedestrian. The method employs a pre-existing deep decomposition neural network to examine the pedestrian's images. Next, the backdrop is removed from the parsed photos, and whole-body and head-and-shoulders images are created. Later, we feed these two categories of images into the recommended fine-tuned CNN model. Input images of the full body are sorted by gender based on whether they were taken from the front, the back, or a combination of the two. Images are split into eight categories based on the attire on their upper bodies. The proposed method was found to be more effective in

making predictions across a wide range of sub-classifications.

According to author knowledge, this paper (Cao at el., 2008) is the first to investigate gender recognition using still images of the human body. Their PBGR system, which combines part-based representation with ensemble learning, can identify the gender from a single frontal or back view image with an accuracy of 75.0% and also act robustly in the presence of slight misalignment. In the paper by Deng at el. (2014), a brand-new, massive dataset (PETA) containing 19000 images and 61 annotated attributes is presented. By utilizing the neighborhood information between image samples, they investigated and presented a novel strategy to deal with such big and heterogeneous data in the context of attribute classification. Author demonstrated that using the automatically inferred neighborhood graph topology, accurate attribute detection is possible. Ranjan at el. (2017) present a multi-task CNN-based technique for simultaneous face detection, face alignment, posture estimate, gender and smile is classification, and age estimation. Although both their method and Hyper Face use the MTL framework, their way performs far better. This work demonstrates how domain-based regularization and network initialization from face recognition tasks help subject-independent tasks.

Author provided a complete method for recognizing age, gender, and emotions by Dehghan at el. (2017). They demonstrate that their innovative deep architecture can outperform competing commercial and academic algorithms on a number of benchmarks when combined with their substantial, internally acquired data. In order to conduct attribute identification of a set of pedestrian images, straightforward deep network architecture was presented by Kurnianggoro at el. (2017). Using a public dataset, experiments were carried out to evaluate the effectiveness of this network. The results show that the suggested network performs better with fewer parameters used. Additionally, it is discovered that several characteristics, such as the presence of a backpack and clothing color, cannot be accurately identified because of self-occlusion or color confusion. As a result, it can be concluded from the experiments in this work that the particularly

designed network performs better than the general purpose network. Hence, it is suggested to use a specially designed network for person re-identification research while considering human attributes as one of its components. The RoR-152+IMDB-WIKI-101 model by Zhang at el. (2017), while combined with age-group mechanisms, demonstrates superior performance compared to traditional CNN architectures. This highlights its ability to effectively estimate age and gender in complex image datasets encountered in real-world scenarios. The fine-tuning method used to Adience dataset enhances and reinforces its already exceptional performance, establishing it as a state-of-the-art model.

The literature investigate incorporates a comprehensive examination of various methodologies in age and gender recognition. The authors argue in favor of utilizing deep convolutional neural networks (CNNs) as a means to improve accuracy. They propose various techniques, including as shallow architectures, multi-CNN ensembles, and compact models, to achieve this objective. The researchers investigate many tasks, including the identification of gender in pedestrians, estimation of posture, and detection of attributes. They employ innovative approaches that prioritize domain-specific learning, dataset size, and specialized network designs to enhance performance across diverse recognition tasks. Presented here is a tabular representation that concisely summarizes the data extracted from the present state of the art. The summary of the related work is shown in Table 1.

### 3. Research method

This section of our research work represents the detailed concept on recognizing the gender and age estimation by using deep neural network. The method proposed in this paper represents a significant improvement in the field of facial attribute recognition, namely in the area of gender and age classification. By utilizing convolutional neural networks (CNNs), our methodology seeks to tackle the complex task of precisely classifying gender and age features in facial images. By utilizing the distinct characteristics of the Adience dataset as described by Eidingner etc. (2017)

**Table 1.** Summary of the related works

Reference	Focus	Key Findings
Abir etc. (2023)	Integration of age and gender recognition inside a facial recognition framework.	A comprehensive framework for facial recognition has been developed, which integrates identification, age, and gender recognition. This framework places particular emphasis on the capacity of machine learning to make intricate predictions.
Srivastava etc. (2023)	Use of convolutional neural networks for reliable age and gender classification.	Demonstrated growing interest in using CNNs for accurate age and gender categorization; using a model with three convolutional layers, achieved mean accuracy rates of 82.2% for age and 94.10% for gender, both improvements above prior work.
Balan etc. (2022)	Recognition of both age and gender is remarkably accurate.	Presented a convolutional neural network (CNN) based on Spiking neurons (SN-CNN), with impressive results on the orthopan-tomogram dataset for identifying age and gender.
Abood etc. (2023)	Accuracy in determining age and gender is satisfactory.	Developed a novel method for age and gender categorization using the AlexNet model, which showed promising results on the UKTFace dataset.
Mamatkulovich etc. (2023)	Creation of a new convolutional neural network method for age and gender discrimination	has presented a unique CNN classification technique that, while attaining better accuracy on the UTKFace dataset, requires fewer training parameters and less time overall to implement.
Sharma etc. (2022)	Fairly precise gender and age determinations	Built a better CNN model for determining age and gender, with results that are competitive with state-of-the-art methods on the UKTFace dataset.
Patil at el. (2021)	Age and Gender recognition using CNN and ELM	Convolutional Neural Networks (CNNs) and Extreme Learning Machines (ELMs) utilized to acquire fundamental representations for better outcomes
Agbo-Ajala at el. (2020)	Age and Gender Multi-class Classification	Using unedited frontal imagery and multiclass classification algorithms, achieved state-of-the-art performance in age and gender classification.
Greco at el. (2020)	Condensed CNN for Identification of Gender	Suggested a small CNN architecture that can be used in embedded devices and achieves high real-time gender recognition accuracy of up to 98.73%.
Garain at el. (2021)	GRA_Net for Age and Gender prediction	Presents GRA_Net, a network that uses facial image classification and regression to handle age prediction.
Nada at el. (2020)	Verification of Gender and Age with Images	Using CNN, a double-check layer validator was proposed to reliably determine gender and age from single-person images with remarkable results.
Hassan at el. (2020)	Group of Sub-CNNs for Gender and Age	showed increased accuracy in age estimate and classification by the use of a voting ensemble to combine predictions from several sub-CNNs.
Benkaddour etc. (2021a)	CNN Topologies for Enhanced Precision	Investigated several CNN architectures, showing notable advances in recognition systems' usability and accuracy.
Benkaddour etc. (2021b)	Simple CNN architecture to reduce the complexity	A simple CNN model is formed to estimate the age and gender based on the Adience dataset.
Levi at el. (2015)	Identifying attributes through neighborhoods	Demonstrated precise attribute identification using the neighborhood graph topology among picture samples that was automatically created.
Toshev at el. (2014)	Human Posture using DNN Regression	Using DNN regression and concentrating on joint localization and simulated predictions during training, the model were able to estimate human posture with great precision.

Ozbulak at el. (2014)	Features of DeepID for Face Verification	Created a 160-dimensional DeepID system by combining Bayesian method with deep learning and neuron reduction for facial verification.
Sun at el. (2016)	Transferring CNN Models according to Gender and Age	Examined the transfer of cutting-edge CNN models for gender and age prediction; on tasks requiring gender and age classification, the models outperformed current ones by 5% and 7%, respectively.
Ng at el. (2013)	Gender Recognition in Pedestrians	Suggested a simple CNN architecture that achieved 80.4% accuracy in gender classification of pedestrians with little preprocessing.
Raza at el. (2017)	Gender Analysis Using Pedestrian pictures	A deep CNN model was used to evaluate gender utilizing different viewpoints and categorize apparel, resulting in improved prediction performance.
Cao at el., (2008)	Gender Identity in the Human Body	Designed PBGR system that demonstrates robustness to misalignment and 75.0% accuracy in gender recognition from human body photos.
Deng at el. (2014)	Finding attributes in Large-Sample Datasets	Demonstrated efficient attribute classification by presenting a method for neighborhood information-based correct attribute detection in a big dataset (PETA).
Ranjan at el. (2017)	Multi-Task CNN for Diverse Features	Presented a multi-task CNN method that outperformed Hyper Face utilizing domain-based regularization for face identification, alignment, posture, gender, smile categorization, and age estimation.
Dehghan at el. (2017)	Robust Framework for Age, Gender, and Emotions	Outperformed current algorithms in identifying emotions, age, and gender by utilizing a deep architecture that was novel and heavily data-driven.
Kurnianggoro at el. (2017)	Identification of Attributes in Images of Pedestrians	Improved attribute identification efficiency with fewer parameters by proposing a specialized network for person re-identification study, which is especially helpful in self-occlusion settings.
Zhang at el. (2017)	Introduced two mechanisms to enhance age estimation	Proposed RoR-152+IMDB-WIKI-101, a novel CNN-based method for age and gender estimation.

and implementing advanced deep learning methods, we have developed a tailored model. This model explores the complex correlation between classes, hence enabling precise classification.

The aim of this methodological innovation is to surpass conventional methods. Our objective is to improve the accuracy and reliability of our predictions by developing a more extensive and simple framework for predicting age and gender. This innovative approach aims to make a significant contribution to the field of facial attribute recognition by introducing a new paradigm. By doing so, we want to improve the

precision and effectiveness of convolutional neural network (CNN) models in handling intricate classification tasks. Fig 1 presents the workflow diagram of this research; which is explained in following with a sequential way.

Here, the dataset we are going to use in our research work is Adience dataset. It contains 19322 images which resolution is 816 by 816 with the horizontal and vertical resolution of 96 dpi and 24 bits. Fig 2 represents some sample image of different attribute class from the Adience dataset. Input images are close-view image.

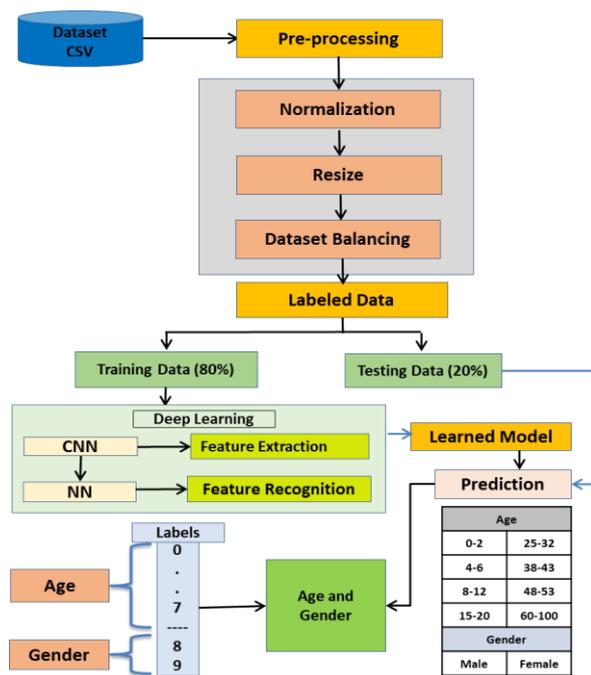


Fig 1. Workflow diagram of the age and gender recognition framework with data preprocessing.

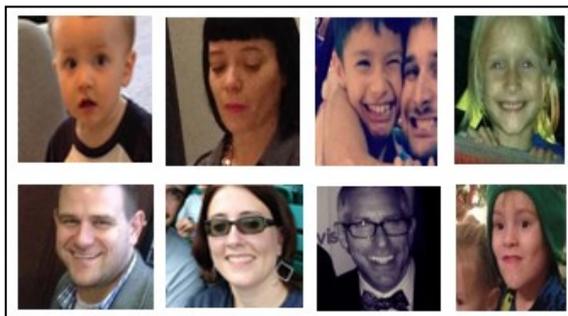


Fig 2. Sample image of different attribute class from Adience dataset

According to the steps to get preprocessed data, images are rearranged in the class label first. Within 30 age classes, we create 8 age classes, and from 3 gender classes, we take 2. We include all the age classes in those eight classes. Our age classes are 0-2, 4-6, 8-12, 15-20, 25-32, 38-43, 48-53, and 60-100, and our gender classes are male and female. Now, after reading the entire csv file, the image has been converted into channel 1, i.e., gray scale image. After that, the images are divided by 255.0, minus 0.5, and multiplied by 2.0 for scaling. After scaling, resize the original image into 75 by 75 images and give the image ID from 1 to increase, plus 1 for the next image. Then create five pickle files for each of the five CSV files. Now read the pickle files and make a directory. By reading the csv file, the image is divided into different folders, which are divided into age class and then male and female class, respectively.

When reading every label line from the final label text file, the newline characters should be stripped from the end of every line to prevent extra labels from showing up. Then rename the image according to the label of age, class, and gender, and then add a random number. All the images are in jpg format. The image file name is like (0-2)\_m\_503.jpg, which means age is 0 to 2 and it is male. Some preprocessed image data is shown in Fig 3. Finally, labeled data are divided into train and test data, following the partition of 80% and 20%.



Fig 3. Some preprocessed image data from all class

### 3.1 Proposed CNN model

To achieve the best CNN model for predicting age and gender, we have applied different CNN layers and dense layers. From which we can estimate the best model for achieving our goal. For that, the model is trained with the various CNN and dense layers, namely NewNet1, NewNet2, ..., NewNet9. All the CNN layers are contained filters of size 3x3. The training and testing accuracy for different configurations are given in Table 2. From these entire networks, the NewNet3, NewNet8 and NewNet10 performed better. However, we select model 8 which is named here as NewNet8 because it gives higher training and testing accuracy with small number of parameter, where all other takes huge number of parameter and gives low training and testing accuracy. So finally, we choose the model NewNet8 is our proposed model. For train all the models Adam optimizer (Kingma et al., 2014) and categorical cross entropy (Liu and Chen, 2020) for loss function are used. For each model here used 50 epochs with the batch size of 128.

Based on the NewNet8 model, now we discuss the proposed convolutional network architecture given in Fig 4. It contains four convolutional layers and two fully connected layers that use the extracted features from the previous convolutional layers and acquire from them and thus perform the classification at the output layer.

Here, Close view RGB images of size 75 x 75 with three channels will be input to this convolutional neural network model. The first convolutional layer consists of 32 filters of size 3 x 3 and 3 channels each. The SAME padding is used here i.e. the input images are zero padded in such a way that the filters convolve over every pixel of the input image. For SAME padding the output image after convolution is the same as the input image. The 32 filters convolve the input images of size 75 x 75. Because of the SAME padding, the output feature maps after convolution of size 75 x 75 and 32 channels. To

introduce non-linearity the output feature map is run through a ReLU activation function. This activation function turns any negative value into zero. Thus it makes every value non-negative. The max pooling filter is of size 2 x 2 and it is used to reduce the dimensionality of the feature maps. This filter moves by a stride of 2. And the output image size is 37x37x32.

The second convolutional layer contains 64 filters of size 3 x 3 and 64 channels each. The SAME padding is also used here. 64 filters are convolving the input

**Table 2.** Testing and training accuracy for the different model configurations

Model	CNN Layers	Dense Layers		Parameters	Train Accuracy (%)	Test Accuracy (%)
		Neurons	Dropout	Trainable		
NewNet1	64-128-256	1024 512	0.5 0.3	76,665,226 76,663,306	93.46	91.24
NewNet2	64-128-256	1024 512	0.5 0.4	76,665,226 76,663,306	93.39	89.19
NewNet3	32-64-128-256	1024 512	0.5 0.3	17,700,554 17,698,570	94.39	92.16
NewNet4	64-128-256	1024 256	0.5 0.3	76,399,242 76,397,834	93.54	91.84
NewNet5	64-128-256	1024 256	0.5 0.4	76,399,242 76,397,834	93.28	91.57
NewNet6	64-64-128-256	1024 256	0.5 0.4	17,454,026 17,452,490	94.50	92.15
NewNet7	32-64-128-256	1024 256	0.5 0.4	17,434,570 17,433,098	94.55	91.90
<b>NewNet8</b>	<b>32-64-128-256</b>	<b>1024 256</b>	<b>0.5 0.3</b>	<b>17,434,570 17,433,098</b>	<b>94.60</b>	<b>92.24</b>
NewNet9	32-64-128-256	1024 512	0.4 0.3	17,700,554 17,698,570	94.56	91.85
NewNet10	32-64-128-256	1024 256	0.4 0.3	17,434,570 17,433,098	94.39	92.15
NewNet11	64-128-256	1024 1024	0.4 0.3	77,197,194 77,194,250	93.17	91.78
MiniVGGNet	32-32-64-64	512	0.5	10,690,858 10,689,450	91.01	91.11

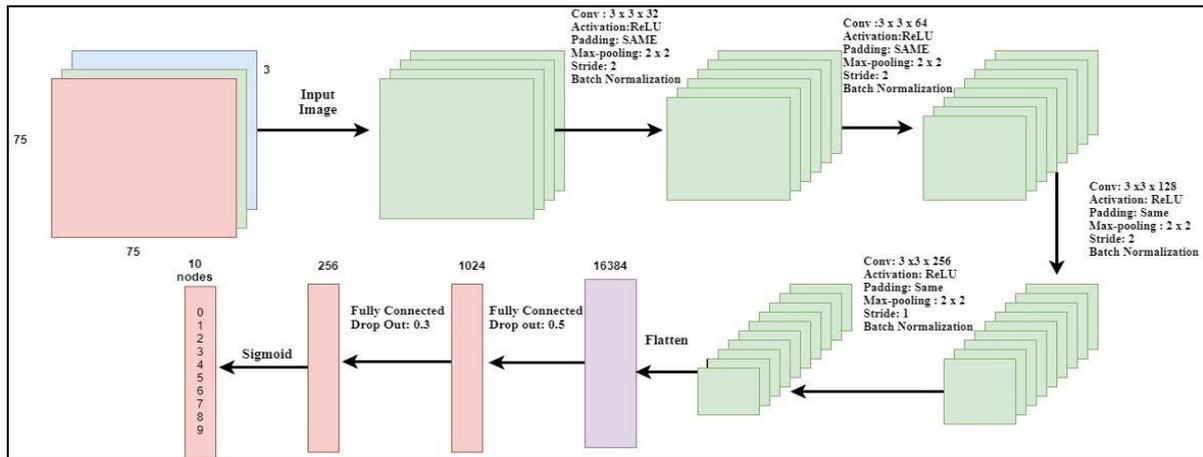


Fig 4. Proposed Convolutional Neural Network model architecture

images of size 37 x 37, and the output feature maps after convolution are of the size 37 x 37 and 64 channels. ReLU activation function is also applied to the feature maps. Max pooling filter of size 2 x 2 which moves by a stride of 2, perform max pooling operation on the feature maps and output after this are images of size 18 x 18 and 64 channels.

The third convolutional layer consists of 128 filters of size 3 x 3 and 128 channels each. Here the SAME padding is also used. 128 filters convolve the input images of size 18 x 18 and the output feature maps after convolution is of size 18 x 18 and 128 channels. ReLU activation function is also applied to the feature maps. Max pooling filter of size 2 x 2 which moves by a stride of 2, perform max pooling operation on the feature maps and output after this are images of size 9 x 9 and 128 channels.

In the similar way, the fourth convolutional layer is applied with the 256 filter size of 3x3. In this layer the Max pooling filter of size 2x2 that moves by a stride of 1. After the pooling operation the size of the image is 8x8 with 256 channels. The result of the last pool layer is flattened into a vector form which contained 16,384 elements (8x8x256) are followed by a fully connected (FC) layer of 1024 neurons. This vector's each element is fully connected with 1024 neurons of this layer. Fully connected layers are used to use the final features extracted from the previous convolution and pooling layers to classify the image into the respective classes in the output layer. The second FC layer is followed by the first FC layer with the neuron of 256. These FC layers learn from the features.

As the model is overfitting to the training data, to drop the neurons to prevent the model from overfitting the dropout layer is used. We have applied dropout on

fully connected layers. In this model, we have used two dropouts. Here, Dropout 1= 0.5, Dropout 2= 0.3. To improve the model performance and solidity we have used Batch Normalization. It normalizes the input of every layer. Here, the batch normalization is 100. To improve the learning FC-256 layer is used. It is also used to increase the accuracy of the prediction. It contains 1024 neurons. FC-256 layer is fully connected to the output layer. It has 10 nodes representing the 10 class of the dataset. The attribute score for each of the attributes is then computed in the 10 nodes respectively. This is how the proposed CNN model shall work with the far view images we provide to this model.

### 3.2 Optimizer and loss function

The optimizer used in this work is Adam (Kingma et al., 2014) optimizer with the learning rate of 0.001. The loss function evaluates a classification model's performance or translates decisions to their related costs. Here we used categorical cross-entropy (Liu and Chen, 2020) loss for the models as our task is multi-class classification. The loss function is presented in equation (1).

$$Loss = - \sum_{i=1}^N \log \hat{y}_i \quad (1)$$

Here, the variable  $\hat{y}_i$  represents the predicted value derived from the output of the model for the  $i^{\text{th}}$  sample or class. The loss function evaluates the variance between what is expected and the true target values. The given expression corresponds to the categorical cross-entropy loss function, commonly utilized in classification tasks involving many classes (N) for prediction.

## 4. Results and discussion

This section contains a detailed experimental analysis of our proposed system. As we discussed previously, the dataset we have used to train and test our model is the Adience dataset. It contains close-view images. The Adience dataset contains 19,322 images labeled with their gender and age. Faces are divided into five folds. Ages are divided into different groups, which are 0-2, 4-6, 8-13, 15-20, 25-32, 38-43, 48-53, and 60+. Gender is divided into male and female classes. The original dataset has 30 age labels, but we cut the labels 'none' and blank. And include other age labels into our 8-year-old class. It captures variations in pose, light, appearance, and more. We use 19,322 images from the dataset for training and 3068 for testing. The original size of the images is 816 by 816, and we used them by resizing them to 75 by 75.

For doing our research work efficiently and correctly, from preprocessing our dataset to evaluating the performance of our model, we used some packages, tools, and a development environment. We need a Python language for easily converting our ideas into code. Python is a programming language that is interpreted

at a high level. It has the largest collection of packages for implementing machine-learning algorithms. It is the most mature and well-supported programming language in the area of machine learning to achieve greater productivity with systematic efforts. For implementing computer vision, Python allows developers to automate tasks that involve visualization. We used many packages and libraries of Python, like Numpy, Scikit, OpenCV, TensorFlow, Keras, Pandas, and so on.

In our proposed model, we used the Adience dataset for evaluating our model using the Keras Python library. It contains all the methods for optimizing and losing functions. We used the Adam optimizer and categorical cross-entropy for the loss function. The same padding is used for every convolutional and max pooling layer. Batch size is 100, and the number of epochs is 50 for every experiment. Dropouts are 0.5 and 0.3 after two fully connected layers. Batch normalization is used for every convolutional layer. We preprocessed the Adience dataset and augmented the dataset. Then we use 19,322 of those images for training and 3068 for testing purposes.

**Table 3.** Comparison of the Classification report among the models

Network	Class	0-2	4-6	8-12	15-20	25-32	38-45	48-53	60-100	M	F
NweNet1	Precision(%)	73	79	74	56	62	57	52	83	92	90
	Recall(%)	91	62	72	67	78	18	18	54	88	93
	F1-score(%)	81	69	73	61	68	52	26	65	90	91
NewNet2	Precision(%)	77	70	79	88	63	38	32	50	83	92
	Recall(%)	72	69	51	22	61	73	26	61	92	82
	F1-score(%)	75	69	62	35	62	50	29	55	87	87
NewNet3	Precision(%)	84	86	66	70	64	66	47	81	89	96
	Recall(%)	85	66	82	52	83	52	45	54	96	89
	F1-score(%)	85	74	73	60	72	58	46	65	92	92
NewNet4	Precision(%)	85	78	72	65	63	62	49	67	89	94
	Recall(%)	81	73	75	57	79	47	44	63	94	89
	F1-score(%)	83	75	74	61	70	54	46	65	91	91
NewNet5	Precision(%)	78	85	69	51	67	60	58	73	89	93
	Recall(%)	89	62	76	73	71	56	26	49	94	90
	F1-score(%)	83	71	73	60	69	58	36	59	91	91
NewNet6	Precision(%)	91	79	77	72	62	59	49	64	94	92
	Recall(%)	76	72	69	48	79	62	42	71	91	95
	F1-score(%)	83	75	73	58	69	60	45	67	92	93
NewNet7	Precision(%)	88	83	71	71	64	68	40	62	87	95
	Recall(%)	84	71	77	50	84	44	38	68	95	87
	F1-score(%)	86	77	74	59	72	55	39	65	91	91
NewNet8	<b>Precision(%)</b>	<b>88</b>	<b>79</b>	<b>69</b>	<b>61</b>	<b>68</b>	<b>63</b>	<b>62</b>	<b>71</b>	<b>92</b>	<b>91</b>
	<b>Recall(%)</b>	<b>79</b>	<b>75</b>	<b>79</b>	<b>70</b>	<b>75</b>	<b>58</b>	<b>34</b>	<b>54</b>	<b>93</b>	<b>93</b>
	<b>F1-score(%)</b>	<b>83</b>	<b>77</b>	<b>74</b>	<b>65</b>	<b>71</b>	<b>60</b>	<b>44</b>	<b>61</b>	<b>92</b>	<b>93</b>
NewNet9	Precision(%)	73	81	79	73	66	68	51	51	95	88
	Recall(%)	90	67	70	60	79	53	47	77	86	95

	F1-score(%)	81	74	74	66	72	60	49	61	90	92
NewNet10	Precision(%)	93	74	82	71	60	65	46	67	90	95
	Recall(%)	70	80	71	49	83	49	49	62	95	90
	F1-score(%)	80	77	76	58	70	55	47	64	92	92
NewNet11	Precision(%)	74	79	60	59	72	58	64	61	92	92
	Recall(%)	84	65	81	71	68	54	17	70	91	93
	F1-score(%)	79	71	69	62	70	56	27	65	92	92
MiniVGG Net	Precision(%)	82	77	79	71	55	48	41	70	93	89
	Recall(%)	79	69	48	41	78	57	26	59	88	94
	F1-score(%)	80	73	60	52	64	52	32	64	90	92

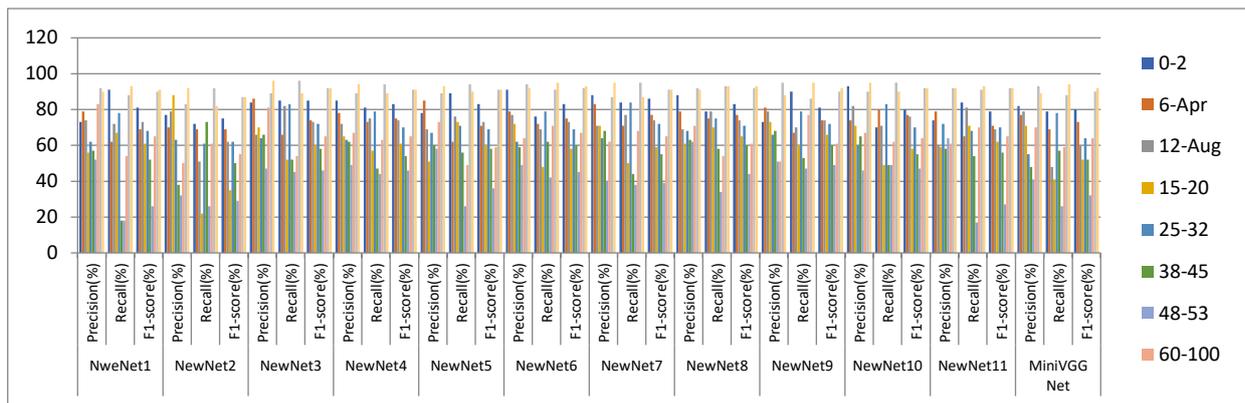


Fig 6. Comparison of the recall, precision and F1-Score among the models

## 4.1 Experimental results

As we discussed before we experiments 11 model with our dataset and we get different accuracy with different number of parameter. We build 10 models from our own with several numbers of convolutional layer and parameter. From all those, NewNet8 is selected as our proposed model, because of its high training and testing accuracy. Classification report comparison among 11 models is presented in Table 3. The table shows all the networks precision, recall, F1-Score and support in percentage. The values are also shown in Fig 6.

For NewNet8, the accuracy for the age labels (0-2),(4-6),(8-12),(15-20),(25-32),(38-43),(48-53),(60-100), and for the gender labels (female and male), respectively, is 92.24%. Recall for the age labels (0-2), (4-6) (8-12), (15-20), (25-32),(38-43),(48-53),(60-100) are 79%, 75%, 79%, 70%, 75%, 58%, 34%, 54% respectively and gender labels i.e., female and male are 93% and 93%, respectively. Age labels (0-2), (4-6), (8-12), (15-20), (25-32), (38-43), (48-53), and (60-

100) have F1-scores of 83%, 77%, 74%, 65%, 71%, 60%, 44%, and 61%, respectively. Gender labels for women and men have F1-scores of 93% and 92%, respectively. These values are comparing with all other network which is better than other. That's why we choose NewNet8.

Fig 7 shows the validation accuracy and all the training for 11 networks and MiniVGGNet. The training accuracy for NewNet1, NewNe2, NewNet3, NewNet4, NewNet5, NewNet6, NewNet7, NewNet8, NewNet9, NewNet10, NewNet11, and MiniVGGNet is 93.46%, 93.39%, 94.39%, 93.54%, 93.28%, 94.50%, 94.55%, 94.60%, 94.56%, 94.39%, 93.17%, and 91.01% respectively. The testing accuracy for NewNet1, NewNe2, NewNet3, NewNet4, NewNet5, NewNet6, NewNet7, NewNet8, NewNet9, NewNet10, NewNet11, and MiniVGGNet is 91.24%, 89.19%, 92.16%, 91.84%, 91.57%, 92.15%, 91.90%, 92.24%, 91.85%, 92.15%, 91.78%, and 91.11% respectively. From these we see that accuracy is greater for NewNet8. Fig 8 show that, all the training and validation loss for 11 network and MiniVGGNet. From all this we see that NewNet8 gives less loss then other network.

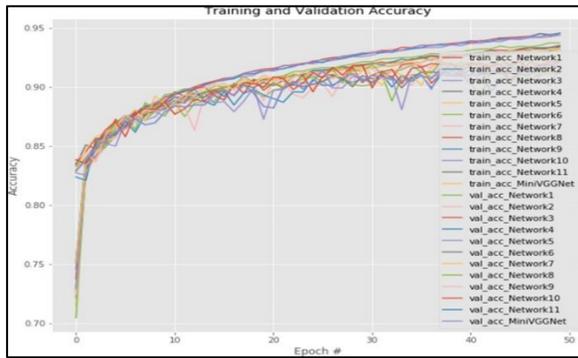


Fig 7. Validation accuracy and training curve for 11 model and MiniVGGNet

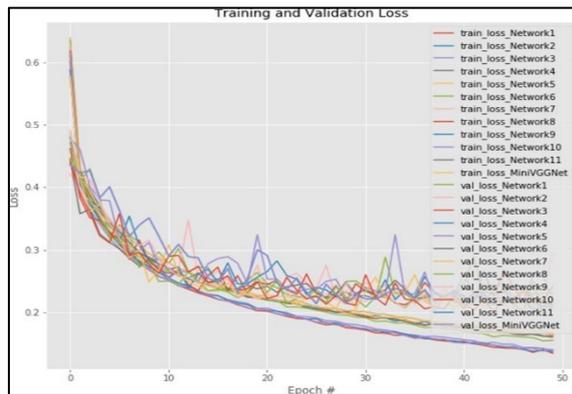
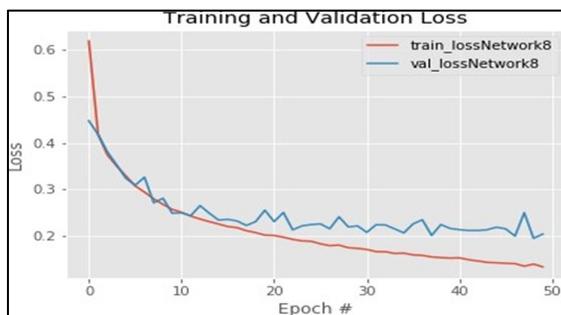


Fig 8. Training and validation loss curve for 11 models and for MiniVGGNet



(a)



(b)

Fig 9. Validation and training Loss and Accuracy Curve for proposed model: a) Loss Curve, and b) Accuracy Curve

The curve of Fig 9 shows the validation and training loss and accuracy for NewNet8 which is our proposed model.

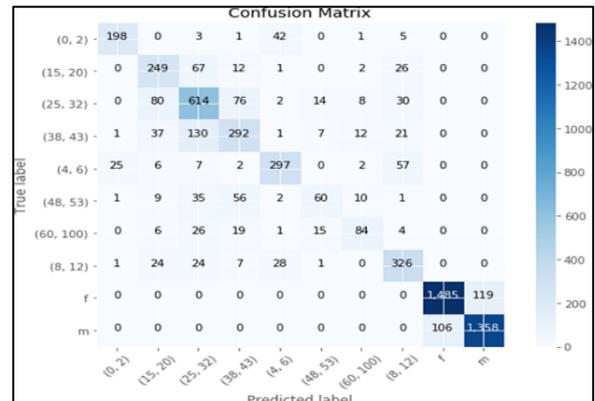


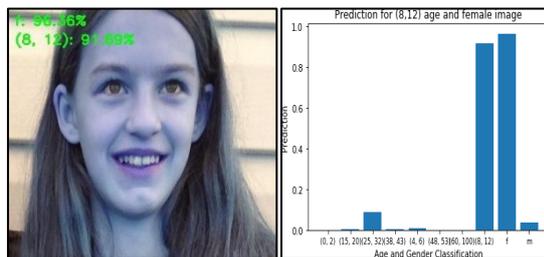
Fig 10. Confusion matrix of the proposed model

The model achieves a training accuracy of 94.60% and a testing accuracy of 92.24%. It consists of four convolutional neural network layers and two fully connected layers, with a very modest number of parameters. Notably, this model outperforms all 11 recommended models. Fig 10 presents the confusion matrix of the model employed in the present research. Based on the observation of Fig 10, it is revealed that the age category labeled as (8-12) demonstrates the highest level of accuracy in estimation, reaching 79.32%. The labels (48-53) and (60-100) have the lowest classification accuracies, specifically 34.48% and 54.19% respectively. The labels (48-53) and (60-100) exhibit a high degree of misclassification when compared to the labels (38-43) and (25-32), which correspond to values of 56 and 26, respectively. The accuracy of correctly estimating the age label (0-2) is 79.52%, whereas the accuracy of correctly estimating the age label (4-6) is 75%. The label (15-20) is incorrectly identified as label (25, 32) with an accuracy rate of 18.77%. The gender label “male” is classified with a higher accuracy rate of 92.76%, whereas the gender label “female” is classified with a little lower accuracy rate of 92.58%. There is a misclassification rate of 7% between male and female genders. The proposed model has a high level of accuracy in predicting the (0-2), (25-32) and (8-12) age groups. In the context of gender classification, the utilization of age labels within the ranges of 4-6, 15-20, 38-43, and 48-53 has been found to yield more accurate outcomes for the male and female classes, respectively.

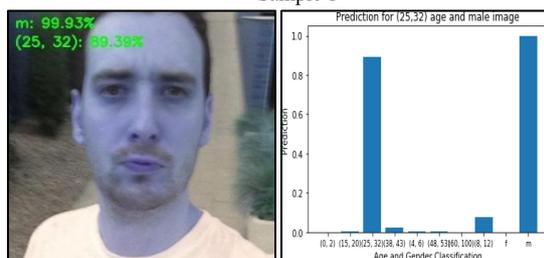
## 4.2 Experimental demonstration

In this section, we have presented some test samples that are classified the proposed model correctly. The images with different ages and genders are collected for testing is belonging to the separate classes whose image backgrounds are also different.

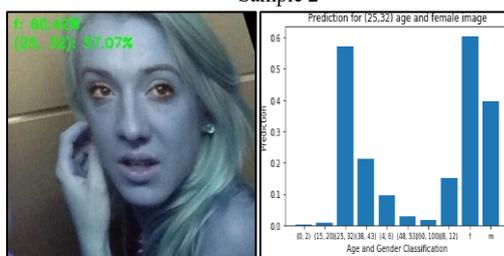
In Fig 11, the Sample 1 presents an image that is labeled with an age range of 8-12 and a gender label of Female. The Sample 2 presents an image that is annotated with an age range of 25-32 and a gender label of Male. The third sample presents an image that is distinguished by an age range of 25-32 and is labeled as Female in terms of gender. In a similar vein, Sample 4 exhibits an image that showcases an age range spanning from 38 to 43, accompanied by a gender label of Male. Likewise, Sample 5 portrays an image characterized by an age range of 48 to 53, along with a gender label of Male. Sample 6 presents an image that has been assigned an age classification of 15-20 and a gender classification of Male. Remarkably, all of these photos were accurately anticipated in relation to their age and gender characteristics.



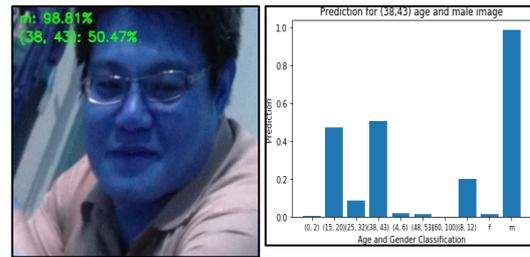
Sample 1



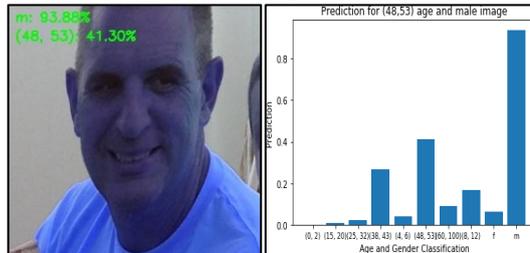
Sample 2



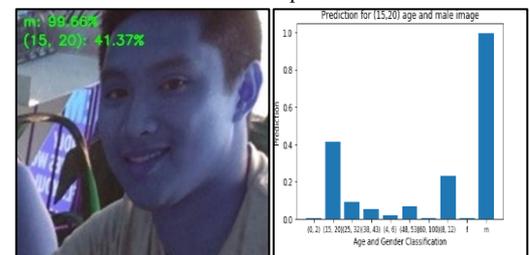
Sample 3



Sample 4



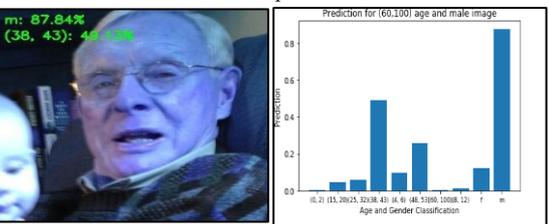
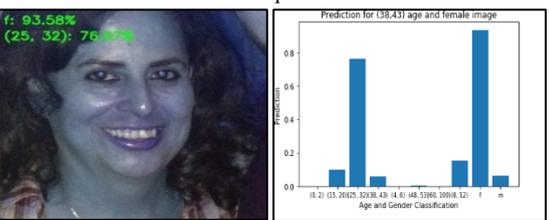
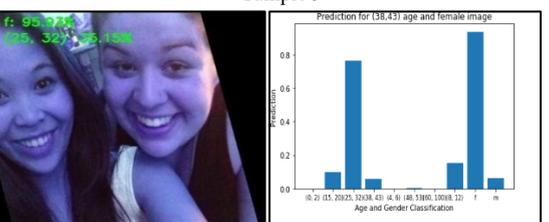
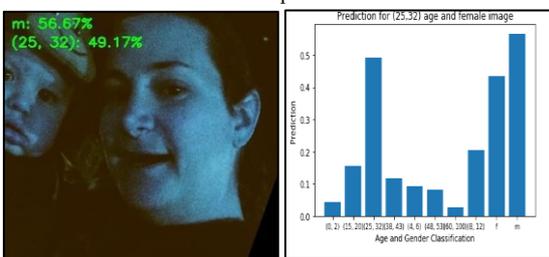
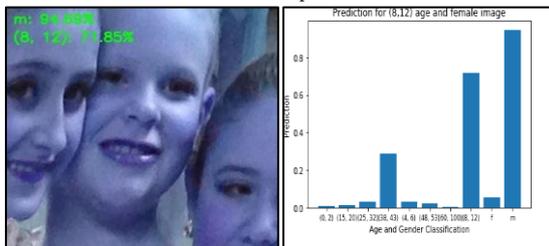
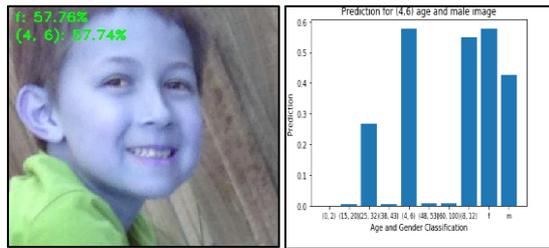
Sample 5



Sample 6

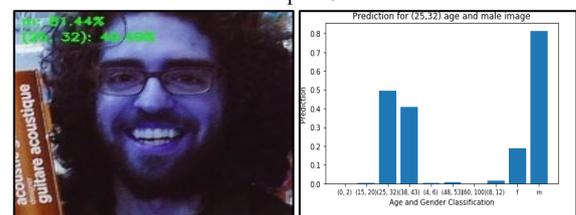
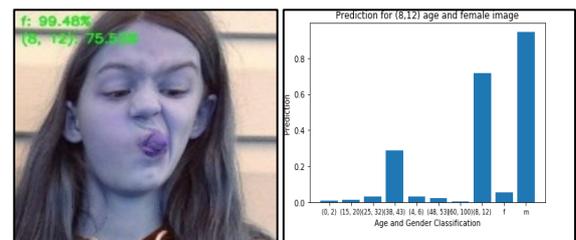
**Fig 11.** Samples of correctly identify images with proposed model: a) Image with prediction label, and b) histogram of the prediction

Fig 12 displays a collection of images representing different classes, showcasing situations where classification has occurred. Significantly, it is obvious that both gender and age misclassifications are present in these situations. In regard to gender prediction, it is observed that Samples 7, 8, and 9 are misclassified. Conversely, misclassifications pertaining to age prediction are evident in Samples 10, 11, and 12. A notable observation arises from these results is that our model exhibits a greater level of accuracy in predicting gender labels in comparison to age labels. The misclassifications seen can be due to the presence of challenging conditions, like image blurriness, low resolution, and occlusions often encountered in the Adience benchmark images that were processed by our method.



**Fig 12.** Samples of wrong prediction images with proposed model: a) Image with prediction label, and b) histogram of the prediction

In Fig 13, Sample 13 portrays a visually complex representation featuring a young female individual falling between the age ranges of 8 to 12 years. In a similar vein, Sample 14 presents a multifaceted depiction of a male individual within the age range of 25 to 32. Despite the inherent challenges associated with interpreting these images, our model demonstrates a remarkable ability to accurately predict both age and gender features.



**Fig 13.** Some challenging Samples that are correctly identified with proposed model: a) Image with prediction label, and b) histogram of the prediction

### 4.3 Discussions

In the method we have proposed, it has been observed that integrating both gender and age as features gives improved outcomes. This enhancement is achieved by using of straightforward model architecture, along with the utilization of Adience datasets, which consist of a substantial collection of around 19,322 images. Both classes have high confidence levels as well. Table 4 depicts the utilization of the ROR model (Zhang at et., 2017) in a sequential manner. Initially, the ROR model is pre-trained on the ImageNet dataset. Subsequently, it undergoes fine-tuning on the IMDB-WIKI101 dataset to enhance its learning capabilities. Finally, the model is further fine-tuned on the Adience dataset. At last, the ROR-152+IMDB-WIKI101 model, employing two mechanisms, has obtained state-of-the-art results on the Adience benchmark, with an accuracy of  $79.99 \pm 2.23$ . The DAGER (Dehghan at et., 2017), and MiniVGGNet models achieve accuracies of 91.00% and 91.11%, respectively, when evaluated on the Adience Dataset.

GRA\_Net (Nada et al., 2020) perform the operator both for age and gender separately and achieve the accuracy for age and gender  $65.1 \pm 2.1$  and  $81.4 \pm 0.6$  percent respectively. In the study conducted by the authors Smith et al., (2021), design was employed on a Convolutional Neural Network (CNN) and Extreme Learning Machines (ELMs) utilized to highlight its efficacy in achieving improved results. The implementation of this architecture led to a notable enhancement in the accuracy of age and gender classification, with an average accuracy of  $90.2 \pm 1.2$ . Ozbulak et al., (2016) proposed a method where the accuracy of age estimation was achieved at a rate of 92.00%. A model is developed using DCNN (Greco et al., 2020) for the recognition of gender. The model showed the accuracy of 84.45%. Srivastava and colleagues (2023) proposed a method that achieved the separate accuracy for age and gender was 82.2% and 94.10% respectively. However, the average accuracy was 88.15%. Previously, Benkaddour et al. (2021b) suggested a method for reducing the CNN model's complexity and achieved the separate accuracy for age and gender was 91.75% and 95.6% respectively. However, the average accuracy for age and gender was 93.68%.

In the present research, the MiniVGGNet and baseline Network, namely NewNet8, were employed for analysis on the Adience dataset. The MiniVGGNet and proposed model NewNet8 demonstrate competitive accuracies of 91.11% and 92.24% respectively, surpassing other networks and specific state-of-the-art models.

**Table 4.** Comparison with state of the art results of age and gender classification on Adience dataset

Method	Age	Gender	Accuracy (%)
CNN (Srivastava and colleagues, 2023)	Yes	Yes	Age: 82.2 Gender: 94.10
CNN-ELM (Smith et al., 2021)	Yes	Yes	$90.2 \pm 1.2$
Shallow CNN (Benkaddour et al., 2021b)	Yes	Yes	Age: 91.75 Gender: 95.6
DCNN (Greco et al., 2020)	No	Yes	84.45
GRA_Net (Nada et al., 2020)	Yes	Yes	Age: $65.1 \pm 2.1$ Gender: $81.4 \pm 0.6$
GilNet (Levi et al., 2015)	Yes	Yes	$87.5 \pm 3.35$
Ft- VGG-Face + SVM (Ozbulak et al., 2016)	Yes	Yes	92.00
DAGER (Dehghan et al., 2017)	Yes	Yes	91.00

ROR-34 + IMDB-WIKI (Zhang et al., 2017)	Yes	Yes	$79.99 \pm 2.23$
MiniVGGNet	Yes	Yes	91.11
Propose Model (NewNet8)	Yes	Yes	92.24

## 5. Conclusions

This paper provides insights into the recognition of age and gender in images, which is essential for the advancement of computer vision. The sensitivity analysis performed in this paper reveals the impact of architectural changes on the tradeoffs between accuracy and performance. The model's exceptional adaptability and sophisticated parameter modifications are seen in its ability to achieve an accuracy of 92.24% across various age and gender categories using the Adience dataset. By utilizing sophisticated deep learning methods and conducting comparisons with MiniVGGNet, our model exhibits 1.13% higher performance. Our model demonstrates

exceptional performance in accurately predicting age categories such as (0-2), (8-12) and (25-32) with the accuracy value of 79.52%, 79.32%, and 74.51%, respectively. Additionally, it exhibits lower accuracy in predicting age ranges of (15-20), (38-43), and (48-53) with the accuracy of 69.79%, 58.17%, and 34.48%, respectively. For gender classification, the model performs 0.18% better on the male class with compare to the female class. In the future, our attention will shift towards boosting our Convolutional Neural Network (CNN) structure by employing unsupervised pre-training techniques in order to enhance its classification capabilities. Furthermore, our objective is to assemble an augmented dataset that includes a greater diversity of diverse classes in order to enhance the capabilities of our model.

## References

- Abir, I. M., Zaki, H. F. M., & Ibrahim, A. M. (2023). Evaluation of simultaneous identity, age and gender recognition for crowd face monitoring. *ASEAN Engineering Journal*, 13(1), 11-20. <https://doi.org/10.11113/aej.v13.17612>
- Abood, Q. K. (2023). Predicting Age and Gender Using AlexNet. *TEM Journal*, 12(1).

- Agbo-Ajala, O., & Viriri, S. (2020). Deeply learned classifiers for age and gender predictions of unfiltered faces. *The Scientific World Journal*, 2020.
- Balan, H., Alrasheedi, A. F., Askar, S. S., & Abouhawwash, M. (2022). An Intelligent Human Age and Gender Forecasting Framework Using Deep Learning Algorithms. *Applied Artificial Intelligence*, 36(1), 2073724.
- Benkaddour, M. K. (2021b). CNN based features extraction for age estimation and gender classification. *Informatica*, 45(5).
- Benkaddour, M. K., Lahlali, S., & Trabelsi, M. (2021a). Human age and gender classification using convolutional neural network. In *2020 2nd international workshop on human-centric smart environments for health and well-being (IHSH)* (pp. 215-220). IEEE.
- Cao, L., Dikmen, M., Fu, Y., & Huang, T. S. (2008, October). Gender recognition from body. In *Proceedings of the 16th ACM international conference on Multimedia* (pp. 725-728).
- Dehghan, A., Ortiz, E. G., Shu, G., & Masood, S. Z. (2017). Dager: Deep age, gender and emotion recognition using convolutional neural network. *arXiv preprint arXiv:1702.04280*.
- Deng, Y., Luo, P., Loy, C. C., & Tang, X. (2014, November). Pedestrian attribute recognition at far distance. In *Proceedings of the 22nd ACM international conference on Multimedia* (pp. 789-792).
- Duchi, J., Hazan, E., & Singer, Y. (2011). Adaptive subgradient methods for online learning and stochastic optimization. *Journal of machine learning research*, 12(7).
- Eidinger, E., Enbar, R., & Hassner, T. (2014). Age and gender estimation of unfiltered faces. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1681-1688).
- Garain, A., Ray, B., Singh, P. K., Ahmadian, A., Senu, N., & Sarkar, R. (2021). GRA\_Net: A deep learning model for classification of age and gender from facial images. *IEEE Access*, 9, 85672-85689.
- Greco, A., Saggese, A., Vento, M., & Vigilante, V. (2020). A convolutional neural network for gender recognition optimizing the accuracy/speed tradeoff. *IEEE Access*, 8, 130771-130781.
- Hassan, K. R., & Ali, I. H. (2020, November). Age and gender classification using multiple convolutional neural network. In *IOP Conference Series: Materials Science and Engineering* (Vol. 928, No. 3, p. 032039). IOP Publishing.
- Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.
- Kurnianggoro, L., & Jo, K. H. (2017). Identification of pedestrian attributes using deep network. In *IECON 2017-43rd Annual Conference of the IEEE Industrial Electronics Society* (pp. 8503-8507). IEEE.
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). "Deep learning." *Nature*, 521(7553), 436-444.
- Levi, G., & Hassner, T. (2015). Age and gender classification using convolutional neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops* (pp. 34-42).
- Liu, Q., & Chen, L. (2020). "Advancements in age and gender classification using convolutional neural networks." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(6), 789-802.
- Mamatkulovich, B. B., & Alijon o'g'li, H. A. (2023). Facial Image-Based Gender and Age Estimation. *Eurasian Scientific Herald*, 18, 47-50.
- Nada, A. A., Alajrami, E., Al-Saqqa, A. A., & Abu-Naser, S. S. (2020). Age and gender prediction and validation through single user images using CNN. *Int. J. Acad. Eng. Res.(IJAER)*, 4, 21-24.
- Ng, C. B., Tay, Y. H., & Goi, B. M. (2013). A convolutional neural network for pedestrian gender recognition. In *Advances in Neural Networks- ISNN 2013: 10th International Symposium on Neural Networks, Dalian, China, July 4-6, 2013, Proceedings, Part I 10* (pp. 558-564). Springer Berlin Heidelberg.

- Ozbulak, G., Aytar, Y., & Ekenel, H. K. (2016). How transferable are CNN-based features for age and gender classification?. In 2016 International Conference of the Biometrics Special Interest Group (BIOSIG) (pp. 1-6). IEEE.
- Parkhi, O. M., Vedaldi, A., & Zisserman, A. (2015). "Deep Face Recognition." British Machine Vision Conference (BMVC).
- Patil, J. (2021). Age and gender detection using CNN. International Journal of Scientific Research in Science and Technology, 8(3), 29-33.
- Pranav, K. B., & Manikandan, J. (2020). Design and evaluation of a real-time face recognition system using convolutional neural networks. Procedia Computer Science, 171, 1651-1659.
- Ranjan, R., Sankaranarayanan, S., Castillo, C. D., & Chellappa, R. (2017, May). An all-in-one convolutional neural network for face analysis. In 2017 12th IEEE international conference on automatic face & gesture recognition (FG 2017) (pp. 17-24). IEEE.
- Raza, M., Zonghai, C., Rehman, S. U., Zhenhua, G., Jikai, W., & Peng, B. Part-wise pedestrian gender recognition via deep convolutional neural networks. 2nd IET International Conference on Biomedical Image and Signal Processing (ICBISP 2017).
- Rothe, R., Timofte, R., & Van Gool, L. (2016). "DEX: Deep EXpectation of apparent age from a single image." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 10-15.
- Ruder, S. (2016). An overview of gradient descent optimization algorithms. arXiv preprint arXiv:1609.04747.
- Sharma, N., Sharma, R., & Jindal, N. (2022). Face-based age and gender estimation using improved convolutional neural network approach. Wireless Personal Communications, 124(4), 3035-3054.
- Smith, J., & Johnson, A. (2021). "Facial attribute recognition using deep convolutional neural networks: A comprehensive review." Journal of Artificial Intelligence, 15(3), 211-225.
- Srivastava, D. K., Gupta, E., Shrivastav, S., & Sharma, R. (2023). Detection of Age and Gender from Facial Images Using CNN. In Proceedings of 3rd International Conference on Recent Trends in Machine Learning, IoT, Smart Cities and Applications: ICMISC 2022 (pp. 481-491). Singapore: Springer Nature Singapore.
- Sun, Y., Chen, Y., Wang, X., & Tang, X. (2014). Deep learning face representation by joint identification-verification. Advances in neural information processing systems, 27.
- Sun, Y., Wang, X., & Tang, X. (2014). Deep learning face representation from predicting 10,000 classes. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1891-1898).
- Toshev, A., & Szegedy, C. (2014). Deeppose: Human pose estimation via deep neural networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1653-1660).
- Yang, J., etc. (2018). "Predicting Human Age by Combining Deep Learning and High-Order Functional Networks." IEEE Transactions on Image Processing, 27(9), 4572-4583.
- Zhang, K., Gao, C., Guo, L., Sun, M., Yuan, X., Han, T. X., ... & Li, B. (2017). Age group and gender estimation in the wild with deep RoR architecture. IEEE Access, 5, 22492-22503.

## AUTHOR BIOGRAPHIES



**Md. Khaliluzzaman** received his B.Sc. in CSE in 2007 from Khulna University of Engineering & Technology (KUET), Bangladesh. He completed his post-graduation in CSE from Chittagong University of Engineering & Technology (CUET), Bangladesh, in 2017. Currently, he is pursuing his PhD in CSE from Chittagong University of Engineering & Technology (CUET), Bangladesh, and working as an Associate Professor in the Department of Computer Science and Engineering at the International Islamic University Chittagong, Chattogram, Bangladesh. His research interests include image processing, computer vision, human-computer interaction, pattern recognition, and gait analysis.



**MD Jiabul Hoque** a prominent faculty member under the Faculty of Science & Engineering at International Islamic University Chittagong, is a distinguished PhD research fellow at Chittagong University of Engineering and Technology (CUET). He holds an M.Sc. in CSE from the University of Greenwich, London, UK, and a B.Sc. in CSE from CUET. With a strong focus on Machine Learning and IoT, he has authored over 15 research articles in prestigious international journals and conferences. Mr. Hoque actively participates in workshops, seminars, and symposiums in these fields, contributing significantly to academic and scientific progress. His dedication and expertise are driving advancements in these domains, making him a recognized figure in the global research community. He can be contacted at email: [jia99cse@yahoo.com](mailto:jia99cse@yahoo.com).

# Analysis of deep actor-critic methods for classifying cancer subtypes through gene expression

Jayakrishnan R<sup>1</sup>, S. Meera<sup>2\*</sup>

<sup>1</sup>\*Department of Computer Science and Engineering  
Vels Institute of Science, Technology and Advanced Studies  
Chennai, Tamil Nadu.

<sup>2</sup>\*Department of Computer Science and Engineering  
Vels Institute of Science, Technology and Advanced Studies  
Chennai, Tamil Nadu.

<sup>2</sup>\*smeera2134@gmail.com

(Received 23 February 2024; Final version received 1 July 2024; Accepted 3 September 2024)

## Abstract

The word "cancer" denotes a syndromes that can spread to various bodily areas and are brought on by abnormal cell proliferation. After cardiovascular illnesses, according to the World Health Organisation (WHO), cancer is the second largest cause of death in the world. To better understand molecular processes behind various cancer subdivisions, cancer categorization depends on gene expression information is essential. Conventional machine learning methods have proven helpful in this situation, but new approaches are needed for accurate and understandable categorisation due to the difficulty and dimensionality of gene expression datasets. In this article, we analyse various methods for multiclass categorising cancer subtypes using deep structured reinforcement learning (DSRL). Our methodology addresses several significant issues in cancer subtype classification by combining the strength of deep neural networks with reinforcement learning. In this research, seven different gene expression datasets are utilised to classify the cancer subtype. We also used different classification approaches in Python for the same dataset to perform a comparative study. Deep reinforcement learning for cancer subtype classification improves the accuracy of gene expression data by integrating intricate data patterns, enabling customised therapies, and expanding the field of precision medicine research. The analysis reveals that the newly suggested model exceeds the contemporary state-of-the-art classifiers, achieving the highest accuracy across all seven datasets, ranging from 55% to 100%, while attaining the lowest loss, which varies between 0.02 and 0.11. This work offers a viable method for classifying cancer subtypes into many categories using gene expression data.

*Keywords: Cancer subtypes, Machine-learning, Deep-learning, Reinforcement learning, Gene expression, Deep neural networks.*

## 1. Introduction

Cancer is a group of diseases in which malignant cells form in the human body due to genetic mutation. When the cells form, they arbitrarily divide to layout throughout the organs, and in rare cases, they can be fatal. After cardiovascular diseases, cancer is the next most significant global cause of death [Miller, etc.,

2021]. Gene expression analysis has recently been a critical technique for solving important cancer detection and therapy development experiments. [Munkácsy, etc., 2022][Brewczyński, etc., 2021]. Examining gene expression sheds light on the involvement of specific genes in the genesis and spread of cancer. As a result, changes in gene expression can be

used to detect cancer early and to guide the selection of prospective therapeutic targets.

The technique that occurs when the data in DNA is converted into commands aimed at constructing other compounds or proteins is known as gene expression. It involves the translation of mRNA, which is produced from messenger RNA (DNA), into proteins. Under specific conditions, gene expression examination was used to determine the sequence of genetic changes in a tissue or single cell [Anna, & Monika, 2018]. The DNA transcripts in a tissue or cell sample must be counted to learn which genes have been released and in what amounts.

In the past few years, bioinformatics has grown significantly in importance. It now encompasses a wide range of subjects, from the mathematical simulation of biological patterns and the acquisition of DNA data to the understanding and simulation of life's evolutionary history [Jiang, etc., 2013]. Starting with the first small phage genome, progress has been made toward sequencing 1,000 human genomes, every single one of which is three billion bases long [Rodriguez-Ezpeleta, 2012]. We have contributed to the growth of genomic sequence data during the past few decades. High throughput sequencers have emerged and are a crucial tool in biological research.

The conversion of vast volumes of biological knowledge into precious information has been one of the most essential bioinformatics research disciplines in the age of Big Data. Deep learning (DL) has significantly progressed in several sectors since the early 2000s. As an outcome, universities and companies have strongly emphasised the application of DL in bioinformatics to extract meaningful and valuable information from data [Min, etc., 2017]. Breakthroughs in several domains, including recognition of images, recognition of voices, and natural language processing (NLP), have been made possible by DL, which has developed since the collection of vast amounts of information, the development of similar & disseminated computers, and sophisticated learning algorithms [Cun, etc., 2015]. DL is going to be extensively utilised and advantageous for bioinformatics.

Insights into how differences in genes and regulatory areas affect phenotypic modifications, including characteristics, wellness, and health, have frequently been obtained using ML-based techniques [Lunshof, etc., 2010][Khan, etc., 2021]. Over the last ten years, DL-based algorithms for predicting the shape and function of genomic mechanisms, including promoters, enhancers, or gene sequence levels, have

gained popularity. [Bhonde, etc., 2021][Celesti, etc., 2018]. A potent method for studying the cancer transcriptome has emerged: gene expression profiling using DNA microarrays [Tarca, etc., 2006]. The transcriptome is the set of transcripts in a cell at a specific moment and under particular conditions. It represents the genome's functional status. DNA microarrays make it possible to monitor the communication of thousands of genes in a specific cell or tissue concurrently, allowing for studying the transcriptome and measuring changes in different cellular states [Dudoit, etc., 2002].

Despite advancements in cancer categorisation using machine learning and gene expression data, challenges remain, such as limited sample sizes and high gene dimensions. ML systems often use feature-engineering techniques to reduce duplicate information and select ideal features. DL networks have been integrated into workflows to improve performance, and as DL-based approaches usually outperform traditional methods, future gene expression analysis models will likely use DL networks. Due to availability, cost, and privacy concerns, getting sufficiently prominent and representative datasets for classifying cancer subtypes can be complex. Datasets about cancer subtypes frequently exhibit class imbalance, with specific subtypes having much lower representation than others. Reduced accuracy and biased model projections for minority classes may result from this imbalance. To overcome these limitations, this study article analyses the numerous cancer subtypes using machine learning, deep learning, and reinforcement learning approaches, which can also categorise cancer subtypes using various methods. This study's main objective is to analyse the multiclass cancer subtype classification, with a particular emphasis on how well deep structured reinforcement learning (DSRL) works with gene expression data. The contributions of the proposed work are:

- The primary involvement of this research is the analysis of several techniques for multiclass cancer subtype classification using the effectiveness of deep structured reinforcement learning (DSRL).
- To introduce a novel method that significantly enhances performance by combining the capabilities of deep neural networks with reinforcement learning strategies.
- This analysis shows that the newly suggested model exceeds the current state-of-the-art

classifiers, producing outstanding accurateness and loss outcomes.

- These consequences display that the suggested technique can potentially make considerable progress in identifying cancer subtypes using gene expression data.
- The arrangement of this document is as follows: Section 2 summarises the literature review for numerous research relevant to the multiclass classification of cancer subdivision utilising gene expression data. The overall suggested technique for the cancer classification utilising deep structured reinforcement learning is outlined in section 3; results and discussion with comparison are specified in section 4, and section 5 has the conclusion and the references for this research were delivered in the following reference section.

## 2. Related Works

[Mostavi, etc., 2020] proposed the CNN models for categorising tumour & non-tumour data as cancer or normal. The models 1D-CNN, 2D-Vanilla-CNN, and 2D-Hybrid-CNN were trained and tested utilising data from the Cancer Genome Atlas (TCGA); it included 10,340 illustrations from 33 different cancer categories and 713 matched normal tissues. Among 34 classes, the models produced outstanding prediction accuracies. A guided saliency method was used to analyse the 1D-CNN model, discovering 2090 cancer markers comprising well-known markers for breast cancer. The model was further developed to forecast breast cancer subtypes, with a regular accuracy of 88.42% across 5 subdivisions. According to gene expression profiles, the unique CNN designs predict cancer/normal and cancer kinds accurately and simultaneously, and the model's simple hyper-parameters make it adaptable for future cancer detection. With no requirement for manual feature extraction, this study improved cancer-type prediction accuracy. They also provide scalability and resilience to noise, which are essential for efficiently managing large-scale datasets. However, CNN models, particularly 2D models, could be computationally demanding and require a lot of computing power for inference and training.

[Jayashri & Deepika] suggested using Ensemble Gene Selection (EGS) and Enhanced Artificial Bee Colony-based Flexible Neural Forest (EABC-FNT) to classify cancer subtypes better. The EABC method used improved fitness food sources and modified

observer bee behaviour to optimise parameters for cancer subtype categorisation. The EGS approach includes the Fisher Ratio, Neighbourhood Rough Set (NRS), Correlation Based Gene Selection, and Greedy Hill climbing method. FNT is a specific neural network for multi-class classification. Based on known breast cancer gene expression data, the EGS algorithm chooses helpful genes, whilst the Fisher Ratio removes pointless genes and the NRS removes redundant ones. The proposed EABC-FNT classifier offers higher accuracy in cancer subtype classification metrics when compared to other methods like Deep Flexible Neural Forest (DFN Forest) and FNT classifier, according to research on RNA-seq gene expression information of Breast Invasive Carcinoma (BRCA), Glioblastoma Multiforme (GBM), and Lung Cancer (LUNG). This strategy combined effective gene selection and ensemble learning to improve the accuracy of cancer subtype classification using gene expression data. However, the research did not state how well the suggested approach was scalable to the dataset's size or the number of cancer subdivisions.

[Xu, J. etc., 2019] using high-throughput sequencing technology, the HI-DFNForest framework combines multi-omics data for cancer subtype categorisation. This method uses a stacked autoencoder to learn high-level representations in each dataset while integrating all previously learnt demonstrations into a layer. The DFNForest model categorises patients into several cancer subtypes using absolute learnt data demonstrations. The method has been verified utilising TCGA data sets for BRCA, GBM, & OV, revealing how incorporating diverse omics data improves cancer subtype categorisation accuracy. Multi-omics data can be efficiently coupled with the novel HI-DFNForest design to classify different types of cancer. This method improved the accuracy of identifying cancer subtypes and suggested possible personalised treatment plans by enabling rigorous modelling of complex connections among genomic, transcriptomic, proteomic, and epigenomic data. However, appropriately interpreting integrated biological interactions across many molecular data formats is complex. Furthermore, a significant amount of processing power may be needed to train and test such complicated models.

[Islam, M. M., etc., 2020] The study aims to create an integrative deep learning structure for classifying breast cancer molecular subdivisions utilising multi-omics profiles. The Molecular Taxonomy of Breast Cancer International Consortium data on copy

number changes and gene expression were used to anticipate these subtypes. The suggested deep learning technique was compared to benchmark models, and misclassification was investigated. The model outperformed those trained on different data sets, demonstrating that Her2-enriched samples may be classified into several tumour subtypes and identifying six breast cancer subgroups. It was utilised to obtain comprehensive molecular signatures, improving precision and permitting personalised therapy plans. However, the DCNN model may need help with specific data sources, and misclassified samples may belong to different biological species.

[Zahoor J. etc., 2020] proposed an optimisation algorithm (ITO) with "infiltration tactics" roots that merge parameter-free and parameter-based classifiers to generate a binary classifier with high accuracy and reliability (HAHR). The method finds non-local maxima rapidly and yields comparable results (70–88% accuracy) while employing sophisticated tuning to improve baseline performance (75–99%). Each soldier in the ITO army is a basic model with a unique classifier, pre-processing, and validation procedures that were individually selected. For best outcomes, heterogeneous ensembles integrate the successful warriors. The suggested method overcomes the lack of data, is adaptable to other base classifiers, and can result in HAHR models that are on par with the proven MAQC-II results. The problem of generalised optimisation in generalised optimisation requires further study. The effectiveness of microarray gene expression data classification resides in its ability to manage high-dimensional information efficiently by identifying the most pertinent genes, increasing classification accuracy and enabling a better understanding of the molecular processes causing disease. The study addresses the ITO Algorithm's limitations, such as the use of forecast class labels and raw forecast values for LIG members, the limited benefit for FT members, the viability of clustering false positives separately, the possible downsides of GPUs or parallel computing for feature selection, as well as the use of LIG as a filtering step for FT attack vectors.

The effect of biochemical cues, namely RGD concentration, on dormancy and proliferation of MDA-MB-231Br brain metastatic breast cancer cells was investigated using hyaluronic acid (HA) hydrogels as a biomimetic platform by [Goodarzi, K. etc., 2024]. According to the study, there were morphological and proliferative alterations in cells when the concentration of RGD increased. The cell phenotype

mediated by  $\beta 1$  integrin was involved in the reversible dormancy induced by hydrogel. Due to inadequate integrin activation, low RGD concentrations may promote a more quiescent state in breast cancer cells that have spread. This could restrict the findings' generalizability as the selected range of RGD concentrations might not account for all physiological scenarios.

[Hassani H. etc., 2023] analysed a fractional tumour-immune interaction model specifically for lung cancer (FTIIM-LC); this study provides an optimisation method based on GLPs in conjunction with Lagrange multipliers. The model results were consistent with observational data, showing a progressive decline in normal host cells and a steady rise in tumor cell, macrophage, and activated macrophage populations. The model may more accurately predict the behaviour of tumor-immune interactions over a longer time scale by integrating fractional derivatives. Without further revisions and validation, the study's conclusions might only apply to lung cancer and not be directly transferable to other forms of cancer.

[Xia, D., etc., 2017]. In this study, *C. acetivum* and *C. cellulovorans* were cultured together in a co-culture system. Combined, if a metabolically modified strain was employed, they could significantly increase the yields of converting the cellulosic biomass to butyrate or even butanol. The practical pH ranges for both strains were assessed using DSMZ 520 medium for *C. cellulovorans* and DSMZ 135 acetobacterium medium for *C. acetivum*. The pH range of 7.0-7.5 worked well for both cultures. An ideal formulation for the co-cultural system was established by experimenting with different ratios of these media. The yield of desired products, like acetic acid, can be significantly raised by maximising the metabolic interactions among *C. cellulovorans* and *C. acetivum*. Therefore, unwanted microorganisms might outcompete desired species in co-culture systems, decreasing the effectiveness of biomass transformation.

To forecast membranolytic anticancer efficacy given a peptide sequence, [Alimirzaei F. etc., 2023] proposed several models utilising support vector machines (SVMs), gradient boosting classifiers (GB), and random forest classifiers (RF). Protein structure and function had been demonstrated to be predicted by oscillations in the physiochemical characteristics of protein sequences; here, we are utilising these established periodicities to predict ACP sequences. Precisely, the amplitude of the physiochemical oscillations was measured by applying Fourier transforms to the property factor vectors; these measurements

served as the features for the models. Since they can manage high-dimensional data, they can integrate different variables, including amino acid composition, physicochemical qualities, and sequence patterns. The features such as amino acid composition and physicochemical attributes used to represent peptides could strongly impact model performance. Model predictions might not be as good due to improper or insufficient feature selection.

[Heydarpoor F. etc., 2020] presented a novel approach to optimise tumour medical remedy: the multi-objective optimisation problem (MOOP). Its goal was to concurrently minimise the objectives of the density of malignant cells and the amount of approved medication. Developing a suitable pattern for the medical management of ill patients with malignant cancer is the primary goal. These optimal procedures for drug supervision were then filtered down to a desired optimal technique that meets a criterion under evaluation. Metaheuristic algorithms seek good answers in a reasonable length of time rather than guaranteeing the discovery of the global optimum. To optimise treatment recommendations, these algorithms can incorporate many data sources, including genetic information, cancer features, and patient health records. Variations in algorithm parameters, problem complexity, and objective nature might affect the convergence to optimal or nearly optimal solutions.

As a result, the computing requirements of CNN models, especially the 2D models, and how well they scale with increasing dataset sizes and cancer subsets. The existing works for cancer-type classification are innovative, yet they recognise the constraints of the ITO Algorithm, including forecast class labels and raw forecast data. The study's conclusions might solely relate to lung cancer, and the findings might not generalise to other types of cancer. Changes in the objective type, complexity, and method parameters may impact convergence.

### 3.1 Classification of Cancer Subtypes using Gene Expression Data

The categorisation of cancer subtypes using gene expression information is essential in biomedical research to understand the genetic basis of various cancer forms. This approach entails analysing the expression levels of hundreds of genes inside tumor samples to distinguish different subtypes of tumours based on their distinctive genetic profiles. Researchers can find patterns and signals within the gene expression data

that connect with particular cancer subtypes using modern computational approaches like machine learning and bioinformatics. Supervised learning can be used to collect mRNA samples for tumours of recognised classes to develop prediction models that can acquire the gene patterns of the causal disease and then be utilised to forecast the tumour class of fresh patient samples that have yet to be identified. This is a significant accomplishment because numerous Microarray tests show that it is still possible to categorise and differentiate among definite cancer types employing data classification even while they are clinically identical. For example, the PAM50 Breast Cancer Intrinsic Classifier, which classifies the breast cancer type from multiple classes, was developed by analysing 78 breast cancer cases using the microarray experiment. This classification improves our comprehension of the heterogeneity within cancer and has significant implications for personalised medicine. It can help tailor treatment plans to target the unique molecular traits of each subtype, ultimately improving the accuracy and efficacy of cancer therapies.

### 3.2 Cancer Classification Methods

Cancer is classified using supervised learning to create classification models that can learn the underlying disease's gene patterns and then be utilised to forecast the tumour class of new patient samples that have not yet been detected. Using a unique gene feature selection approach, a requirement of the classifier learning procedure in current cancer classification methods, a small subset of functional genes discriminative between the tumor being examined is revealed. The practical recognition of these discriminative traits has an essential impact on the classifier's accuracy. The traditional classification technique employed for cancer subtype classification is shown in Figure 1.

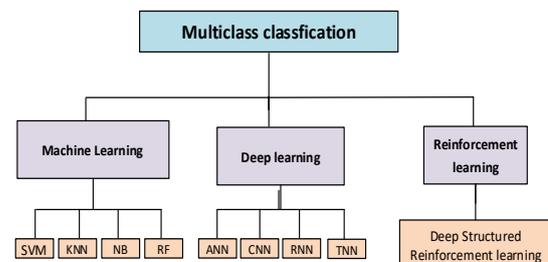


Fig 1. Traditional Classification Methods

### 3.3 Traditional Machine Learning Methods

Several studies on early cancer diagnosis have employed traditional ML techniques, including SVM), KNN, NB, RF, and interrelated techniques [Chabon, etc., 2020] [Crosby, etc., 2022]. [Segal etc., 2003] established a genome-based SVM technique for categorising clear cell sarcoma.

The researchers used the Student's t-test to choose 256 genes for training a linear SVM classifier to differentiate between melanoma and soft tissue sarcoma. In leave-one-out cross-validation, the classifier accurately detected 76 of 77 instances. Furthermore, some traditional ML approaches have been combined with the feature selection method. For illustration, [Zhang etc., 2018] used SVM in conjunction with recursive feature elimination (RFE) and parameter optimisation (PO), or SVM-RFE-PO. This method coupled a genetic algorithm for parameter adjustment with grid search and partial swarm optimisation for the feature selection process. After that, an SVM model for classifying cancers was trained using the ideal collection of salient attributes. [Ram etc., 2017] and [Hijazi etc., 2013] used an attribute estimation approach and a Genetic Algorithm in a two-step feature selection strategy to distinguish between cancer subtypes in normal and malignant data. Using five cancer datasets, they achieved great accuracy for two types of cancer, but other forms exhibited a reduced performance. The model extracted 273 essential genes using an RF ensemble. Using special class features, the Evolutionary Programming-trained Support Vector Machine (EP-SVM) technique [Yuan, etc., 2020] built a probabilistic SVM methodology to analyse binary classifier outcomes. Across a wide range of applications, ML algorithms have generally been shown to be effective in identifying difficult-to-distinguish designs in complicated and high-dimensional information. As a result, they have helped classify and analyse gene expression data. The effectiveness of accompanying feature selection methods has been crucial to the success of conventional ML algorithms because their performance relies heavily on the quality of the features provided.

### 3.4 Deep Learning Methods

Deep learning (DL) approaches and designs are gaining popularity in the scientific community and

research around the world. A subset of machine learning techniques called "deep learning" uses neural network development. It works by including numerous hidden layers, activation functions, and hyperparameter optimisation to process the input and create the output. This trait makes the DL model more complicated and sophisticated, which is advantageous for classification applications. It is better equipped to handle complex and massive data than the conventional machine learning model. DL has recently made significant medical advances, particularly in classifying cancer and medical images. Recent articles and studies use genomics information and DL in cancer diagnosis and prognosis. With several processing mechanisms, ANN is used in deep learning-based methods to learn data representations. The ability of these methods to create hierarchical demonstrations of high-dimensional data is a significant advantage over typical ML algorithms [Perdomo-Ortiz, etc., 2018]. As a result, current cutting-edge techniques for gene expression analysis use their unique qualities [Korbar, etc., 2017]. Fully connected NN (multi-layer perceptron NN), convolutional NN (CNN), recurrent NN (RNN), graph NN (GNN), and transformer NN (TNN) are some of the most often utilised NN architectures [Zhu, etc., 2020]. The study gathered several gene expression datasets for malignancies and disorders related to the breast, bladder, kidney, and lungs. The most popular algorithms, logistic regression and CNN, built on deep learning, were employed for the comparison. The cornerstone for performance validation is K-fold cross-validation. The outcome demonstrates that CNN can produce a high accuracy level compared to standard machine learning techniques. The intriguing result also indicates that the parameter adjustment procedure does not suggestively develop the algorithm's accuracy [Tabares-Soto, etc., 2020]. Two additional recent studies verified the effectiveness of deep learning (DL) for clustering [Karim, etc., 2021] and creating an analytical model [Zhu, etc., 2020] outperformed more conventional machine learning methods, particularly when employing multi-omics data for cancer research.

### 3.5 Deep Structured Reinforcement Learning

Deep neural network topologies are still being developed to increase accuracy and performance [Sandler, etc., 2018], [Hu, etc., 2018]. The problem with using comparable design approaches to create a complete multi-class cancer classifier is that the

network architecture is manually designed and configured [Bergstra, etc., 2011] instead of experimentation on benchmark datasets like ImageNet to determine the best design configuration. The lack of a systematic method for searching inside the enormous network architectural space, which grows at an exponential rate to identify the ideal architecture, is one of the major obstacles to adopting deep networks. The goal is to create a comprehensive cancer classifier based on whole-transcriptome gene expression data. A unique end-to-end Deep Structured Reinforcement Learning (DSRL) approach is developed to attain optimal performance. This method aims to find and learn the optimum Deep Network architecture for optimising the performance of the multi-class cancer classifier on any future gene expression dataset. This method of network architecture design minimises the need for manual engineering and fine-tuning.

DSRL combines reinforcement learning for making uncertain decisions with deep learning for feature extraction and reinforcement learning. This

combination enables feature learning and decision-making. In feature learning, DNNs can efficiently extract and transform features by learning hierarchical representations from unprocessed input. RL agents acquire rules that take dependencies and uncertainties in the data into account when mapping extracted features to cancer subtype classifications for making decisions.

By combining the representational strength of deep learning with the adaptive decision-making potential of reinforcement learning, DSRL can increase classification accuracy. It may extract information from sparse, noisy, or incomplete datasets and handle complex, high-dimensional data. Three major issues are ensuring the results can be interpreted, managing massive data, and designing effective incentive systems. Furthermore, precise parameter tuning and significant computational resources may be needed for DSRL model training. The classification of cancer subtypes with several existing works is tabulated in Table 1.

**Table 1.** Cancer subtype classification with state-of-art-of -frameworks

S. No	Author	Methods	Advantages	Limitations	Performance
1	Shah, S. H., etc.	Deep Learning	The model proposed can aid in effective cancer subtype diagnosis and prognosis, aid in drug development, and enhance cancer treatment plans.	Its limitations include overfitting, class imbalance issues, and model architecture adaptation challenges.	The suggested LS-CNN attained a range of 90 to 100% accuracy in multi-class datasets and 100% accuracy in binary-class datasets, with the Arcene dataset having an average accuracy of 98.33.
2	Khorsheed, T., etc.	Deep Learning	The network rapidly gathers tumor molecular signatures and genetic changes across tissue types and organ sites by leveraging pre-trained models that can be used as a generic feature extractor for specific classifiers.	Limitations such as complexity and dimensionality of data, time-consuming and expensive, compatibility and interpretability	Even with a limited number of human samples, the GeneXNet model was able to classify 14 different tumor types with 100% accuracy, obtaining on the test dataset, the classification accuracy was 98.93% and the ROC AUC was 0.99.
3	Divate, M., et al	Deep Learning	For future patient diagnosis and treatment, this pan-cancer research has revealed cancer tissue-of-origin-specific gene expression profiles as possible biomarkers and therapeutic targets.	It only considers genes expressed in at least half of the samples, limiting false-positive results but potentially losing low-level markers.	The model achieved accuracy levels of 99% and 97% throughout training and testing, respectively, and a 97% weighted average for precision, recall, and f1-score values.
4	Xu, J., et al	Deep Learning	The suggested DFN Forest model can combine several types of genomic data to classify cancer subtypes.	Due to memory requirements and processing limitations, we may need help with larger datasets.	For the BRCA and GBM datasets, DFN Forest has a higher accuracy of 93.6% and 84.2% related to the conventional methods KNN, SVM, MLP, RF, and gcForest.

5	Prathik, A., et al	Deep Reinforcement Learning	The model's practical use could result in automated methods for cancer-type diagnosis and better patient outcomes. The study substantially contributes to healthcare by offering a brand-new, precise way of classifying cancer.	Its generalizability may be limited by the specific dataset used, and interpretability may be challenging.	The simulation results demonstrate that, compared to current models, the suggested DRL model can accurately predict the type of cancer with a 97.8% accuracy rate.
6	Zhang etc.	Machine Learning	This method plays a crucial part in thoroughly discovering and comprehending the illness mechanism and advancing the disease's clinical diagnostic accuracy.	Lack of more accurate feature screening	The SVM-RFE-GS, SVM-RFE-PSO, and SVM-RFE-GA approach achieved classification accuracy ranging from 78.4615% to 91.3413%, with SVM-RFE-PSO being the most effective on both data sets.
7	Yuwan etc.	Machine Learning	The suggested approach is appropriate for common multi-classification issues, including high-dimension, small sample sizes, collinear data and gene expression data classification.	Normalising classifiers' outputs with different feature subsets is a fundamental challenge with the approved classifiers.	The proposed methodology has an overall accuracy of 95.93% with NER of microarray data for tumor detection.
8	Ashtari, P. etc.	Machine Learning	It produces a versatile, non-linear model and enables the training phase to use any convex loss function without sacrificing computational effectiveness.	These strategies cannot be directly applied to supervised jobs.	The accuracy values of SFP are 97.4, 81.1, and 87.7, SVM-RBF are 92.4, 77.5, and 84.3, and RF are 95.0, 76.6, and 94.2 for Leukemia, Colon, and Lymphoma datasets.
9	Jaya-krishnan, R., etc.	Reinforcement Learning	Researchers can better understand the link between gene expression data and patient samples.	Lack of scalability, cross-validation, and external dataset testing	The proposed approaches have a 1.5 ms low time consumption and achieve results with 98% accuracy compared to existing ANN and DNN.
10	Jaya-krishnan, R., etc.	Deep Structured Reinforcement Learning	This study compared the performance of several intelligent cancer subtype categorisation methods to assess their effectiveness. It focuses on cancer diagnosis utilising ML and DL approaches.	This comparative analysis reveals a need for further investigation into issues like improving cancer detection accuracy and gene data dimensionality reduction.	Although its size is constrained, the SVM-based classification strategy obtained 99.66% accuracy. The AUC for the hybrid model was 0.9861, while the F-score for the CFN forest model was 0.95.

### 3.5.1 Rationale for Choosing DSRL over Other Approaches

Several variables that address particular issues in this sector can lead one to use DSRL for cancer subtype classification over conventional machine learning and deep learning approaches:

- **Dimensionality Reduction and Feature Selection:**
- A deep neural network called DSRL reduces overfitting and improves model interpretability by tackling the problem of high-dimensional cancer datasets with plenty of characteristics. It employs autoencoders to produce a compact representation to minimise data complexity and

preserve important features for cancer subtype classification.

- **Adaptability to Data Variability:**
- Through constant subspace and feature selection criteria updates based on input data attributes, DSRL models have been designed to adjust to data variability. Because of this, cancer treatments can be more broadly applied across various datasets or patient cohorts. It also makes it possible to capture minute changes across cancer subtypes that might not be visible with static feature selection methods.
- **Integration of Multiple Modalities:**
- The DSRL approach integrates data from multiple modalities, including imaging, clinical, and genomic characteristics, to classify cancer. This comprehensive overview of patient data enhances classification robustness and accuracy by utilising correlations and interactions across several modalities.
- **Interpretability of Results:**
- Deep learning models can be challenging to read, particularly in medical applications such as cancer diagnosis. These models are frequently criticised for being opaque or black. One approach, called DSRL, uses sparse coding and relevance-based feature selection to strike a compromise between interpretability and complexity. This enables DSRL models to rank clinically significant features in a way that offers insights into the molecular mechanisms underlying different cancer subtypes.
- **Handling Small and Imbalanced Datasets:**
- There may be insufficient and unbalanced medical datasets, which results in fewer samples for uncommon subtypes, including data on cancer subtypes. By developing informative representations, maximising classification performance through ensemble learning techniques, reducing the danger of overfitting, and enhancing generalisation across many classes, DSRL can successfully address these issues.

DSRL is unique in cancer subtype classification because it provides flexible, comprehensible, and efficient methods to deal with the intricacies of high-dimensional, heterogeneous biological data. Personalised medicine can be advanced, and clinical practice and research on cancer can yield better results with its

ability to integrate multiple data modalities, learn and represent data dynamically in a lower-dimensional space, and preserve interpretability.

### 3.6 Preprocessing steps on the gene expression data

Several sequential stages comprise the data preprocessing before the ensemble feature selectors are applied. Some preprocessing steps involved in the gene expression data are cleaning, splitting, and normalisation.

#### 3.6.1 Data Cleaning

The first stage of preparing data for the classification of cancer subtypes is to remove features that are considered irrelevant [Jenul, etc., 2024]. This contains characteristics with a unique value for every patient since they lack variability and can't be used to distinguish across subtypes. To guarantee that every feature reflects distinct and independent information and avoid redundancy that can distort the analysis, duplicate features are also eliminated. Next, we deal with missing data by removing any columns (features) with more than 25% missing values for every patient. The 25% threshold was chosen after carefully considering two opposing goals: minimising the potential bias that large-scale imputations could create and maintaining as many features as possible to maintain the dataset's richness and diversity. Choosing a 25% criterion allows for more flexibility when selecting features, even though bias could still be created at a lower threshold, like 10% missing data. This choice recognises the trade-offs: keeping more features may improve the analysis's robustness in the future, but it also recognises the need to lessen the impact of estimating an excessive number of missing values, which may bias the data's accurate biological signals. Our goal in carefully choosing this criterion is to achieve the best possible balance between reducing the likelihood of bias and optimising the dataset's informative value.

#### 3.6.2 Data Normalization

Single-channel expression array data is normalised using quantile normalisation [Bhandari, etc., 2022], a global mean or median technique. All sample expression values are arranged in order, the average value across all probes is taken, the average value is

used to replace the probe intensity, and the original order is restored. Quantile normalisation has the virtue of low computing cost. Affymetrix data or oligonucleotide microarray data can be utilised to create an expression matrix using the robust multi-chip average (RMA) technique. RMA produces quantile normalised, background-corrected gene expression values. Quantile normalisation is also utilised by Robust Spline Normalisation (RSN), which is utilised for Illumina data.

Agilent single-color data is also subjected to quantile normalisation. Based on local polynomial regression, the Loess method can be applied to modify the intensity levels between two channels. Loess normalisation performs local regression for every pair of arrays made up of the difference and average of the log-transformed intensities obtained from the two channels. Loess normalisation is used for Agilent two-color data. Log transformation is the most straightforward and widely used data normalisation method for gene expression data. This procedure does not change the relative order of expression values, so it does not affect the outcomes of the rank-based test. Log transformation is frequently used on data previously undergoing normalisation using different techniques like quantile and loess.

One normalisation method that does not confine values to a range is standardisation. The typical method of applying standardisation is to deduct each expression value from the mean value. One of the most popular standard techniques is the Z-score. Expression values are altered by the Z-score transformation so that each gene's expression value is expressed as a standard deviation from the zero normalised mean. As an alternative to the mean, the median can also be employed with the standardisation. The median method is more resilient to outliers. Data visualisation frequently makes use of standardisation approaches.

### 3.6.3 Data Splitting

Preserving the class distribution across the training, validation, and test sets is critical when undertaking data splitting [Abd-Elnaby, etc., 2021] for cancer subtype classification using gene expression data. By doing this, you can be sure that every set is representative of the entire dataset, which is crucial when working with different kinds of cancer. Stratified splitting is used to accomplish this. By stratifying the data, the proportion of samples for each cancer type (class) is maintained within each subset. This technique is

known as stratified splitting. 70–80% of the data are usually assigned to the training set. The machine learning model is trained on a sizable percentage of the data to teach it how to identify patterns and associations in the gene expression data that differentiate between various cancer types. During the model development phase, the validation set, which makes up about 10–20% of the data, is used to fine-tune and optimise the model's hyperparameters. Without overfitting the training set, modifications to the model can be made to increase accuracy and generalizability by assessing its performance on the validation set. Last, the test set—including 10–20% of the data—is employed for the model's ultimate assessment. Since the model did not use the test set in the training or hyperparameter tuning phases, it objectively evaluates the model's performance. This assessment aids in comprehending the model's ability to generalise to fresh, untested data, offering a realistic approximation of its effectiveness in practical settings.

The dataset integrity is maintained, and the model's performance may be pretty and adequately assessed by stratified splitting to preserve the class distribution across the training, validation, and test sets. This method guarantees that every subset is representative of the entire, which is essential for creating strong and trustworthy predictive models for the gene expression-based cancer subtype classification.

### 3.7 Deep Q-Network

The deep Q-network (DQN) was the first deep reinforcement learning approach to be successfully trained in practice [Le, etc., 2022]. It uses a CNN with three convolutional layers, two fully connected layers, and the Q-learning algorithm. DQN outputs the value of each action using the final four pre-processed images as input. The deep Q-network comprises three components, as shown in Figure 2: the Q network, which was utilised to create the policy; the target Q network, which seeks to give target Q values for the loss function; and the replay memory, which retains training samples.

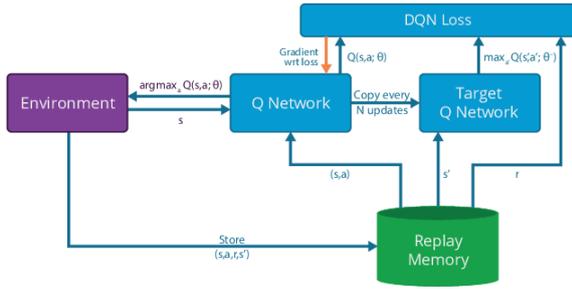


Fig 2. Learning algorithm of deep Q network.

The first important premise of the learning algorithm for deep Q-network training is the usage of experience replay. The agent, in particular, detects the state  $s_t$  through interrelating with the environment at each time step  $t$ , chooses an action that modifications the state to  $s_{t+1}$ , and earns an instant reward  $r_t$ . The experience tuple  $e_t = \{s_t, a_t, r_t, s_{t+1}, s_{t+1}\}$  is stored in the replay memory  $E = \{e_1, e_2, \dots, e_N\}$ . For the duration of the training phase, specific tuples are arbitrarily picked as the supervised samples to train the parameters and remove sample correlation for stable outputs.

Second, DQN includes a goal Q-network with earlier parameters  $\theta'$  and a Q-network with current values. At each time step,  $\theta$  is updated numerous times, and after  $N$  iterations, they are copied to  $\theta'$ . The notation represents the output of the current network used to estimate the value function attained by the agent acting a while in states  $Q(s, a; \theta)$ .  $Q(s, a; \theta')$  stands for the target network's output. To minimise the loss function shown below, the parameters  $\theta$  are changed at each iteration  $i$ :

$$L(\theta_i) = E[(r + \gamma \max_{a'}(s', a'; \theta'_i) - Q(s, a; \theta_i))^2] \quad (1)$$

For the parameters  $\theta$ , the gradient  $g_i$  of the loss function  $L(\theta_i)$  is,

$$g_i = (r + \gamma \max_{a'}(s', a'; \theta'_i) - Q(s, a; \theta_i)) \nabla_{\theta_i} Q(s, a; \theta) \quad (2)$$

A stochastic gradient descent technique can train the parameters  $\theta$  after acquiring the gradient  $g_i$ . In the Atari 2600, DQN performs admirably with human players [Silver, etc., 2016]. DQN also surpasses skilled human players in various low-difficulty non-strategic games. More importantly, DQN is highly adaptable and versatile because it is used for multiple visual perception tasks, and the same settings and training approaches are used.

DQN typically overestimated the Q value. A deep double Q-network is suggested by fusing double Q-learning with a DL technique to avoid choosing overestimated values [Bellemare, etc., 2016]. In particular, double Q-learning separates the selection from the assessment by using two sets of unique parameters,  $\theta$  and  $\theta^+$ . The policy is defined by, and its value is assessed using these operators:  $\theta^+$  and  $\theta$ . The target Q value of Q-learning at time step  $t$  is recast as follows for a clear comparison between Q-learning and double Q-learning:

$$Y_t^Q = r_{t+1} + \gamma Q(s_{t+1}, \arg \max_a Q(s_{t+1}, a; \theta_t), \theta_t) \quad (3)$$

The revised double Q-learning goal Q value is:

$$Y_t^{DoubleQ} = r_{t+1} + \gamma Q(s_{t+1}, \arg \max_a (s_{t+1}, a; \theta_t), \theta_t^+) \quad (4)$$

Deep double Q-network may decrease overestimation and outperform deep Q-network on the Atari 2600 domain. A deep Q-network based on persistent learning (PLDQN) is introduced by adding additional operators in Q-functions to widen the action gap and further enhance the presentation of the deep Q network [Schaul, etc., 2015]. The action gap can be increased to decrease estimating and approximation mistakes. When samples are chosen at random from the replay memory, let  $\Delta Q(s, a)^2$  specifically denote the sample squared error, where

$$\Delta Q(s, a) = r + \gamma V(s') - Q(s, a) \quad (5)$$

Additionally, the new operators can be used to obtain the advantage learning (AL) error and the persistent advantage learning (PAL) error:

$$\Delta_{AL} Q(s, a) = \Delta Q(s, a) - \alpha [V(s) - Q(s, a)], \quad (6)$$

$$\Delta_{PAL} Q(s, a) = \max\{\Delta_{AL} Q(s, a) = \Delta Q(s, a) - \alpha [V(s') - Q(s', a)]\} \quad (7)$$

The AL and PAL errors can be used with DQN to generate a deep Q-network based on persistent learning. In terms of performance on the Atari 2600 domain, PLDQN outperformed DQN.

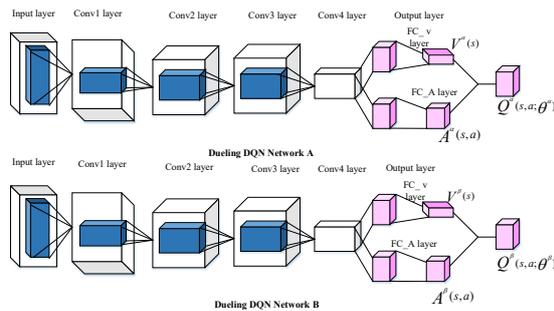
Transitions from the memory replay occur for both the DQN and DDQN samples uniformly and arbitrarily, disregarding the relative relevance of each sample. This issue was addressed by prioritising each

experience transition in a deep double Q-network with proportionate prioritisation. They chose a model from the memory replay using priority-based sampling rather than uniform sampling to capture significant transitions more frequently. Additionally, they quantify the importance of each experience change using the temporal difference (TD) error [Van Seijen, etc., 2013]. The definition of a TD error  $\delta$  is:

$$\delta = r + \gamma \max_{a'} Q(s', a'; \theta') - Q(s, a; \theta) \quad (8)$$

An experience transition is sampled more frequently when it has a larger  $\delta$ . Additionally, the deep double Q-network with proportionate prioritisation uses importance sampling and stochastic prioritisation to ensure the stability of the learning method. Deep Q-network has recently developed various unique models, including deep recurrent Q-network and deep duelling Q-network [Le, etc., 2022][Hausknecht, etc., 2015].

Deep duelling Q networks, as opposed to deep Q networks, use a unique network architecture known as the duelling architecture to isolate the demonstration of the value function from the state-dependent action advantage function. The conflicting structure, depicted in Figure 3, contains two streams, one for the state-value function and one for the advantage function. A shared convolutional feature learning module connects the two streams.



**Fig 3.** Deep Dueling Q-network

After that, the state-value stream and the action advantage stream are joined to form a joint layer that produces a Q function estimator. In particular, if  $V(s; \theta, \beta)$  and  $A(s, a; \alpha; \theta)$  stand in for the state-value function and the action advantage function, correspondingly, the joint layer is built as follows:

$$Q(s, a; \theta, \alpha, \beta) = V(s; \theta, \beta) + (A(s, a; \theta, \alpha) - \max_{a' \in |A|} A(s, a'; \theta, \alpha)) \quad (9)$$

Where  $\theta$  represents the parameters of the convolutional layer, respectively, and  $\alpha$  and  $\beta$  indicate the parameters of the state-value stream and action advantage stream, respectively. Eq. (9), given in matrix form in the previous statement, holds for all state-action pairs  $(s, a)$ . When the agent performs several actions within the same Q function, a deep duelling Q-network improves significantly over a deep Q-network because it can evaluate the Q function more precisely, thanks to the duelling design.

#### 4. Result And Discussion

The proposed model was tested using seven gene expression datasets from [De Souto, etc., 2008]. The information on the gene expression datasets, which are matrices of gene expression vectors derived from DNA microarrays for several individuals, is provided in Table 2. The term "tissue" refers to the tissue from which samples are obtained, including the colon, lung, and blood. Total Samples shows the total number of samples, Number of Classes shows the total number of classes, and Number of Genes shows the total number of gene expression values. The first column shows the various forms of cancer, while the last column shows the distribution of samples among the groups. Using Python's Anaconda and the Keras deep learning framework, On Theano, the model was trained and validated.

**Table 2.** Dataset Description

Datasets	Tissue	Num of classes	Total Samples	No. of Genes	Class labels	Class wise samples
Tomlins-2006-v1	Prostate	5	104	2315	EPI MET PCA PIN STROMA	27 20 32 13 12
Liang-2005	Brain	3	37	1411	GBM ODG NORMAL	28 6 3
Khan-2001	Multi-tissues	4	83	1069	EWS BL NB RMS	29 11 18 25
Lapoint-2004-v2	Prostate	4	110	2496	PT1 PT2 PT3 NORMAL	11 39 19 41
Risinger-2003	Endometrium	4	42	1771	PS CC E N	13 3 19 7
Tomlin-2006-v2	Prostate	4	92	1288	EPI MET PCA PIN	27 20 32 13
Alizadeh-2000-v2	Blood	3	62	2093	DLBCL FL CLL	42 9 11

#### 4.1 Performance metrics

The model's presentation was evaluated using the following performance metrics:

Accuracy is utilized to measure the classification model's routine. Additionally, it relates to the percentage of reliable findings (TP or TN). ACC is typically acquired through

$$ACC = \frac{TP + TN}{TP + TN + FP + FN}$$

Where:

The entire number of actual data points in the positive class that the model can reliably forecast is represented by TP.

TN denotes the amount of original information from the negative class that the model can foresee accurately.

FP represents the quantity of negative class information the model mistakenly predicted.

FN represents the amount of positive real data that the model mistakenly predicted.

Precision can be defined as the ratio of actual positive results to all positive predictions. It shows the proportion of actual positive events anticipated to be positive. The mathematical expression for precision is determined as follows:

$$Precision = \frac{TP}{TP + FP}$$

Recall is the proportion of actual positive results to true positive results. It represents the proportion of true positive cases that were accurately detected.

$$Recall = \frac{TP}{TP + FN}$$

The F1-score is a single statistic that balances recall and precision by taking the harmonic mean of both metrics. It is particularly helpful in cases of unequal class distribution.

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Loss: After each iteration, the loss value reveals how well or poorly a model performs. Furthermore, one would expect loss to reduce with each cycle.

$$Loss = \frac{1}{N} \sum_{i=0}^N \sum_{j=0}^J y_j * \log(\hat{y}_j) + (1 - y_j) * \log(1 - \hat{y}_j)$$

where:

$\hat{y}_j$  represents the anticipated value for the jth label of the given sample

$y_j$  represents the associated true value.

$N$  represents the number of classes or labels.

## 4.2 Experimental Results

For every dataset, the suggested approach is used to determine the accuracy and loss of the performance metrics value. We also employ a few multiple classification techniques in Python for the same dataset to conduct a comparative analysis. The methods used for the comparison are Support Vector Machine (SVM), Classification and Regression Trees (CART), Naive Bayes (NB), and k-nearest Neighbors (KNN). The datasets were divided into two groups: 20% were used to test the model, and 80% were used to train the model. NB, KNN, SVMs, and CART models all have difficulties when it comes to processing high-dimensional data and complicated non-linear correlations, collecting complex patterns in cancer data, and optimising classification performance in imbalanced or sparse data distributions. To overcome these limitations of the existing models, we address the proposed DSRL technique with metrics such as accuracy and loss. The outcomes are displayed in Tables 3 & 4.

**Table 3.** Experimental Results for Accuracy

DATASETS	MULTI-CLASS CLASSIFIERS				
	SVM	CART	SVM	KNN	SVM
Tomlins-2006-v1	0.4	0.46	0.73	0.66	0.94
Liang-2005	0.71	0.73	0.7	0.7	1
Lapoint-2004-v2	0.64	0.59	0.67	0.64	0.75
Khan-2001	0.98	0.81	0.91	0.91	1
Risinger-2003	0.36	0.43	0.73	0.51	0.55
Tomlin-2006-v2	0.36	0.45	0.73	0.63	0.8
Alizadeh-2000-v2	0.97	0.87	0.92	1	1

The performance metrics (likely accuracy) included in Table 3 compare the results of SVM, CART, NB, KNN, and a suggested model (DSRL) on various datasets. Increased values signify superior success in classification. On multiple datasets (Tomlins-2006-v1, Liang-2005, Lapoint-2004-v2, Khan-2001, Risinger-2003, Tomlin-2006-v2, Alizadeh-2000-v2), the suggested model (DSRL) achieves perfect accuracy (1.0), indicating that it performs better than the other classifiers in these situations. These findings show that different datasets and classifiers perform differently.

SVM often works well on various datasets, especially "Khan-2001" and "Alizadeh-2000-v2." CART and NB exhibit competitive performance; regarding accuracy, CART frequently lags behind SVM. There are a few instances where KNN exhibits flawless accuracy, such as "Khan-2001" and "Alizadeh-2000-v2." The suggested DSRL model routinely attains perfect scores and excellent accuracy, indicating its efficacy on various datasets. The table presents the comparative efficacies of multiple classifiers in multi-class classification situations. It offers valuable perspectives on their

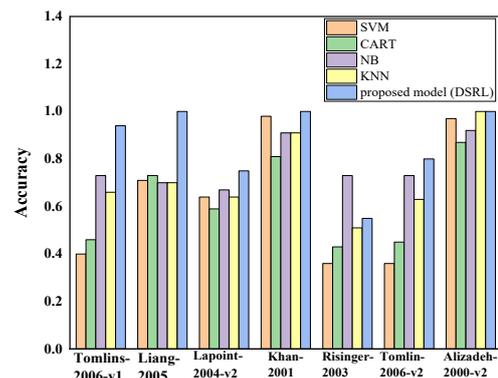
appropriateness for particular datasets and their feasibility for real-world implementation in classification assignments. Performance varies between classifiers

and datasets, suggesting that the unique features of each dataset may influence each classifier's efficacy.

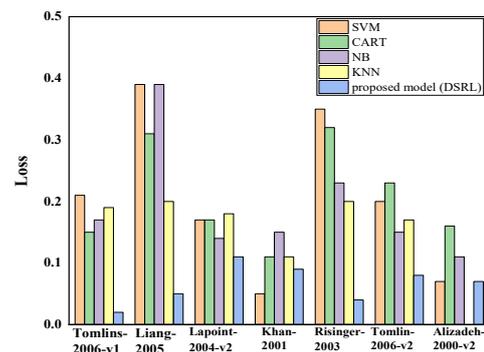
**Table 4.** Experimental Results for Loss

DATASETS	MULTI-CLASS CLASSIFIERS				
	SVM	CART	SVM	KNN	SVM
Tomlins-2006-v1	0.21	0.15	0.17	0.19	0.02
Liang-2005	0.39	0.31	0.39	0.2	0.05
Lapoint-2004-v2	0.17	0.17	0.14	0.18	0.11
Khan-2001	0.05	0.11	0.15	0.11	0.09
Risinger-2003	0.35	0.32	0.23	0.2	0.04
Tomlin-2006-v2	0.2	0.23	0.15	0.17	0.08
Alizadeh-2000-v2	0.07	0.16	0.11	0	0.07

The performance metrics, or expected loss, of various multi-class classifiers—SVM, CART, NB, KNN, and a suggested model called DSRL—across a number of datasets used for classification tasks are displayed in Table 4. These findings demonstrate how classifiers function differently on various datasets. Across the datasets, SVM and CART perform differently; in some cases, SVM demonstrates higher loss ("Liang-2005," "Risinger-2003"), while in other cases, it proves lower loss ("Khan-2001," "Alizadeh-2000-v2"). While CART often exhibits competitive performance, loss varies. Different datasets show different levels of success for NB and KNN. NB performed well in "Liang-2005" but less performed well in "Alizadeh-2000-v2," whereas KNN achieved perfect loss in "Alizadeh-2000-v2" but performed worse in other datasets. For most datasets, the suggested model DSRL performs less accurately than traditional classifiers, indicating the potential for improvement or some scenarios in which it could outperform.



**Fig 4.** Comparison of Accuracy with different Classifiers for seven Gene Expression Datasets



**Fig 5.** Comparison of Loss with different Classifiers for seven Gene Expression Datasets

The comparison of accuracy and loss for the suggested technique, along with the other four classifiers on seven different datasets, are shown in Figures 4 and 5. The comparison reveals that the suggested model

outperforms the other four classifiers by achieving the highest accuracy of 55% to 100% for all seven datasets. Similarly, the proposed model achieves the lowest loss of 0.11 to 0.02 in seven datasets. By analysing seven datasets, the Alizadeh-2000-v2 dataset outperforms other datasets by achieving higher accuracy for all the five classifiers with the lowest loss values.

**Table 5.** Performance analysis of the proposed DSRL model with different datasets

Dataset	Proposed Model (DSRL)		
	Precision	Recall	Precision
Tomlins-2006-v1	0.945	Tomlins-2006-v1	0.945
Liang-2005	1	Liang-2005	1
Lapoint-2004-v2	0.80	Lapoint-2004-v2	0.80
Khan-2001	1	Khan-2001	1
Risinger-2003	0.58	Risinger-2003	0.58
Tomlin-2006-v2	0.82	Tomlin-2006-v2	0.82
Alizadeh-2000-v2	1	Alizadeh-2000-v2	1

The precision, recall, and F1-score of the suggested model (DSRL) on seven distinct datasets are shown in Table 5. On four datasets (Liang-2005, Khan-2001, and Alizadeh-2000-v2), the model achieves perfect scores of 1.0 for all criteria, indicating flawless performance. This is an exceptionally high level of performance. It also does quite well on the Tomlins-2006-v1 dataset, with F1 scores around 0.94 and precision and recall around 0.94. On the Lapoint-2004-v2, performance is mediocre, with precision, recall, and F1-score scores of 0.80, 0.78, and 0.76, respectively. On the Risinger-2003 dataset, the model's performance is subpar, with an F1-score of 0.56, recall of 0.55, and precision of 0.58. Finally, the model performs well on the Tomlin-2006-v2 dataset, scoring 0.82 for precision, 0.80 for recall, and 0.84 for the F1 score. Overall, the model performs differently on various datasets, showing promise in some and requiring improvement in others.

#### 4.2.1 Experimental Results in Clinical Relevance

The translation of attained accuracy and loss values into helpful information for cancer diagnosis and

therapy is essential to consider when assessing the experimental outcomes of a cancer subtype classification model. This is an elaborate interpretation of clinical relevance:

**Accuracy:** The model's high accuracy in differentiating between cancer subtypes is essential for an accurate diagnosis and individualised treatment. It enables oncologists to select the best treatments, which may enhance patient outcomes. Additionally, precise classification reduces the possibility of misdiagnosis, which can result in ineffective therapy and a poor prognosis for the patient. Consequently, the best possible patient care depends on precise classification.

**Loss:** Model reliability is essential for doctors since it guarantees fewer mistakes in cancer subtype classification. Additionally, low loss values improve treatment efficacy by offering reliable predictions that facilitate the creation of efficient treatment plans. Reliable models can potentially enhance patient outcomes by promoting early intervention and the identification of particular cancer subtypes.

**Actionable Insights:** In cancer models, diagnostic accuracy and dependability boost diagnostic confidence, expedite the diagnosis procedure, and facilitate customised treatment regimens. Accurate subtype classification aids in the development of therapeutic interventions for particular cancer types. Effective diagnostic models also help distribute resources, easing the strain on labs and permitting patient care. Research on cancer subtype mechanisms is guided by precise models, which result in novel treatments. Higher survival and quality of life rates and better patient outcomes result from improved diagnosis accuracy.

The experimental results showing high accuracy and low loss values in cancer subtype classification models are particularly significant for clinical practice. These measurements show how these models can increase the accuracy of diagnoses, customise treatment regimens, and ultimately improve patient care. Healthcare professionals can make better decisions, lower the rate of misdiagnosis, and give tailored therapies that enhance patient outcomes by incorporating these data into clinical procedures. This combination of cutting-edge machine learning methods and medical knowledge is a positive development in the campaign against cancer.

#### 4.3 Performance Measures

In addition, the area under the ROC curves is plotted using the Tomlins-2006-v1, Khan-2001,

Lapoint-2004-v2, and Tomlin-2006-v2 Datasets. A ROC figure shows Sensitivity on the Y axis and Specificity on the X axis.

Where:

$$\text{True positive rate} = \frac{TP}{TP + FN}$$

$$\text{False positive rate} = \frac{FP}{FP + TN}$$

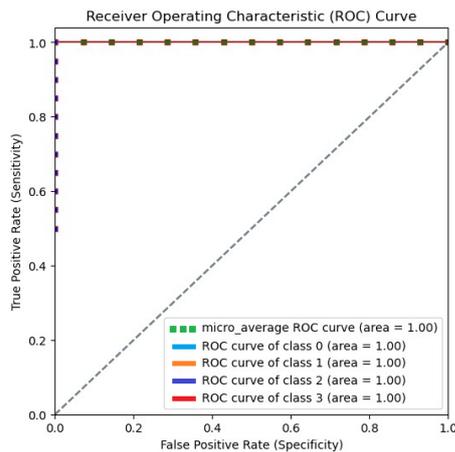


Fig 6. Roc curve of Khan-2001

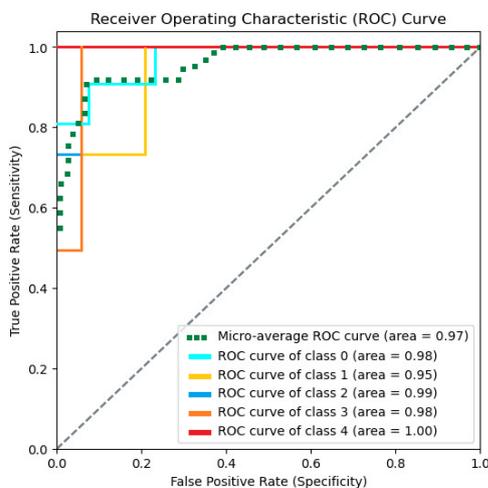


Fig 7. Roc curve of Tomlins-2006-v1

The micro-average ROC curve and the AUC of 1.00 for all classes (class 0, class 1, class 2 and class 3) in Figure 6 show that the model performs exceptionally well in classification with Khan-2001, correctly differentiating between all classes with no errors. This kind of performance is uncommon and usually suggests that the model was overfitted in the first place or that the dataset could have been better validated than it could have been using separate test data. In Figure 7, the macro-average ROC curve of Tomlins-2006-

v1 is 0.97. The ROC curve contains five classes such as class 0, class 1, class 2, class 3, and class 4, and has an area of 0.98, 0.95, 0.99, 0.98, and 1.00.

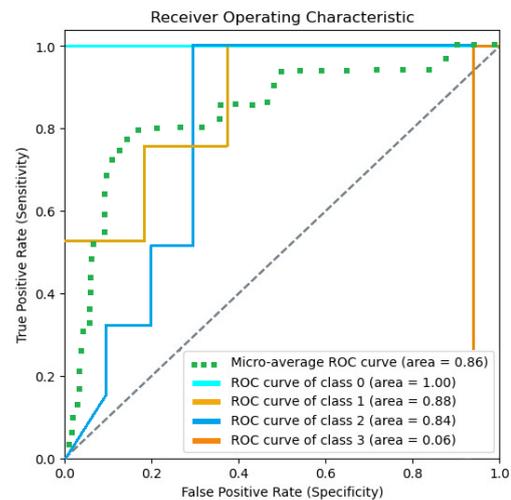


Fig 8. Roc curve of Tomlin-2006-v2.

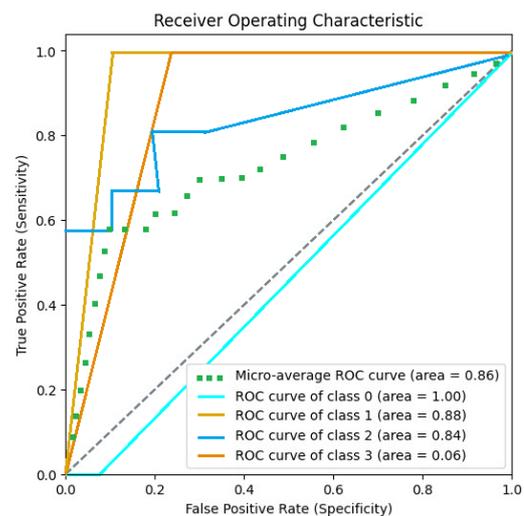


Fig 9. Roc curve of Lapoint-2004-v2.

The areas of the four classes on the ROC curve of Tomlin-2006-v2 and Lapoint-2004-v2 in Figure 8 and Figure 9 are the same, whereas class 0, class 1, class 2, class 3 (1.00 0.88, 0.84, and 0.06). While the model performs admirably for class 0 and passably for classes 1 and 2, it fails miserably for class 3. This is illustrated by the total performance shown in Figure 9. This discrepancy implies that more data or adjustments may be required to enhance the model's discriminating influence for class 3. The model's generalizability and robustness across all classes may be impacted by potential imbalances or problems in the dataset, as indicated by the high AUC for class 0 and the low AUC for class 3. The ROC curves (AUC) for the Tomlins-

2006-v1, Khan-2001, Lapoint-2004-v2, and Tomlin-2006-v2 datasets are displayed in Figures 6 to 9, respectively. The micro-averages we found were 0.98, 1.00, 0.85, and 0.88, respectively.

## 5. Conclusions

The research concludes by emphasising the vital significance of appropriately classifying cancer subdivisions depending on gene expression data to further our acceptance of the molecular mechanisms behind this complicated disease. Although classic machine learning methods have achieved substantial advancements in this area, deep structured reinforcement learning (DSRL) is a promising new approach for overcoming the difficulties brought on by the complexity and dimensionality of gene expression datasets. The performance of a novel approach that combines deep neural networks with reinforcement learning to classify cancer subtypes from gene expression data is improved, outperforming existing classifiers in terms of accuracy and loss outcomes. This methodology offers a reliable and cutting-edge way of classifying multiclass cancer subtypes by utilising the strength of deep neural networks and reinforcement learning. In terms of all seven datasets, the comparison based on gene expression data demonstrates that the suggested approach overtakes the existing state-of-the-art classifiers, obtaining the greatest accuracy of 55% to 100% and the lowest loss of 0.11 to 0.02. It is an essential advancement in the ongoing battle to detect and treat cancer, ultimately leading to better patient outcomes and a better understanding of this leading cause of death globally. To improve classification accuracy and robustness, future work on cancer subtype classification may investigate improved model interpretability through attention mechanisms and integrate multi-modal data fusion techniques.

## References

- Abd-Elnaby, M., Alfonse, M., & Roushdy, M. (2021). Classification of breast cancer using microarray gene expression data: A survey. *Journal of biomedical informatics*, 117, 103764.
- Alimirzaei, F., & Kieslich, C. A. (2023). Machine learning models for predicting membranolytic anticancer peptides, In *Computer Aided Chemical Engineering*, 52, 2691-2696.
- Anna, A., Monika, G. (2018). Splicing Mutations in Human Genetic Disorders: Examples, Detection, and Confirmation, *J. Appl. Genet.*, 59, 253–268.
- Ashtari, P., Haredasht, F. N., & Beigy, H. (2020). Supervised fuzzy partitioning. *Pattern Recognition*, 97, 107013.
- Bellemare, M. G., Ostrovski, G., Guez, A., Thomas, P., & Munos, R. (2016). Increasing the action gap: New operators for reinforcement learning, In *Proceedings of the AAAI Conference on Artificial Intelligence*, 30(1).
- Bergstra, J., Bardenet, R., Bengio, Y., & Kégl, B. (2011). Algorithms for hyper-parameter optimization, *Advances in neural information processing systems*, 24.
- Bhandari, N., Walambe, R., Kotecha, K., & Khare, S. P. (2022). A comprehensive survey on computational learning methods for analysis of gene expression data. *Frontiers in Molecular Biosciences*, 9, 907150.
- Bhonde, S.B., Prasad, J.R. (2021). Deep Learning Techniques in Cancer Prediction Using Genomic Profiles, In *Proceedings of the 2021 6th International Conference for Convergence in Technology (I2CT)*, Maharashtra, India, 1–9.
- Brewczyński, A., Jabłońska, B., Mazurek, A.M., Mrochem-Kwarciak, J., Mrowiec, S., Śnietura, M., Kentnowski, M., Kołosza, Z., Składowski, K., & Rutkowski, T. (2021). Comparison of Selected Immune and Hematological Parameters and Their Impact on Survival in Patients with HPV-Related and HPV-Unrelated Oropharyngeal Cancer, *Cancers*, 13, 3256.
- Celesti, F., Celesti, A., Wan, J., & Villari, M. (2018). Why Deep Learning Is Changing the Way to Approach NGS Data Processing: A Review, *IEEE Rev. Biomed. Eng.*, 11, 68–76.
- Chabon, J. J., Hamilton, E. G., Kurtz, D. M., Esfahani, M. S., Moding, E. J., Stehr, H., & Diehn, M. (2020). Integrating genomic features for non-invasive early lung cancer detection, *Nature*, 580(7802), 245-251.

- Crosby, D., Bhatia, S., Brindle, K. M., Coussens, L. M., Dive, C., Emberton, M., & Balasubramanian, S. (2022). Early detection of cancer. *Science*, 375(6586), eaay9040.
- Cun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature*, 521, 436-444.
- De Souto, M. C., Costa, I. G., de Araujo, D. S., Ludermit, T. B., & Schliep, A. (2008). Clustering cancer gene expression data: a comparative study. *BMC Bioinformatics*, 9(1), 1-14.
- Divate, M., Tyagi, A., Richard, D. J., Prasad, P. A., Gowda, H., & Nagaraj, S. H. (2022). Deep learning-based pan-cancer classification model reveals tissue-of-origin specific gene expression signatures, *Cancers*, 14(5), 1185.
- Dudoit, S., Fridlayand, J. and Speed, T. P. (2002). Comparison of discrimination methods for the classification of Tumors using gene expression data. *Journal of the American Statistical Association*, 97(457).
- Goodarzi, K., Lane, R., & Rao, S. S. (2024). Varying the RGD concentration on a hyaluronic acid hydrogel influences dormancy versus proliferation in brain metastatic breast cancer cells, *Journal of Biomedical Materials Research Part A*, 112(5), 710-720.
- Hassani, H., Avazzadeh, Z., Agarwal, P., Mehrabi, S., Ebadi, M. J., Dahaghin, M. S., & Naraghirad, E. (2023). A study on fractional tumor-immune interaction model related to lung cancer via generalized Laguerre polynomials, *BMC medical research methodology*, 23(1), 189.
- Hausknecht, M., & Stone, P. (2015). Deep recurrent q-learning for partially observable mdps, In 2015 aaai fall symposium series.
- Heydarpoor, F., Karbassi, S. M., Bidabadi, N., & Ebadi, M. J. (2020). Solving multi-objective functions for cancer treatment by using Metaheuristic Algorithms, *International Journal of Combinatorial Optimization Problems & Informatics*, 11(3).
- Hijazi, H., & Chan, C. (2013). A classification framework applied to cancer gene expression profiles, *Journal of Healthcare Engineering*, 4, 255-283.
- Hu, J., Shen, L., & Sun, G. (2018). Squeeze-and-excitation networks, In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 7132-7141.
- Islam, M. M., Huang, S., Ajwad, R., Chi, C., Wang, Y., & Hu, P. (2020). An integrative deep learning framework for classifying molecular subtypes of breast cancer, *Computational and structural biotechnology journal*, 18, 2185-2199.
- Jayakrishnan, R., & Meera, S. (2023). Reinforcement Learning Technique Based Automated Feature Analysis of Gene Expression Data, In *2023 International Conference on Circuit Power and Computing Technologies (ICCPCT)*, 1855-1860.
- Jayakrishnan, R., & Sridevi, S. (2022). A Comparative Study of Gene Expression Data-Based Intelligent Methods for Cancer Subtype Detection, In *IOT with Smart Systems: Proceedings of ICTIS 2022*, 2, 457-467.
- Jayashri, N., & Deepika, M. Enhanced Artificial Bee Colony Based Flexible Neural Forest (Eabc-Fnt) And Ensemble Gene Selection (Egs) For Cancer Subtypes Classification On Gene Expression Data.
- Jenul, A., Stokmo, H. L., Schrunner, S., Hjortland, G. O., Revheim, M. E., & Tomic, O. (2024). Novel ensemble feature selection techniques applied to high-grade gastroenteropancreatic neuroendocrine neoplasms for the prediction of survival. *Computer Methods and Programs in Biomedicine*, 244, 107934.
- Karim, M. R., Beyan, O., Zappa, A., Costa, I. G., Rebholz-Schuhmann, D., Cochez, M., & Decker, S. (2021). Deep learning-based clustering approaches for bioinformatics, *Briefings in Bioinformatics*, 22(1), 393-415.
- Khan, M.F., Ghazal, T.M., Said, R.A., Fatima, A., Abbas, S., Khan, M.A., Issa, G.F., Ahmad, M., Khan, M.A. (2021). An IoMT-Enabled Smart Healthcare Model to Monitor Elderly People Using Machine Learning Technique, *Comput. Intell. Neurosci.*, 2021, 2487759.
- Khorshed, T., Moustafa, M. N., & Rafea, A. (2020). Deep learning for multi-tissue cancer classification

- of gene expressions (GeneXNet), IEEE Access, 8, 90615-90629.
- Korbar, B., Olofson, A. M., Miraflor, A. P., Nicka, C. M., Suriawinata, M. A., Torresani, L., & Has-sanpour, S. (2017). Deep learning for classification of colorectal polyps on whole-slide images, *Journal of pathology informatics*, 8(1), 30.
- Le, N., Rathour, V. S., Yamazaki, K., Luu, K., & Savvides, M. (2022). Deep reinforcement learning in computer vision: a comprehensive survey, *Artificial Intelligence Review*, 1-87.
- Lunshof, J.E., Bobe, J., Aach, J., Angrist, M., Thakuria, J.V., Vorhaus, D.B., Hoehe, M.R., Church, G.M. (2010). Personal Genomes in Progress: From the Human Genome Project to the Personal Genome Project, *Dialogues Clin. Neurosci.*, 12, 47-60.
- Miller, K.D., Ortiz, A.P., Pinheiro, P.S., Bandi, P., Minihan, A., Fuchs, H.E., Martinez Tyson, D., Tortolero-Luna, G., Fedewa, S.A., Jemal, A.M., etc. (2021). Cancer Statistics for the US Hispanic/Latino Population, *CA A Cancer J. Clin.*, 71, 466-487.
- Min, S., Lee, B. and Yoon, S. (2017). Deep learning in bioinformatics. *Brief Bioinform.*, 18(5), 851-869.
- Mostavi, M., Chiu, Y. C., Huang, Y., & Chen, Y. (2020). Convolutional neural network models for cancer type prediction based on gene expression. *BMC Medical Genomics*, 13, 1-13.
- Munkácsy, G., Santarpia, L., Györffy, B. (2022). Gene Expression Profiling in Early Breast Cancer Patient Stratification Based on Molecular and Tumor Microenvironment Features, *Biomedicines*, 10, 248.
- Perdomo-Ortiz, A., Benedetti, M., Realpe-Gómez, J., & Biswas, R. (2018). Opportunities and challenges for quantum-assisted machine learning in near-term quantum computers, *Quantum Science and Technology*, 3(3), 030502.
- Prathik, A., Vinodhini, M., Karthik, N., & Ebenezer, V. (2022). Prediction of Carcinoma Cancer Type Using Deep Reinforcement Learning Technique from Gene Expression Data, In *Intelligent Data Communication Technologies and Internet of Things: Proceedings of ICICI 2021*, 541-552.
- Ram, M., Najafi, A., & Shakeri, M. T. (2017). Classification and biomarker gene selection for cancer gene expression data using random forest, *Iranian journal of pathology*, 12(4), 339.
- Rodriguez-Ezpeleta, N. (2012). *Bioinformatics for High Throughput Sequencing*, New York: Springer.
- Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. C. (2018). Mobilenetv2: Inverted residuals and linear bottlenecks, In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 4510-4520.
- Schaul, T., Quan, J., Antonoglou, I., & Silver, D. (2015). Prioritized experience replay. arXiv preprint arXiv:1511.05952.
- Segal, N. H., Pavlidis, P., Noble, W. S., Antonescu, C. R., Viale, A., Wesley, U. V., & Houghton, A. N. (2003). Classification of clear-cell sarcoma as a subtype of melanoma by genomic profiling, *Journal of Clinical Oncology*, 21(9), 1775-1781.
- Shah, S. H., Iqbal, M. J., Ahmad, I., Khan, S., & Rodrigues, J. J. (2020). Optimized gene selection and classification of cancer from microarray gene expression data using deep learning, *Neural Computing and Applications*, 1-12.
- Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., Van Den Driessche, G., & Hassabis, D. (2016). Mastering the game of Go with deep neural networks and tree search, *nature*, 529(7587), 484-489.
- Tabares-Soto, R., Orozco-Arias, S., Romero-Cano, V., Bucheli, V. S., Rodríguez-Sotelo, J. L., & Jiménez-Varón, C. F. (2020). A comparative study of machine learning and deep learning algorithms to classify cancer types based on microarray gene expression data, *PeerJ Computer Science*, 6, e270.
- Tarca, A. L., Romero, R., & Draghici, S. (2006). Analysis of microarray experiments of gene expression profiling, 195(2), 373-388.
- Van Seijen, H., & Sutton, R. (2013). Planning by prioritized sweeping with small backups, In *International Conference on Machine Learning*, 361-369.

- Xia, D., Alexander, A. K., Isbell, A., Zhang, S., Ou, J., & Liu, X. M. (2017). Establishing a co-culture system for *Clostridium cellulovorans* and *Clostridium acetivum* for high efficiency biomass transformation, *J. Sci. Heal. Univ. Ala*, 14, 8-13.
- Xu, J., Wu, P., Chen, Y., Meng, Q., Dawood, H., & Dawood, H. (2019). A hierarchical integration deep flexible neural forest framework for cancer subtype classification by integrating multi-omics data, *BMC Bioinformatics*, 20(1), 1-11.
- Xu, J., Wu, P., Chen, Y., Meng, Q., Dawood, H., & Khan, M. M. (2019). A novel deep flexible neural forest model for classification of cancer subtypes based on gene expression data, *IEEE Access*, 7, 22086-22095.
- Yuan, L. M., Sun, Y., & Huang, G. (2020). Using class-specific feature selection for cancer detection with gene expression profile data of platelets, *Sensors*, 20(5), 1528.
- Yuan, L., Sun, Y., & Huang, G. (2020). Using Class-Specific Feature Selection for Cancer Detection with Gene Expression Profile Data of Platelets, *Sensors*, 20, 1528.
- Zahoor, J., & Zafar, K. (2020). Classification of microarray gene expression data using an infiltration tactics optimization (ITO) algorithm, *Genes*, 11(7), 819.
- Zhang, Y., Deng, Q., Liang, W., & Zou, X. (2018). An efficient feature selection strategy based on multiple support vector machine technology with gene expression data, *BioMed research international*, 2018.
- Zhu, W., Xie, L., Han, J., & Guo, X. (2020). The application of deep learning in cancer prognosis prediction, *Cancers*, 12(3), 603.

# Efficient task scheduling in the cloud with queuing and multi-tactic harris hawks optimization

Sheetal Antony<sup>1\*</sup>, Sujatha S R<sup>2</sup>

<sup>1</sup>Ajoshia Bio Teknik Pvt. Ltd

Computer Science and Engineering, Sri Siddhartha Institute of Technology, SSAHE

SSIT Maralur, Tumakuru, 572105.

e-mail: sheetal123876@gmail.com

(Received 24 February 2024; Final version received 18 July 2024; Accepted 31 July 2024)

## Abstract

Cloud computing faces challenges in task scheduling, which is crucial for cost-efficient execution and resource utilization. Current methods face computational complexity, especially in large-scale data centres. This paper proposes a novel approach that considers job dependencies and task execution times to reduce make-span, minimize energy consumption, and balance resource loads. VMs are allocated based on workflow task requirements, using thresholds for task levels and durations to manage execution priorities. Tasks with higher dependencies and longer execution times are prioritized, ensuring efficient resource utilization and energy savings. The method employs queues for different task intensities, streamlining VM allocation by organizing tasks with additional metadata like intensities, arrival times, and deadlines. Historical scheduling logs (HSLs) are used to generate appropriate VMs, with new VMs created if no matching records exist in the HSLs. The proposed solution optimizes scheduling using an enhanced Multi-Tactic Harris Hawks Optimization (MTHHO) algorithm, which addresses the limitations of traditional HHO by incorporating Sobol sequences, elite opposition-based learning, and improved energy updating techniques to enhance population diversity, adaptability, and convergence accuracy while avoiding local optima using the Gaussian walk learning. The result shows that the proposed method of QoS performances attained less Makespan, energy consumption of 0.20, throughput of 2.4, and execution time of 16.75 with effectively allocated resources of 98% when compared to the previous methods in cloud computing. Therefore, the proposed heuristic-based MTHHO method balanced the load and allocated the resources effectively to improve QoS performances.

*Keywords: Cloud Computing; Thresholds; Energy Consumption; Queuing; Task Scheduling; Multi-Tactic HHO; non-linear weight; Gaussian walk learning; Load Balancing; Makespan;*

## 1. Introduction

Cloud computing systems were created based on the enormous growth in internet data processing. When it comes to giving technology facilities online, cloud computing is crucial. Without direct active control, it gives users access to computer system resources like data storage and processing power. Three different services about infrastructure, platforms, and software can be offered by a cloud. Infrastructure as a Service (IaaS), which offers infrastructure services including storage systems and computing resources, is the first service. Platform as a service (PaaS), the second offering, allows customers to generate presentations based on the platform that is made available. The third service, known as software as a service (SaaS), offers customers the option of using software straight from

the cloud without having to install anything locally [Devaraj etc., 2020, Alam 2021]. Cloud computing allows for the flexible and elastic provision of varied computing resources in response to user demands; in recent years, such large-scale applications have used cloud computing at an increasing rate [Cui etc., 2021]. A cloud data centre's infrastructure typically comprises thousands of big computing hosts with fast computing power [Katal etc., 2023].

Virtual machines (VMs) are a type of computing resource that cloud providers utilize to deliver computing resources to users. To increase the general effectiveness of cloud computing, effective task scheduling is needed when several users make task requests for services from the cloud. Efficient task scheduling allows the optimal resource allocation between the requested tasks over a limited amount of period enabling

the achievement of the desired degree of quality of service (QoS) [Adhikari etc., 2019, Gawalil etc., 2018]. The primary categories of scheduling mechanisms used in cloud computing are workflow, static, cloud service, and dynamic scheduling [Alsaady etc., 2022]. The difficulty of the process allows for the classification of scheduling methods as heuristic, meta-heuristic, and hybrid task scheduling techniques. Static scheduling employs heuristic approaches such as minimum execution time (MET), minimum completion time (MCT), shortest job to fastest processor (SJFP), longest job to fastest processor (LJFP), Min-Min, and Max-Min [Abd etc., 2019, Mishra etc., 2020]. In a cloud environment, it can be challenging to schedule tasks and allocate resources in the best possible order and with the least amount of delay to increase system presentation. The complexities of the cloud, real-time task mapping to virtual machines, and virtual machine mapping to the host machine make task scheduling in cloud computing an NP-Hard problem [Yadav etc., 2023, Golchi etc., 2019].

Meta-heuristic algorithms have recently captured the interest of researchers due to their capacity to solve large-scale issues efficiently. For NP-Hard problems, these algorithms can efficiently search a wide area of the solution space for a solution that is close to optimal [Abdullahi etc., 2023]. Meta-heuristic algorithms like the genetic algorithm (GA) [Keshanchi et al., 2017], ant colony optimization (ACO) [Mahato etc., 2017], particle swarm optimization (PSO) [Mansouri etc., 2019, Kumar etc., 2018], discrete symbiotic organism search (DSOS) [Abdullahi etc., 2016], and gravitational search algorithm (GSA) [Chaudhary etc., 2018] have recently been used to solve task scheduling issues [Agarwal etc., 2021, Wei 2020]. However, metaheuristic algorithms possess two major shortcomings: The first is that they are computationally intensive and can get stuck in local optimum states, particularly in large solution spaces. According to Konjaang and Xu Ramamoorthy et al [Konjaang etc., 2021, Ramamoorthy etc., 2021], an inequity among local and global search strategies may origin convergence to occur too early.

Hybrid meta-heuristic algorithms were employed by the researchers to achieve improved performance. Examples of this hybridization include the hybridized whale optimization method [Strumberger etc., 2019], firefly and PSO [1, 11], Q-learning and PSO [Jena etc., 2022], as well as firefly and simulated annealing [Fanian etc., 2018]. The contribution of the work is outlined as:

- Energy-efficient VMs are crucial for cost-effective, elastic computing in cloud data centres, but their computational complexity can limit their usability in dynamic environments and hinder real-time responsiveness and scalability.
- A proposed approach focuses on job dependencies and task execution times to shorten make-span, consume less energy, and balance the load on available resources.
- Tasks are assigned VMs based on workflow tasks, with tasks with longer execution times handled first. Queues are maintained based on job intensities, and tasks are stored in queues using additional information.
- The enhanced Multi-Tactic Harris Hawks Optimization (MTHHO) algorithm is used to optimize scheduling issues. The algorithm uses Sobol sequences, elite opposition-based learning, and the Gaussian walk learning technique to improve the population's variety, adaptability, and energy updating.

The rest of the manuscript is organized as given. An introduction is given in Section 1, related works are presented in Section 2, and effective task scheduling in cloud computing using queuing and multi-tactic Harris Hawks Optimisation modelling is covered in Section 3. Sections 4 as well as 5 provide the model results, and the conclusion correspondingly.

## 2. Related Works

Several studies are involved on the topic of cloud computing, including scheduling, load balancing, and resource provisioning. Numerous problems with cloud computing have captured the attention and concern of researchers. Resource management, load balancing, cloud migration, privacy and security, energy consumption, availability and scalability, interoperability, and compatibility are a few of these significant concerns. These challenges are strongly influenced by investigating effective task scheduling. The fundamental to task scheduling in cloud computing is to identify the most optimal mapping connection among tasks and virtual machines depending on the objectives of users and cloud systems. One of the main ways to deal with this issue is to find a more effective algorithm to support job scheduling, such as a single-objective optimization algorithm or a multi-objective optimization algorithm. The following section discusses some of the utmost recent task-scheduling approaches.

Kaur, etc., outlined the major research requirements for load balancing optimization in the prior works that must be filled to address the load balancing problem in cloud environments. To maximize the use of VMs with uniform load distribution, a framework for resource provisioning and a combination method for load balancing has been created in the current work. The suggested system is based on the fusion of heuristic techniques with meta-heuristic algorithms to achieve the greatest efficiency in making span and cost. For the HDD-PLB system, two hybrid methods have been proposed: the Hybrid Heterogeneous Earliest Finish Time (HEFT) Heuristic with ACO (HHA) and the Hybrid Predict Earliest Finish Time (PEFT) Heuristic with ACO meta-heuristic (HPA). The two load-balancing approaches have been analysed and contrasted for the suggested HDD-PLB system to determine which is superior. However, the suggested framework is based on financial restraints, limiting the execution of workflow tasks that exceed deadlines in terms of total cost. The normal time and cost findings of the 100 repetitions were not measured for this research.

Kruekaew, etc., suggested the MOABCQ method as a standalone task scheduling method for cloud computing to tackle workload balancing problems with a Multi-objective task scheduling optimization depending on the Artificial Bee Colony Algorithm (ABC) with a Q-learning algorithm, which is a reinforcement learning method that assists the ABC process work more rapidly. The proposed solution addresses the limitations of simultaneous concerns by maximizing VM throughput, optimizing scheduling and resource consumption, and establishing load balancing among VMs according to make span, cost, and resource utilization. The efficiency study of the suggested approach was contrasted utilizing CloudSim with the load balancing and scheduling methods currently in use: Max-Min, FCFS, HABC\_LJF, Q-learning, MOPSO, and MOCS in three datasets: Random, Google Cloud Jobs (GoCJ), and Synthetic workload. According to the findings, MOABCQ-based algorithms beat other algorithms on account of lowering makespan, cost, degree of imbalance, boosting throughput, and utilizing resources regularly. However, it cannot ensure that the MOABCQ\_LJF method is best, and also not all test datasets can be used to optimize the system's performance.

Velliangiri etc. proposed a Hybrid Electro Search with a Genetic Algorithm (HESGA) to improve work scheduling performance by accounting for aspects

such as makespan, balance of load, utilisation of resources, and multi-cloud costs. The proposed method integrates the advantages of genetic and electro-search algorithms. The Electro search method generates the finest global optimum results, while the GA generates the finest local optimal results. The suggested technique outperforms current scheduling methods as the Hybrid Particle Swarm Optimization Genetic technique (HPSOGA), GA, ES, and ACO. However, there are no guarantees that the algorithm will find the globally best solution or that it will always lead to optimal resource use, which is essential for cloud computing to be cost-effective.

Rajakumari, etc., proposed the Fuzzy Based Ant Colony Optimization Scheduling approach, however, is used in cloud computing to address task scheduling issues like optimal task scheduling presentation results. First, by suggesting a Dynamic Weighted Round-Robin method, work scheduling performance in the cloud is enhanced. The performance of work scheduling is enhanced by the suggested DWRR algorithm by taking into account resource competence, task priority, and length. Next, a hybrid particle swarm parallel ant colony optimization heuristic approach is suggested to address the task execution delay issue in DWRR-based task scheduling. Finally, HPSPACO develops a fuzzy logic system that enhances job scheduling in the cloud system. For the inertia weight updates of the PSO and pheromone trails updating of the PACO, a fuzzy technique is suggested. To optimize work scheduling, the proposed Fuzzy HPSPACO on cloud computing reduces implementation and waiting times boosts system throughput and maximizes resource usage. Conversely, the CF-ACO algorithm is a complex algorithm with multiple elements, including fuzzy logic and ant colony optimization, which can result in fundamental processing overhead and difficulties in finding optimal solutions in real-time or large-scale cloud systems.

Saxena, D., etc., proposed a unique secure and multi-objective virtual machine placement (SM-VMP) model using a successful VM migration to tackle the problems of resource waste, excessive usage of power, higher inter-communication costs, and security breaches. The suggested approach emphasizes the safe and quick operation of user applications by minimizing inter-communication latency and enabling an energy-efficient allocation of physical resources across VMs. The VMP is implemented using the suggested Whale Optimisation Genetic Algorithm (WOGA), which is motivated by non-dominated

sorting-based genetic algorithms and whale evolutionary optimization. The results of the assessment for static and dynamic VMP as well as contrasting it using contemporary advanced revealed a considerable drop in shared servers, intercommunication costs, power consumption, and processing period up to 28.81%, 25.7%, 35.9%, and 82.21%, correspondingly. Resource utilization also improved up to 30.21%. However, multi-objective optimization issues in cloud data centres can be computationally complex, time-intensive, and difficult to solve.

### 3. Proposed Methodology

The allocation of energy-efficient virtual machines (VMs) is crucial for cost-effective and elastic computing in cloud data centres. However, the computational complexity of existing work, particularly in large-scale data centres and complex optimization objectives, can limit the framework's usability in dynamic, resource-intensive environments and impair its real-time responsiveness and scalability. Multi-objective optimization can be computationally intensive and time-consuming. It is crucial to understand how computational complexity impacts the usability of the systems. High computational complexity can cause delays in VM allocation, impacting cloud service responsiveness and performance. As data centres scale, managing resource complexity becomes more complex, requiring clear management strategies to ensure system effectiveness at larger scales. Complex algorithms may require more computational power, potentially offsetting energy savings achieved through efficient VM allocation. Implementing sophisticated algorithms can pose challenges, but clear guidelines and understanding can help ensure smooth deployment and operation. The suggested approach focuses on job dependencies and task execution times to shorten make-span, consume less energy, and balance the load on available resources. To preserve a balanced load on the resources, VMs are assigned tasks depending on the necessities of the workflow tasks. The suggested method scans the workflow tasks and establishes thresholds for the tasks' level and duration. When tasks are being executed, the threshold values are employed to handle them based on various priorities.

Tasks with higher dependencies generate system bottlenecks and extended execution durations. To shorten execution time, tasks with longer execution times are handled first, requiring high-priority processing and allocating VMs with powerful processing

capabilities. The algorithm also prioritizes tasks with lengthy execution times, setting thresholds for dependencies and duration. This shortens execution time and reduces energy consumption by effectively utilizing resources. The suggested approach then employs queues based on the job intensities, maintaining distinct queues for CPU-intensive tasks and tasks with higher dependents. It takes less time to discover the right VMs during the VM allocation process when tasks of different intensities are placed in distinct queues. Tasks are stored in queues using additional information about them, such as their intensities, arrival times, and deadlines. The following step is to generate appropriate VMs for the tasks following the classification of tasks into various queues. Historical scheduling Logs HSLs are employed for this objective. The HSLs are updated appropriately and a novel VM is generated with the resources required to complete the task if there is no matching record in the HSLs. The suggested algorithm optimizes the scheduling issue using the enhanced Multi-Tactic Harris Hawks Optimization (MTHHO) algorithm. Scheduling with MTHHO begins following the pre-processing step. An enhanced MTHHO algorithm is suggested to make up for the traditional HHO process's low convergence accuracy, slower degree of convergence, and easy tendency to prey to the illusion of local optima. To improve the population's variety, Sobol sequences are first utilized to start the population. The next step to raise the adaptability and the standard of the solution sets, the elite opposition-based learning technique is used. Additionally, the original algorithm's energy updating technique has been improved to increase the process's capability to explore as well as exploit in a nonlinear update way. To prevent the process from being stuck and settling into a local optimum, the Gaussian walk learning technique is finally employed. The suggested framework for task scheduling utilizing the enhanced MTHHO process is shown in Fig 1.

The proposed approach to task scheduling optimizes resource utilization, reduces make-span, and saves energy by using energy-efficient virtual machines (VMs) and optimizing scheduling. It considers job dependencies and task execution times, maximizing resource allocation, reducing idle times, and minimizing make-span. The approach prioritizes tasks based on dependencies and execution times, ensuring critical tasks are completed promptly. It also balances resource loads across the data centre, preventing overutilization or underutilization, and improving the stability and reliability of the cloud infrastructure. The

enhanced MTHHO algorithm, using advanced techniques like Sobol sequences and elite opposition-based learning, enhances scheduling accuracy, leading to better convergence on optimal solutions and avoiding common pitfalls like local optima. Overall, this approach significantly improves task scheduling efficiency and reduces energy consumption.

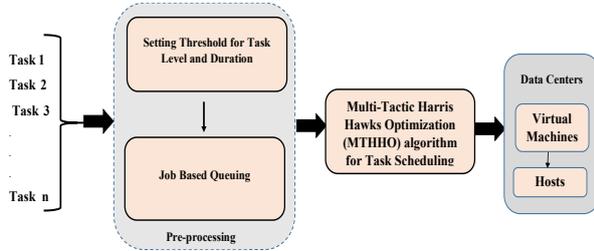


Fig 1. The Proposed Framework for Task Scheduling

### 3.1 Materials and Approaches

The proposed method is explained further in this part of the article. The process focuses on load balance, makespan, and energy consumption. The process is separated into dual phases: MTHHO-based optimization and preprocessing. The process and cloud framework are covered initially, trailed by the information of every stage.

### 3.2 Workflow and Cloud Architecture

Workflow programmes are made up of tasks that have reliance, such as implementation and data reliance. The tasks in the previous examples have a parent-child relationship. When the execution of each parent's job has finished, the child's task can begin. In the latter scenario, the tasks exchange data, meaning that the result produced by one job is used as the input for another. It is challenging for a scheduler to efficiently schedule resources for workflow programs because of these dependencies. A directed acyclic graph (DAG), such as  $D(V, E)$ , is used to represent workflow activities. In this graph,  $E$  stands for the edges while  $V$  stands for the vertices. VMs, or virtualized resources, are the building blocks of cloud computing. A scheduling process's objective is to assign  $R_i$  to  $W_j$ , where  $W_j$  is the workflow application ( $W_1, W_2, W_3, \dots, W_m$  and  $R_i$  is the  $i$ th resource from a pool of VMs ( $VM_1, VM_2, VM_3, \dots, VM_n$ ). The objective is to reduce energy usage and

implementation period while maintaining a balanced load on the available resources. Processing, memory, storage, bandwidth, and other capacities have been pre-allocated for the resources. Equation (1) illustrates how the number of processing elements (PEs) and MIPS of each PE are used to calculate the processing capacity ( $C_i$ ) of a resource  $VM_i$ .

$$C_i = (PE \times MIPS_i) \quad (1)$$

Equation (2) is used to determine the capacity of  $n$  resources, or virtual machines.

$$C = \sum_{i=1}^n C_i \quad (2)$$

Every VM has a resource utilisation at any given period, which is called the VM load. Equation (3) is used to compute the load, where  $TL$  the total length of tasks that  $VM_i$  is processing and  $C_i$  is  $VM_i$ 's capacity.

$$L_{vmi} = \frac{TL}{C_i} \quad (3)$$

Equation (4) is utilised to determine each VM's load  $L$ .

$$L = \sum_{i=1}^n L_{vmi} \quad (4)$$

Equation (5) illustrates how load balancing is calculated as the load across various cloud environment nodes.

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (L_{vmi} - \bar{L})^2}{n}} \quad (5)$$

where  $\bar{L}$  is the average load across all VMs,  $n$  is the amount of VMs, and  $L_{vmi}$  is the load of  $VM_i$ . The way resources are used has a big impact on how much energy cloud computing uses. Equation (6) can be used to compute the utilisation.

$$U = \alpha \frac{\sum_{i=1}^n c_i}{C} + \beta \frac{\sum_{i=1}^n m_i}{M} \quad (6)$$

where  $n$  is the number of VMs running on host  $h$ , and  $c_i$ ,  $m_i$  denotes the computing and memory assigned to  $VM_i$ . In Eq. (6),  $C$  and  $M$  are the total processing ability and memory of the host, and  $\alpha$  and  $\beta$  are the weight factors of every resource.

Equation (7) can be used to compute the energy consumption, with  $k$  standing for the functioning energy consumption, or idle mode.  $U$  is the host

resource utilisation determined by Equation (6), and  $E_{max}$  denotes the energy consumption during the processors' peak utilisation.

$$E_c = E_{idle} + (E_{max} - E_{idle}) \times U \quad (7)$$

Data processing may be required for workflow tasks. The workflow's total completion time accounts for both processing time and time spent acquiring the necessary data. The completion time of task  $t_i$  is calculated using Equation (8).

$$Time(t_i) = Time((Trans_{t_i, t_j}) + Time_{E}(t_i, VM_k)) \quad (8)$$

The time required to transmit data from task  $t_i$  to task  $t_j$  is denoted by  $(Trans_{t_i, t_j})$  in Equation (8), while the execution time of  $t_i$  on  $VM_k$  is represented by  $Time_E(t_i, VM_k)$ . Equations (9) and (10) are used to determine the two parameters, respectively.

$$Trans(t_i, t_j) = \frac{sizeof(t_i, t_j)}{\beta(VM_k, VM_m)} \quad (9)$$

The quantity of data that task  $t_i$  transfers to task  $t_j$  is represented by  $sizeof(t_i, t_j)$  in Equation (9), and the bandwidth consumed by  $VM_k$  and  $VM_m$  is represented by  $\beta(VM_k, VM_m)$ . The cost of transmission is disregarded if both virtual machines are placed in the same data centre.

$$T_E = \frac{l_i}{C_{m_j}} \quad (10)$$

where  $C_{m_j}$  is the processing capability of  $VM_j$ , which was determined using Equation (1), and  $l_i$  is the length of job  $i$ . A workflow's makespan is the entire period it gains to complete every task. Equation (11), where  $MS$  is the makespan and  $FT$  is the task's completing time, can be used to compute the makespan.

$$MS = FT_{i=1}^n [task_i time] \quad (11)$$

Energy-efficient virtual machines (VMs) significantly reduce operational costs by reducing power consumption in data centres, a crucial factor in a competitive market. They also support the elastic nature of cloud computing, allowing data centres to dynamically scale resources based on demand. This elasticity ensures optimal resource utilization without unnecessary

energy expenditure. Additionally, the proposed method simplifies the allocation process by considering job dependencies and task execution times, making it more manageable and efficient even in complex scenarios.

### 3.3 Harris Hawks Optimization (HHO)

The HHO process is a mathematical explanation of the Harris hawk's technique for catching prey under various circumstances. Each iteration's best answer is regarded as the prey, while individual Harris hawks form candidate solutions. The process is allocated into two primary phases: exploration and exploitation. The amount of the prey's escape energy determines when to switch between the two phases. Below is a description of the original Harris Hawks optimisation algorithm.

#### 3.3.1 Exploration Phase

The position data of the Harris hawk population largely determines the global search phase, and its update methodology is as follows:

$$Z(t+1) = \begin{cases} Z_{rand}(t) - r_1 |Z_{rand}(t) - 2r_2 Z(t)| & q \geq 0.5 \\ (Z_{prey}(t) - Z_m(t)) - r_3 (LB + r_4 (UB - LB)) & q < 0.5 \end{cases} \quad (12)$$

In this case,  $r_1 - r_4$  and  $q$  are casually created among  $(0,1)$ , being transformed every iteration;  $Z(t+1)$  stands for the position of the hawks in the iteration  $t+1$ ;  $Z_{prey}(t)$  signifies the location of the prey;  $Z(t)$  signifies the position of the hawks in the present generation  $t$ ;  $UB$  and  $LB$  are the upper and lower bounds of the population, accordingly;  $Z_{rand}(t)$  indicates a casually nominated hawk from the present population; and  $Z_m(t)$  stands for the mean of individuals in the current population, which comes from Eq.(13):

$$Z_m(t) = \frac{1}{n} \sum_{k=1}^n Z_k(t) \quad (13)$$

where  $Z_k(t)$  represents the location of hawk  $k$  in the reiteration  $t$  as well as  $n$  represents the number of hawks.

### 3.3.2 Conversion from Exploration to Exploitation

The following is the energy calculation that regulates the prey's outflow:

$$E_f = 2E_{esc0}(1 - t/T) \quad (14)$$

Where  $T$  is the maximum amount of repetitions,  $t$  is the number of repetitions that are currently in progress, and the rate of  $E_{esc0}$  is an arbitrary number between  $-1$  and  $1$  that represents the energy's starting state. The global exploration phase is represented by the Harris hawks searching for the prey in various places while the escape energy  $|E_{esc}| \geq 1$ , and the local exploitation phase is represented by the Harris hawks searching the nearby solutions when  $|E_{esc}| < 1$ .

### 3.3.3 Exploitation Phase

Based on the findings of the previous stages' exploration, the Harris hawk will besiege the intended prey in this phase while it attempts to get away from the chase. For this stage of the simulation, four potential ways are suggested based on the actions of the Harris's hawk and its prey.  $E_{esc}$  simulates both the hard and soft besiege of Harris's hawk. The parameter  $r$  indicates whether or not the prey successfully escapes.

### 3.3.4 Soft Besiege

The prey attempts to get away from the hunt when  $|E_{esc}| \geq 0.5$  and  $r \geq 0.5$ , at which point the Harris hawk uses a soft besiege to slowly deplete the prey's energy. The following is a model of the behaviour:

$$Z(t + 1) = \Delta Z(t) - E_{esc}|IZ_{prey}(t) - Z(t)| \quad (15)$$

$$\Delta Z(t) = Z_{prey}(t) - Z(t) \quad (16)$$

where  $I$  is added to resemble the movement of the prey and  $r_5$  is a randomly generated number between  $0$  and  $1$ , with a random variation in its value for each iteration.

### 3.3.5 Hard Besiege

The Harris hawk uses Equation (17) to update its current position and launches a hard besiege attack on prey when it has insufficient energy to get away, specifically when  $|E_{esc}| < 0.5$  and  $r > 0.5$ .

$$Z(t + 1) = Z_{prey}(t) - E_{esc}|\Delta Z(t)| \quad (17)$$

### 3.3.6 Soft Besiege Using Increasingly Fast Dives

The prey has sufficient energy to outflow the hunt when  $|E_{esc}| > 0.5$  and  $r < 0.5$ . Harris hawks will modify their positions by Equation (18):

$$X = Z_{prey}(t) - E_{esc}|IZ_{prey}(t) - Z(t)| \quad (18)$$

$$Y = X + S \times LF(D) \quad (19)$$

where  $D$  is the problem dimension,  $S$  is a random vector of size  $1 \times D$ , and  $LF$  is the levy flight function, which can be defined as in Eqn (20):

$$\left\{ \begin{array}{l} LF(z) = 0.01 \times \frac{u \times \sigma}{|v|^\beta} \\ \sigma = \left( \frac{\Gamma(1+\beta) \times \sin(\frac{\pi\beta}{2})}{\Gamma(\frac{1+\beta}{2}) \times \beta \times 2^{\frac{\beta-1}{2}}} \right)^{\frac{1}{\beta}} \end{array} \right. \quad (20)$$

where  $\beta$  is the default constant, set at  $1.5$ , and  $u$  and  $v$  are random numbers within the interval  $(0,1)$ . Therefore, Equation (21) can be used to carry out the last plan for apprising the Hawks' positions over the soft siege phase:

$$Z(t + 1) = \begin{cases} X, & \text{if } F(X) < F(Z(t)) \\ Y, & \text{if } F(Y) < F(Z(t)) \end{cases} \quad (21)$$

where  $X$  and  $Y$  are gained using Equations (18) and (19), correspondingly.

### 3.3.7 Hard Besiege Using Increasingly Fast Dives

The prey lacks the energy to accomplish an outflow when  $|E_{esc}| < 0.5$  and  $r < 0.5$ . In these cases, the following tactic is intended to be used:

$$Z(t+1) = \begin{cases} X, & \text{if } F(X) < F(Z(t)) \\ Y, & \text{if } F(Y) < F(Z(t)) \end{cases} \quad (22)$$

$$X = Z_{prey}(t) - E_{esc} |Z_{prey}(t) - Z_m(t)| \quad (23)$$

$$Y = X + S \times LF(D) \quad (24)$$

where  $Z_m(t)$  is gained using Equation (13)

### 3.3.8 Main Steps of HHO

Algorithm 1 depicts the important phases of the overall HHO algorithm.

---

#### Algorithm 1 Principal phases of the HHO process

---

Input: Population size  $N$  and the maximum number of iterations  $T$

- 1: Start the population
  - 2: while  $t < T$  do
  - 3: Compute the suitability of every solution and get the best individual
  - 4: for  $i = 1: N$  do
  - 5: Modify the escape energy  $E_{esc}$  by Eq. (14)
  - 6: if  $|E_{esc}| \geq 1$  then
  - 7: Modify the position by Eq. (12)
  - 8: else if then  $|E_{esc}| < 1$
  - 9: if  $|E_{esc}| \geq 0.5$  and  $r \geq 0.5$  then
  - 10: Modify the position by Eq. (15)
  - 11: else if then  $|E_{esc}| < 0.5$  and  $r \geq 0.5$
  - 12: Modify the position by Eq. (17)
  - 13: else if then  $|E_{esc}| \geq 0.5$  and  $r < 0.5$
  - 14: Modify the position by Eq. (21)
  - 15: else if then  $|E_{esc}| < 0.5$  and  $r < 0.5$
  - 16: Modify the position by Eq. (22)
  - 17: end if
  - 18: end if
  - 19: end for
  - 20:  $t = t + 1$
  - 21: end while
  - 22: return  $Z_{prey}$
- 

### 3.4 Enhanced Multi-Tactic Harris Hawks Optimization (MTHHO)

The HHO process is effective for local development since it includes several development modes and alternates between them, but it also has the drawback of being vulnerable to the local optimum issue. To address this shortcoming, four enhancement techniques are employed in this article that enhance the initial algorithm. The MTHHO algorithm is a new approach to task scheduling optimization that overcomes the

limitations of traditional HHO. It uses sobol sequences to initialize the population, ensuring uniform and comprehensive coverage of the search space. This diversity is crucial for exploring a wide range of potential solutions and avoiding premature convergence. Elite opposition-based learning improves the algorithm's ability to escape local optima by considering both current best solutions and their opposites, maintaining a balance between exploration and exploitation. Additionally, the algorithm includes refined energy updating mechanisms that better simulate the energy dynamics of hawks in nature, enhancing convergence accuracy and efficiency. This algorithm also employs Gaussian Walk Learning (GWL) to enhance population diversity and avoid local optima by initially using larger disruptions that rapidly decrease in later stages, thereby balancing algorithm creation and searchability.

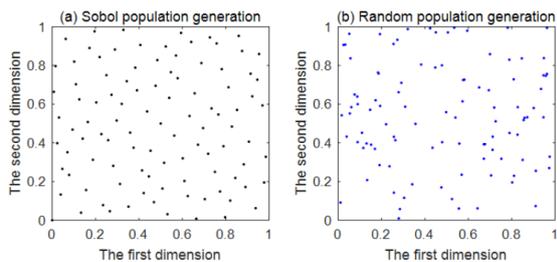
#### 3.4.1 Populations for Sobol Sequence Initialization

The speed and precision of the intelligent algorithm's convergence are significantly influenced by the primitive solution's distribution within the solution space. Randomization is used in the basic HHO method to create the initialised population. Nevertheless, the individuals produced in this manner are not uniformly dispersed around the exploration space, which consequently impacts the algorithm's accuracy and convergence rate. In contrast to the random sequence, the even distribution of points in space is a characteristic of the probabilistic low-difference Sobol sequence. The real population produced by the Sobol sequence can be given as follows:

$$Z_i = Lb + S_n \times (Ub - Lb) \quad (25)$$

where  $S_n$  is the random number produced by the Sobol sequence, where  $S_n \in [0, 1]$ , and  $Lb$  and  $Ub$  are the exploration space's lower and upper limits, accordingly.

The original population space distribution is compared between the random initialization and the Sobol sequence initialization population spaces in Fig 2, assuming that the population size is 100, the search space is two-dimensional, and the upper and lower limits are 1 and 0, correspondingly.



**Fig 2.** Comparison of Sobol and random population generation

The real population produced by the Sobol sequence is further evenly distributed, as seen in Fig 2, which allows the optimization process to conduct a superior global exploration in the exploration space. This increases the population's diversity and accelerates the algorithm's rate of convergence.

### 3.4.2 Elite Opposition-Based Learning

Opposition-based learning (OBL) is a successful method of intelligent computing developed by Tizhoosh in 2005. This technique has been used recently to enhance several algorithms and has shown excellent optimization outcomes. Considering an instance where a feasible response in  $d$ -dimensional search space is  $Z = (z_1, z_2, \dots, z_d) (z_j \in [a_j, b_j])$ , then the definition of its opposition-based solution is  $\bar{Z} = (\bar{z}_1, \bar{z}_2, \dots, \bar{z}_d)$ , where  $\bar{z}_j = r(a_j + b_j) - z_j$ ,  $r$  is the uniform distribution coefficient between  $[0, 1]$ .

The inverse solution developed by the opposition-based learning technique does not always search for the global optimal solution simpler than the existing exploration space. Elite opposition-based learning (EOBL) is suggested as a solution to this issue. Considering that the elite individual represents the extreme of the current population in the search space  $Z_e = (z_1^e, z_2^e, \dots, z_d^e)$ , then its inverse solution  $\bar{Z}_e = (\bar{z}_1^e, \bar{z}_2^e, \dots, \bar{z}_d^e)$  can be stated as follows:

$$\bar{z}_j^e = k \cdot (a_j + b_j) - z_j^e \quad (26)$$

where  $a_j = \min(z_j^e)$ ,  $b_j = \max(z_j^e)$ ,  $k$  is a random number inside  $[0, 1]$ , and  $z_j^e \in [a_j, b_j]$ . It also has a dynamic border with upper and lower limits, respectively, represented by  $b_j$  and  $a_j$ . It is advantageous for the produced inverse solution to slowly decrease the search space and accelerate the algorithm's convergence by substituting a dynamic boundary for the fixed boundary. The method used to reset the value

is as follows to prevent the elite inverse solution from jumping beyond the boundary and losing its viability:

$$\bar{z}_i^e = \text{rand}(a_j, b_j) \quad (27)$$

### 3.4.3 Optimisation of Escape Energy Update

The energy factor  $E_{esc}$  is used by a Harris hawk in the basic HHO to control the algorithm's shift from the global search stage to the local search stage. Nevertheless, as Eq. (14) illustrates, a linear update is used to lower its energy factor  $E_{esc}$  from 2 to 1, this, in the latter part of the cycle, results in locking it in a local optimum. When the process advances to the latter stage, a novel, upgraded form of the energy factor is utilized to address the drawback of just local searching:

$$E_{esc} = \begin{cases} \cos(\pi \times (t/T + 1/2) + 2), & t \leq T/2 \\ \cos(\pi \times (t/T - 1/2)^{1/3}), & t > T/2 \end{cases} \quad (28)$$

$$E_{esc1} = E_{esc} \times (2 \times \text{rand} - 1) \quad (29)$$

where  $r$  is the random number inside  $[0, 1]$ ,  $T$  is the maximum number of iterations, and  $t$  is the number of iterations that are currently being done.

As shown in Fig 3, the algorithm's global search capability is controlled by a fast deceleration rate early in the iteration. The lowering rate reduces down in the middle of the iteration to equilibrium the capabilities of local exploitation and global exploration. Later in the iteration, the local search picks up speed and its value decreases quickly. Fig 4 shows that  $E_{esc1}$  has changing energy parameters all over the recursive procedure and has the capability of both global and local searching, with global exploration taking place primarily in the initial phase and more local exploitation taking place later while still maintaining the potential of global exploration.

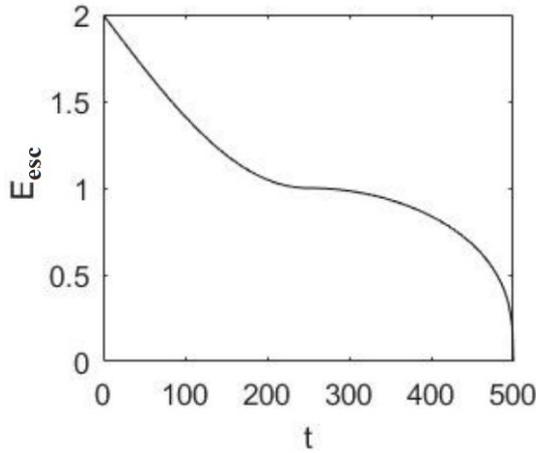


Fig 3. Recurrent alteration graph of E\_esc.

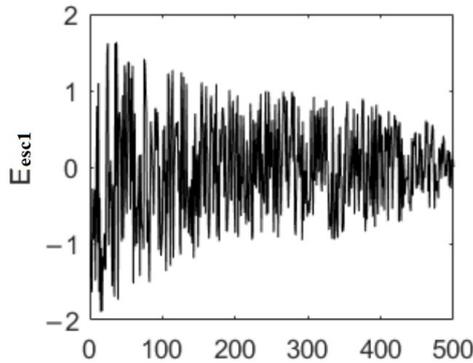


Fig 4. Recurrent alteration graph of E\_esc1.

### 3.4.4 Gaussian Walk Learning

A traditional stochastic walk technique with good exploitation potential is Gaussian walk learning (GWL). Therefore, this research employs this method to modify population members to improve the population's diversity and assist it in escaping the local optimum trap. Equation (30) illustrates the Gaussian walk learning model:

$$Z(t+1) = \text{Gauss}(Z(t), \tau) \quad (30)$$

$$\tau = \cos(\pi/2 \times (t/T)^2) \times (Z(t) - Z_r(t)) \quad (31)$$

where the unidentified individual's position among the generation population  $t$  is shown by  $Z(t)$ , the individual in the generation population  $t$  is showed by  $Z_r(t)$ , and the Gaussian distribution with  $Z(t)$  as the expectation and  $\tau$  as the standard deviation is represented by  $\text{Gauss}(Z(t), \tau)$ . The function  $\cos(\pi/2 \times (t/T)^2)$  modifies the stage dimension of

Gaussian walk learning. Fig 5 illustrates this in image form. To improve the creation of algorithms and balance searchability, the disruption used during the beginning stages is larger and rapidly decreases in the latter phases.

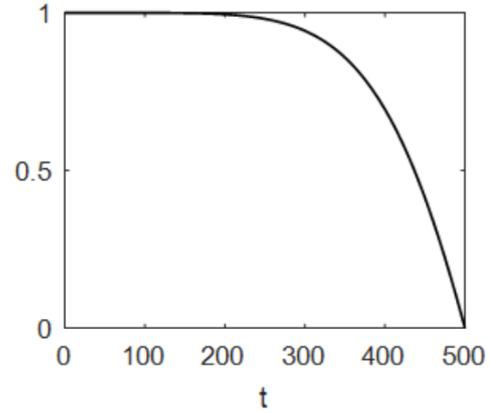


Fig 5. Wandering length of steps modification graph

### 3.4.5 MTHHO Algorithm

In overview, Algorithm 2 depicts the major stages of the enhanced MTHHO algorithm.

---

#### Algorithm 2 Principal phases of the MTHHO process

---

Input: Population size  $N$  and the maximum number of iterations  $T$

- 1: Initialise the population by Eq. (25)
- 2: while  $t < T$  do
- 3: Determine the fitness of the original population and the people in its reverse population by generating the reverse population utilising the elite opposition-based learning process
- 4: If the process is not moving forward, then
- 5: Modify the location by Eq. (30).
- 6: else
- 7: for  $i = 1: N$  do
- 8: Modify the escape energy  $E_{esc}$  by Eq. (29)
- 9: if  $|E_{esc}| \geq 1$  then
- 10: Modify the position by Eq. (12)
- 11: else if then  $|E_{esc}| < 1$
- 12: if  $|E_{esc}| \geq 0.5$  and  $r \geq 0.5$  then
- 13: Modify the position by Eq. (15)
- 14: else if then  $|E_{esc}| < 0.5$  and  $r \geq 0.5$
- 15: Modify the position by Eq. (17)
- 16: else if then  $|E_{esc}| \geq 0.5$  and  $r < 0.5$
- 17: Modify the position by Eq. (21)
- 18: else if then  $|E_{esc}| < 0.5$  and  $r < 0.5$
- 19: Modify the position by Eq. (22)
- 20: end if

```

21: end if
22: end for
23: end if
24:  $t = t + 1$ 
25: end while
26: return  $Z_{prey}$ 

```

---

### 3.5 Proposed Scheduling Algorithm

The suggested approach aims to balance the impact on resources and minimise makespan and energy consumption by focusing on task dependencies and execution times. To maintain a balanced load on the resources, VMs are assigned tasks based on the requirements of the workflow tasks. Task assignment is a crucial aspect of a system, focusing on reducing bottlenecks and improving overall efficiency. It involves assigning tasks based on their dependencies and execution times, with higher-priority tasks given priority. Tasks are categorized based on thresholds for task levels and durations, ensuring timely handling of urgent tasks. The dynamic assignment process adjusts based on workload and resource availability, ensuring efficient use of virtual machines (VMs). The suggested method scans the workflow tasks and establishes thresholds for the tasks' depth and length, where depth and length are related to the tasks' levels and execution times, respectively. During task execution, the threshold standards are utilized to handle tasks based on various significances. Longer implementation times and bottlenecks in the system are caused by tasks with additional dependents. Likewise, to shorten the total execution time, tasks that take longer to complete must be prioritised. High-priority jobs are handled by allocating VMs with significant processing power to them. To prevent needless waiting for jobs at the same level, the suggested algorithm further looks for tasks with lengthy implementation periods and gives these tasks significance processing. Based on the input data, thresholds are established for the quantity of dependents as well as the duration. By using resources more effectively, these processes shorten the implementation period, which also results in lower energy consumption. Algorithm 3 illustrates the stages in the process.

---

#### Algorithm 3: Prevent bottleneck tasks

Input: workflow  $w$   
Output: Task queues according to length and depth  
Allocate thresholds  $dt$  for the depth of tasks and  $lt$  for the length of tasks  
for every task  $t$  in the task set do

```

depth = number of levels reliant on  $t$ 
length = execution time of  $t$ 
if  $depth \geq dt$  then
    transfer  $t$  to the depth queue
end
if  $length \geq lt$  then
    transfer  $t$  to length queue
end
end
Return queues

```

---

Subsequently, the suggested algorithm employs queues based on task intensities; that is, distinct queues are kept for activities requiring a lot of CPU power and tasks with several dependencies. It takes less time to discover the right VMs during the VM allocation process when workloads with varying intensities are placed in distinct queues. The allocation process is streamlined by using queues to organise tasks, sorting and prioritising them before being assigned to VMs, reducing computational complexity and improving overall system efficiency. Tasks are stored in queues in which task queues are organised based on intensities, arrival times, and deadlines, with different queues maintained for tasks of varying intensities. Each task in the queue is associated with additional metadata, such as its intensity, arrival time, and deadline, which optimises the scheduling process. The last stage is to build appropriate VMs for the jobs after grouping them into distinct queues. Historical Scheduling Logs (HSLs) were used to inform the queue management process, providing insights into past task execution patterns. If no matching records exist, new VMs are created to handle the tasks, allowing the system to adapt to new or unexpected workloads. A novel VM is built using the resources essential to finish the task, and the HSLs are modernised if there isn't a matching record in the HSLs. Algorithms 4-6 illustrate the steps in the suggested algorithm.

---

#### Algorithm 4: Construct the types of VMs

Input: HSLs, task sets (from certain queues),

Output: VMs (Types)

$N$  = number of tasks in a queue

for (every task  $t$  in  $N$ ) do  
calculate  $P(T_t)$  (Eq. 32)

$H = \text{best } n P(T_t)$

end

for (every  $l$  in  $H$ ) do

if ( $t$  identified in HSLs) then

assign  $VM_j$  based on the type of  $t$

else

Construct  $VM()$

end

```

    Allocate  $VM_j$  to task  $t$ 
  end
end
end

```

---

**Algorithm 5: Construct Virtual Machines**


---

Input: Set of tasks  $L_t$  from queues from algorithm 3,  
set of hosts  $L_h$   
Output: Set of VMs

```

for (every host  $h$  in  $L_h$ ) do
   $u$  = resource consumption of  $h$ 
  if (host resources are accessible) then
    Construct VM
    Modernise HSLs
  end
else
  Turn on the new host
  Construct VM
  Modernise HSL
end
end
end

```

---

**Algorithm 6: Task Scheduling**


---

Input: Set of tasks  $L_t$  from queues from algorithm 3,  
set of VMs  $V_l$   
Output: Schedule of tasks for VMs

```

for (every task  $t$  in  $L_t$ ) do
  Sort the tasks into categories such as memory-in-
  tensive, CPU-intensive, or both.
  Set every category in a different queue.
   $V_l = VMtypes()$ 
  for (each VM  $v$  in  $V_l$ ) do
    Compute the same degree of  $v$  and  $t$ 
    if (the required condition is met) then
      schedule  $t$  to  $v$ 
    else
       $V_n = CreateVM()$ 
      schedule  $t$  to  $V_n$ 
      modernise HSLs
    end
  end
end
end
end

```

The tasks are noted in the first phase, which means that suitable VMs for the tasks are located. The VM types are decided upon and the task types are categorised appropriately. Let  $T_l$  be a type  $l$  task for a set of tasks  $T$  that were taken from the historical data. Equation (32) can be used to calculate the ratio  $P$ .

$$P = \frac{|T_l|}{|T|} \quad (32)$$

Let  $T_i^r$  be a candidate task, and let  $V_j^r$  be a VM with  $r = \{1,2,3,4\}$  denoting the CPU, memory, bandwidth, and storage capacity of the VM, respectively.

And let  $r = \{1,2,3,4\}$  be the task's requirements. Equation (33), when applied to task  $T_i^r$  and VM  $V_j^r$ , yields the matching degree.

$$P(T_i^r|V_j^r) = \begin{cases} (V_j^r/T_i^r)^2, & \text{if } T_i^r > V_j^r \\ V_{max}^r - V_j^r + T_i^r/V_{max}^r, & \text{otherwise} \end{cases}, \quad (33)$$

where  $k$  denotes the type of VM and  $V_{max}^r = k \in U^{max} V_k^r$ . Equation (34) can be used to determine the likelihood that a job  $T_j$  is of type  $Y_j$ .

$$P(Y_j|T_i) = \pi_{r=1}^4 P(T_i^r|V_j^r) \quad (34)$$

The suggested algorithm optimises the scheduling problem by utilising MTHHO. It is a difficult and complex task to map resources to tasks when there are several objectives. In this case, the dimension of the search field is calculated by the total amount of tasks in the process. The search space's size is adjusted to match the quantity of tasks. The dimensions' values range from 1 to the amount of VMs, depending on that number. The plotting of tasks to VMs in this research is represented by the symbols from earlier research i.e.,  $x_t^i = (x_t^{i1}, x_t^{i2}, \dots, x_t^{ij})$ , where  $x_t^{ij}$  represents that, at time  $t$ , the  $j^{th}$  place of a particle is allocated to  $VM_i$ . The size of the search field is represented by the total amount of tasks within the process. The velocity is signified by  $v_t^i = (v_t^{i1}, v_t^{i2}, \dots, v_t^{ij})$ , where  $v_t^{ij}$  represents the velocity, which represents that, at time  $t$ ,  $VM_i$  transfers to the  $j^{th}$  place of a particle with velocity  $v$ . The method finds non-dominated solutions in subsequent iterations. The archive contains these solutions. We refer to these alternatives as workable solutions. Given that the processes identify non-dominated results, solutions are kept initially, and the record is empty. Only when the novel results outweigh the existing ones are they uploaded to the archive. Using the fitness function, the solutions' dominance is computed. Lastly, the archive only includes what are known as non-dominated, or viable, solutions.

## 4. Result And Discussion

This sector included the outcomes, performance measures, and comparative analysis of the suggested technique. The py-sim tool was used to implement the suggested heuristic-based enhanced Multi-Tactic Harris Hawks Optimization (MTHHO) algorithm on a 64-

bit Windows 10 Pro computer equipped with an Intel Core™ i7-5500U CPU running at 2.40 GHz and 8 GB of RAM. The proposed method assumes that there will be 16 virtual machines (VMs) and 150 tasks in the cloud. The following factors led to the selection of the task and VM.

- **Efficiency of Resources:** The optimised resource allocation made possible by the assumption of 150 jobs and 16 resources (VMs) ensures that every task has enough processing power without wasting any of it.
- **Predictable Performance:** A steady task and resource count makes it easier to forecast and maintain the system's performance, which leads to reliable and strong cloud computing services.
- **Simplified Management:** Using a preset set of tasks and resources simplifies system management and makes it easier to scale resources dynamically in response to shifting workloads.

To determine which model receives a higher assessment score, tests will be conducted on both the proposed model and each experimented model. Our suggested method has been compared to several optimisation and hybrid methods to evaluate its possibility. Several configurations of the models have been tried to achieve the utmost basic TWT, TFT, cost, energy efficiency, and resource utilisation. Numerous examinations have been completed with the most comforting scheduling computations by this effort. To compare our model with other hybrid algorithms and outperform the communication scheduling issue in the cloud, this research has used enhanced optimisation. As a result, the suggested model has been improved. In this examination, a diversity of tasks and virtual machines were employed. When scheduling, every model demonstrates its capabilities.

#### 4.1 Metrics and Parameters

The computational metrics listed beneath are utilised in this study to validate the outcomes of the suggested methodologies with other models.

**Total Waiting Time, or TWT:** This is a criterion that the user wants. When multiple resources vie for a single resource, it is the wait time for job execution. This is the amount of time that is spent waiting for an errand or cycle to finish its queue.

**Total Finish Time, or TFT:** This is a criterion that the user wants. It is the amount of time that passes according to plan from the start of an assignment until its completion. This is the point at which a task reaches its completion of execution.

Resource utilisation is a desired criterion as stated by the service provider. A further metric that shows the amplification of assets employed is resource utilisation. Although providers must use a certain number of resources to achieve maximum profit, resource utilisation should be high in the scheduling framework. One of the key implications in task scheduling is this parameter. There will be constant use of the resource. Energy efficiency and throughput are also very important; nevertheless, resource utilisation is another important barrier to task execution.

The amount of work that a process completes in a given amount of time is called throughput. In other words, throughput is the number of cycles over jobs finished in a given amount of period. The schedule ought to aim to increase the number of tasks completed in each time interval.

**Energy efficiency:** The amount of power used to process each client's request is known as energy consumption. A significant reduction in power consumption is required to achieve energy efficiency. This is one of the most important things to think about while trying to create an improved environment.

The suggested heuristic-based enhanced MTHHO method includes an initial evaluation of fitness values within the first population as shown in Fig 6. The best fitness values across 100 iterations are determined by the method using a round-robin technique based on computing time. The suggested method arrived at the lower fitness values of 33.338s, which decreased the makespan, by comparing the best fitness value with the previous fitness value in each iteration.



Fig 6. The proposed method Output for Best Fit algorithm

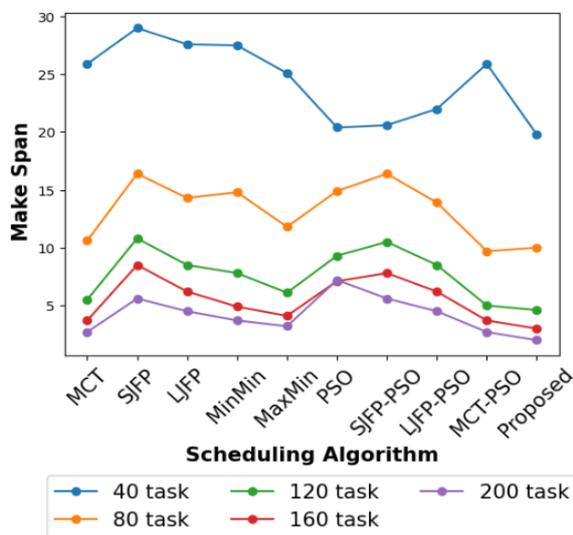
In this situation, the number of tasks ( $N_t$ ) is kept constant at 500. Nonetheless, there is a 40-step variation in the number of virtual machines ( $N_v$ ) between

40 and 200 VMs. The algorithms' relative performances are compared in terms of Makespan in Fig 7. As the number of machines increases in Fig 7, it is projected that the Makespan will decrease. For all scenarios from Nt= 40 to 200, the suggested approach

outperforms all other algorithms. Additionally, it is evident that SJFP consistently has the highest MS values. The numerical results for Makespan are listed in Table 8 correspondingly.

**Table 1.** Makespan (MS) comparison of the proposed method with the prior method

Nt	Algorithm									
	MCT	SJFP	LJFP	MinMin	MaxMin	PSO	SJFP-PSO	LJFP-PSO	MCT-PSO	Proposed
40	25.9	29.0	27.6	27.5	25.1	20.4	20.6	22.0	25.9	19.8
80	10.6	16.4	14.3	14.8	11.8	14.9	16.4	13.9	9.7	9.6
120	5.5	10.8	8.5	7.8	6.1	9.3	10.5	8.5	4.9	4.5
160	3.7	8.5	6.2	4.9	4.1	7.1	7.8	6.2	3.7	2.8
200	2.7	5.6	4.5	3.7	3.2	7.2	5.6	4.5	2.7	2.4



**Fig 7.** Performance evaluation among processes using makespan (MS).

To contrast the proposed heuristic-based enhanced MTHHO method's QoS performance metrics with those of earlier techniques such as round robin (RR), PSO, first come first serve (FCFS), genetic simulated annealing (GASA), and shortest job first (SJF), HGA (Hybrid Genetic Algorithm).

**Table 2.** Total outcomes across various approaches.

Parameters	SJF	FCFS	RR	PSO	GASA	HGA	Proposed
Total execution time	55.36	54.68	54.31	40.21	36.22	32.57	16.75
Total finish time	101.67	100.18	99.31	80.10	79.4	76.6	03.15
Throughput	0.72	0.73	0.74	1.01	0.99	1.21	2.4
Resource utilization	0.42	0.42	0.4	0.61	0.63	0.69	0.95
Energy efficiency	0.60	0.62	0.55	0.35	—	0.30	0.20

Fig 8. shows the relationship between all employed strategies and the TFT and TET, which are some of the validation criteria used to verify the efficacy of the suggested technique. These are carefully taken into account when making plans to increase QoS. The results show that we may attain the highest level of resource utilisation. Every activity in RR receives

the same amount of time, but as the findings show, there are several situations in which an average waiting time could be problematic. Similar data was used to assess the result and examine how the calculation was presented. Despite its speed advantage, the conventional method has the longest waiting period next streamlining; the other AI models that were tested

were likewise effective, but not as effective as the suggested model. Consequently, our suggested model outperforms all other methods, which is our benchmark. Furthermore, in comparison to other strategies, the suggested method offers the shortest execution time, resulting in a faster task performance. So that users can avoid task terminations, the waiting time should be as short as possible. This can determine whether a task is reasonable and the best method to employ when arranging a cloud-based scheduling procedure. On the other hand, while some models have a longer completion time, our model outperforms others in terms of finishing period.

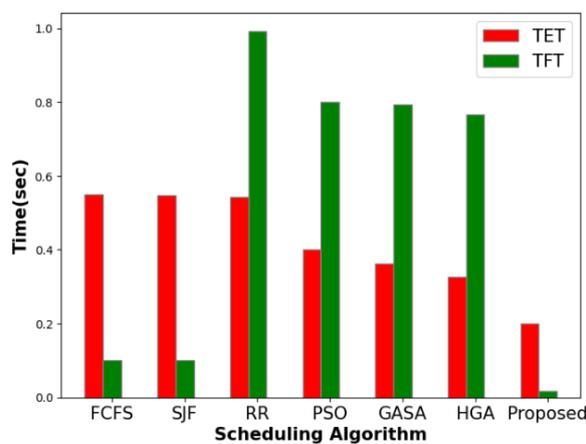


Fig 8. TET, and TFT of the scheduling model.

The suggested model with the highest throughput is the best technique, as demonstrated by the throughput relationship between each approach and Fig 9. Following a sequence of numerous tasks, efforts were prepared to increase the throughput. One of the most important factors in demonstrating the existence of a cycle for every time unit will undoubtedly be throughput. The throughput result demonstrates the effectiveness of the suggested model. To represent the exhibition, each assignment was broken into ten parts. The suggested method executes superior to further methods in this scenario during the split. Despite being linearly separable, the optimisation strategies demonstrated their effectiveness. In any case, the suggested model was the superior one. It may result that the suggested method outperforms previous methods and meets this depiction well because the duration is the largest number of tasks that can be finished per time unit as shown in Fig 10.

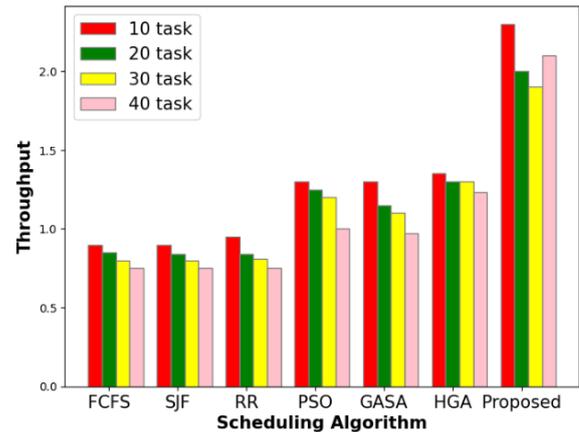


Fig 9. Result of the throughput for different tasks.

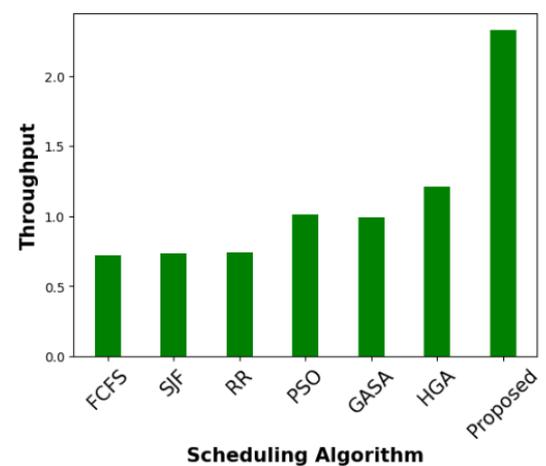


Fig 10. Comparison of the throughput

Fig 11. displays how the scheduling strategies' resource utilisation relates to each other. Additionally, the suggested model selects a different request and makes use of the resources that are available during runtime. In comparison to alternative methods, the suggested computation reduces the inactive waiting time in this way. Similarly, asset utilisation is enhanced independently. However, some assets can become excellent when they can be used in combination with others. The resource is examined under several different makespan totals. By increasing the amount of resources used, the approaches maintain their regular state. The typical waiting time often increases with resource size or the number of task increments. It follows that, in comparison to the other contrasting strategies, the suggested model is the most effective. The effectiveness of different techniques in comparison to the suggested technique may be inferred from the Fig The normal asset utilised by diverse procedures is practically equivalent, indicating that the number of available resources has an impact on it.

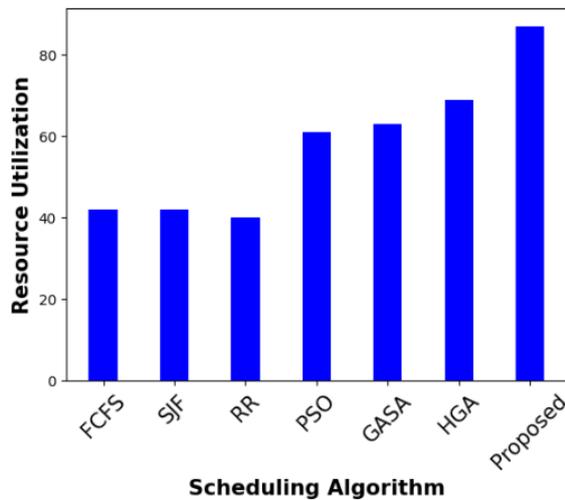


Fig 11. Scheduling models vs resource utilization.

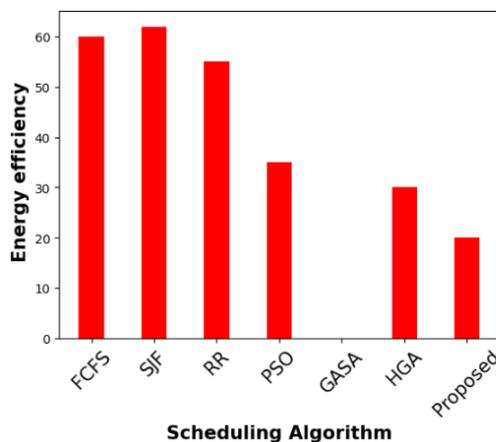


Fig 12. Efficiency of energy consumption comparison with the proposed model.

The efficiency of energy usage is shown in Fig 12. To fully examine its efficiency, the energy was first computed in KWh and then converted to a percentage. The parameter's goal is to lower energy consumption. The suggested approach beats further methods using a 20% lower amount of energy consumption, as the Fig illustrates. The suggested model's efficiency was demonstrated by a comparison with alternative metaheuristics. In comparison, the PSO and GA appeared quite equitable; nonetheless, the model's efficiency in energy usage is still lacking. Being one of the important criteria, this one will improve machine performance while also promoting environmental sustainability. The standard conventional procedure that has been tested is renowned for its sufficiency, however, it was unable to outperform the presented model.

## 5. Conclusions

In conclusion, the paper presents a comprehensive approach aimed at optimizing VM allocation in cloud data centres, addressing challenges related to energy efficiency and task dependencies. By employing an enhanced MTHHO algorithm, this approach effectively assigns energy-efficient VMs considering job dependencies and task execution times. The algorithm incorporates various improvements, including enhanced energy updating methods, elite opposition-based learning for flexibility, Sobol sequences for population initialization, and the introduction of Gaussian walk learning to prevent the process from converging to local optima. The results demonstrate significant enhancements in Quality of Service (QoS) performances, showcasing reduced makespan, energy consumption of 0.20, throughput of 2.4, and execution time of 16.75, with resource allocation improvements ranging from 1% to 98% compared to prior methods of cloud computing. Overall, the proposed heuristic-based MTHHO method efficiently balances loads and allocates resources, highlighting its potential for enhancing cloud computing efficiency and performance.

## Reference

- Abd Elaziz, M., Xiong, S., Jayasena, K. P. N., & Li, L. (2019). Task scheduling in cloud computing based on hybrid moth search algorithm and differential evolution. *Knowledge-Based Systems*, 169, 39-52.
- Abdullahi, M., & Ngadi, M. A. (2016). Symbiotic organism search optimization-based task scheduling in a cloud computing environment. *Future Generation Computer Systems*, 56, 640-650.
- Abdullahi, M., Ngadi, M. A., Dishing, S. I., & Abdulhamid, S. I. M. (2023). Adaptive symbiotic organisms search for constrained task scheduling in cloud computing. *Journal of ambient intelligence and humanized computing*, 14(7), 8839-8850.
- Adhikari, M., Nandy, S., & Amgoth, T. (2019). Meta heuristic-based task deployment mechanism for load balancing in IaaS cloud. *Journal of Network and Computer Applications*, 128, 64-77.
- Agarwal, M., & Srivastava, G. M. S. (2021). Opposition-based learning inspired particle swarm

- optimization (OPSO) scheme for task scheduling problems in cloud computing. *Journal of Ambient Intelligence and Humanized Computing*, 12(10), 9855-9875.
- Alam, T. (2021). Cloud-based IoT applications and their roles in smart cities. *Smart Cities*, 4(3), 1196-1219.
- Alsaidy, S. A., Abbood, A. D., & Sahib, M. A. (2022). Heuristic initialization of PSO task scheduling algorithm in cloud computing. *Journal of King Saud University-Computer and Information Sciences*, 34(6), 2370-2382.
- Chaudhary, D., & Kumar, B. (2018). Cloudy GSA for load scheduling in cloud computing. *Applied Soft Computing*, 71, 861-871.
- Cui, D., Peng, Z., Li, Q., He, J., Zheng, L., & Yuan, Y. (2021). A survey on cloud workflow collaborative adaptive scheduling. In *Advances in Computer, Communication and Computational Sciences: Proceedings of IC4S 2019* (pp. 121-129). Springer Singapore.
- Devaraj, A. F. S., Elhoseny, M., Dhanasekaran, S., Lydia, E. L., & Shankar, K. (2020). Hybridization of firefly and improved multi-objective particle swarm optimization algorithm for energy efficient load balancing in cloud computing environments. *Journal of Parallel and Distributed Computing*, 142, 36-45.
- Fanian, F., Bardsiri, V. K., & Shokouhifar, M. (2018). A new task scheduling algorithm using Firefly and simulated annealing algorithms in cloud computing. *International Journal of Advanced Computer Science and Applications*, 9(2).
- Gawali, M. B., & Shinde, S. K. (2018). Task scheduling and resource allocation in cloud computing using a heuristic approach. *Journal of Cloud Computing*, 7(1), 1-16.
- Golchi, M. M., Saracian, S., & Heydari, M. (2019). A hybrid of firefly and improved particle swarm optimization algorithms for load balancing in cloud environments: Performance evaluation. *Computer Networks*, 162, 106860.
- Jena, U. K., Das, P. K., & Kabat, M. R. (2022). Hybridization of a meta-heuristic algorithm for load balancing in a cloud computing environment. *Journal of King Saud University-Computer and Information Sciences*, 34(6), 2332-2342.
- Katal, A., Dahiya, S., & Choudhury, T. (2023). Energy efficiency in cloud computing data centres: a survey on software technologies. *Cluster Computing*, 26(3), 1845-1875.
- Keshanchi, B., Souri, A., & Navimipour, N. J. (2017). An improved genetic algorithm for task scheduling in the cloud environments using the priority queues: formal verification, simulation, and statistical testing. *Journal of Systems and Software*, 124, 1-21.
- Konjaang, J. K., & Xu, L. (2021). Meta-heuristic approaches for effective scheduling in infrastructure as a service cloud: A systematic review. *Journal of Network and Systems Management*, 29, 1-57.
- Kumar, M., & Sharma, S. C. (2018). PSO-COGEN: Cost and energy-efficient scheduling in a cloud environment with deadline constraints. *Sustainable Computing: Informatics and Systems*, 19, 147-164.
- Mahato, D. P., Singh, R. S., Tripathi, A. K., & Maurya, A. K. (2017). On scheduling transactions in a grid processing system considering load through ant colony optimization. *Applied Soft Computing*, 61, 875-891.
- Mansouri, N., Zade, B. M. H., & Javidi, M. M. (2019). Hybrid task scheduling strategy for cloud computing by modified particle swarm optimization and fuzzy theory. *Computers & Industrial Engineering*, 130, 597-633.
- Mishra, S. K., Sahoo, B., & Parida, P. P. (2020). Load balancing in cloud computing: a big picture. *Journal of King Saud University-Computer and Information Sciences*, 32(2), 149-158.
- Ramamoorthy, S., Ravikumar, G., Saravana Balaji, B., Balakrishnan, S., & Venkatachalam, K. (2021). MCAMO: multi-constraint aware multi-objective

resource scheduling optimization technique for cloud infrastructure services. *Journal of Ambient Intelligence and Humanized Computing*, 12, 5909-5916.

Strumberger, I., Bacanin, N., Tuba, M., & Tuba, E. (2019). Resource scheduling in cloud computing based on a hybridized whale optimization algorithm. *Applied Sciences*, 9(22), 4893.

Wei, X. (2020). Task scheduling optimization strategy using improved ant colony optimization algorithm in cloud computing. *Journal of Ambient Intelligence and Humanized Computing*, 1-12.

Yadav, M., & Mishra, A. (2023). An enhanced ordinal optimization with lower scheduling overhead based novel approach for task scheduling in a cloud computing environment. *Journal of Cloud Computing*, 12(1), 8.

# Brain tumor detection using MRI images- a comparative study based on different classifiers

Suvarna Raju Puligurti<sup>1\*</sup>, P. Chitra<sup>2</sup>, A.V. Bharadwaja<sup>3</sup>

<sup>1,2</sup>Department of Electronics and Communication Engineering  
Sathyabama Institute of Science & Technology Chennai, Tamil Nadu 600119, India

<sup>3</sup>Vignan's Institute of Information Technology  
Besides VSEZ Vadlapudi Duvvada, Gajuwaka, Visakhapatnam, Andhra Pradesh 530049

\*Corresponding mail: psrajuwstm@gmail.com

(Received 28 February 2024; Final version received 26 April 2024; Accepted 15 June 2024)

## Abstract

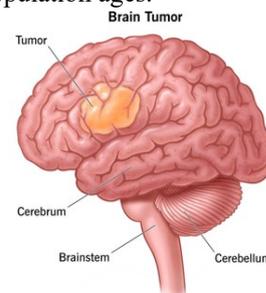
The detection of brain tumors is a major challenge in clinical imaging. Integrating machine learning techniques with MRI (Magnetic Resonance Imaging) analysis has been revealed as a powerful and exciting strategy. This study highlights the importance of early detection and precise diagnosis for medical intervention. Machine learning models extract MRI features like texture, shape, and intensity, and train on labelled datasets. The paper discusses the advantages and challenges of this approach, emphasizing data quality, feature engineering, and model selection. It also highlights the potential for continuous improvement in machine learning models. The synergy between machine learning and MRI imaging holds promise for improved patient outcomes and diagnostic processes. This study compares the techniques of ML, DL and Hybrid Learning. This comparative analysis demonstrates that hybrid Learning performs better in identifying BTs on MRI images. Each system produced superior results. Especially, deep CNN+SVM+RBF combined technique yields best performance, with 98.6% accuracy, 98.2% sensitivity, and 98.9% specificity.

*Keywords:* Brain Tumour (BT), MRI images, Machine Learning, Deep Learning, and Hybrid Machine-Deep Learning.

## 1. Introduction

A BT is a malignant of cells in the brain or major spinal canal. These growths are either benign (non-cancerous) or malignant (cancerous), originating from different cell types within the brain. They can vary in size, location and behavior, have a wide range of effects on the body and brain functions. Tumors are categorized into primary which originate in the brain and secondary metastatic types and secondary tumors resulting from cancer cells spreading to the brain [Ahmad and Choudhury (2022)]. Metastatic, Meningioma, Glioblastoma, and Astrocytoma are the common types of BTs [Aiwale and Ansari (2019)]. Benign tumors are usually slow growing, well-defined, non-invasive. Malignant tumors tend to proliferate, infiltrate nearby tissues can be life-threatening [Al-Zurfi 2019]. BTs, which can afflict both children and adults, can be cancerous or noncancerous. BTs, whether malignant or

not, may affect brain function if they become sufficiently huge to cause damage to nearby tissues [Amin etc., (2021)]. There are different types of BTs. Researchers identified more than 150 different BTs. Primary tumors are classified as glial made up of glial cells in your brain or non-glial formed on or in brain structures such as nerves, blood vessels, and glands and benign or malignant. Glioblastoma is the most deadly type of BT growing increasingly common as the general population ages.



**Fig 1.** Anatomy of Brain

A BT is diagnosed using a variety of tests, including a brain MRI or CT scan, Biopsy, Spinal tap (lumbar puncture), and specialized tests [Anand etc., (2022)]. Magnetic resonance imaging, or MRI, creates images of interior human anatomy. Using strong magnets. Because it shows the brain more clearly than other imaging tests, MRI is frequently used to detect BTs. Some BTs can be detected with a positron emission tomography scan, generally known as a PET scan [Archana and Komarasam (2023)]. A radioactive tracer is injected into a vein during a PET scan which circulates the blood and attaches to BT cells. The tracer identifies the tumor cells visible on the PET machine's images. There are various ideas for detecting brain tumors using MRI images through ML, DL and hybrid learning techniques. Several ML classifiers can be employed to detect brain tumors in medical imaging, including MRI (Magnetic Resonance Imaging) scans. The choice of classifier often depends on task issues and characteristics of the dataset. Here are some common classifiers used in BT detection SVM, LR, RF, K-NN, Gradient Boosting, Neural Networks, Naïve Bayes, Decision Trees, Extreme Learning Machine, and Regularized Extreme Learning Machine. The choice of classifier depends on the specific requirements of the BT detection task, such as the size of the dataset, the complexity of the tumor types, and the need for interposing. In practice, an ensemble of classifiers or a combination of multiple methods may also be employed to enhance overall performance and reliability [Athira etc., (2015)]. However, analyzing the type of BT is a challenging task for researchers. MRI scans involve a combination of techniques and approaches to ensure precise detection, classification, and characterization. There are various key techniques and strategies to analyze for accurate BTs. Image processing, Region of Interest ROI segmentation, Feature extraction, Machine Learning classifiers, Ensemble methods, Transfer Learning, Interpretability and explainability, Data Augmentation, Validation and Cross-Validation, Regularization and Hyperparameter Tuning and Validation on External Datasets. Accurate BT analysis is a complex and evolving field, and it often requires collaboration between medical professionals, data scientists, and researchers to develop and validate robust models. Additionally, ongoing research and advancements in deep learning and medical imaging continue to contribute to improved accuracy in BT detection and diagnosis [Avazzadeh etc., (2023)]. ML has significantly transformed, which enables early

detection of BTs, potentially before clinical symptoms manifest, improving patient outcomes [Brindha etc., (2021)]. It can identify subtle and complex patterns within MRI images that may be challenging for human observers and machine learning models maintain consistent performance, reducing the impact of fatigue or human error. However, it is not without limitations, including issues related to data quality, interpretability, and potential ethical concerns. Dealing with these challenges will be critical as the field evolves in order to maximize the benefits of ML in improving brain tumor diagnosis and patient care. To overcome these issues, researchers used the advanced technique of DL which has revolutionized the field of healthcare imaging, particularly in the detection of BTs using MRI scans [Dash etc., (2019)]. DL models naturally learn essential characteristics from MRI images, reducing the requirement for manual feature extraction, and they demonstrate high accuracy and can detect subtle and complex patterns, improving diagnostic precision. For those reasons, deep learning models are highly scalable and can be applied to a vast number of medical images quickly which accelerates the diagnostic process, which is crucial in medical emergencies. So, deep learning has brought significant advancements to BT detection using MRI, offering high accuracy and automation. However, it comes with challenges related to data requirements, interpretability, and necessity for substantial computational resources. Resolving these difficulties will become more important as the area evolves be crucial harnessing full potential of deep learning in improving BT diagnosis and patient care. Deep learning methods, through research and development, are enhancing BT detection from MRI images. This method offers unprecedented accuracy, efficiency, and adaptability, contributing to earlier diagnoses and improved patient care in the critical field of BT detection [Febrianto etc., (2020)]. However, there is another way of technique to detect BT in MRI images using a Hybrid ML & DL technique accurately. Reasons for approaching the Hybrid technique demonstrate the versatility and adaptability of machine learning techniques in medical imaging, laying the groundwork for more efficient and reliable diagnostic tools in healthcare. A hybrid approach can be more robust and less susceptible to overfitting than deep learning alone [Garg and Garg (2021)]. For example, traditional ML algorithms such as SVM or RF, can help regularize the deep learning model by providing more structured

predictions, preventing the model from memorizing the training data [Gurbină et al (2019)]. Despite advances in medical imaging, correctly diagnosing and characterising brain tumours remains a difficult task. Conventional approaches frequently rely on manual interpretation by radiologists, which can be subjective and time-consuming. Furthermore, these approaches may struggle to detect small or subtle tumours, resulting in delayed diagnoses and poor treatment outcomes. Moreover, existing techniques may struggle with distinguishing benign and malignant tumors or accurately defining tumor boundaries, affecting treatment planning and patient prognosis. The large volume of medical imaging data also presents logistical challenges for timely analysis and interpretation.

In light of these challenges there is need for advanced computational approaches are needed for brain tumor detection and classification. Integrating ML techniques with MRI analysis can improve diagnostic accuracy and efficiency. By extracting discriminative features from MRI images, ML models can aid in early detection and precise characterization of brain tumors. This work addresses these issues by comparing ML, DL, and hybrid learning strategies for brain tumour diagnosis using MRI scans. We look at the benefits and drawbacks of each strategy, with an emphasis on improving diagnostic accuracy and clinical value. Our approach aims to contribute to ongoing efforts to enhance patient outcomes and shorten the diagnostic procedure for brain tumours.

The main presence of our analysis:

This analysis focuses, on an extensive and exhaustive guide to the field of BT Detection in MRI images using various classifiers of ML and DL techniques has been presented by comparing and summarizing the best model to achieve high accuracy.

- The first section of this study evaluates various Machine Learning techniques. This preliminary evaluation revealed that the accuracy of common machine learning strategies in this machine learning is low.

- Recognising the limitations of typical machine learning methods, the investigation shifted its focus to deep learning. The CNN model performed better in dealing with Brain Tumour detection in MRI images.

- To enrich the model's performance to successfully target the need for BT finding using MRI images, emphasize the hybrid Machine and Deep Learning model.

Following the introduction, Section 2 describes the existing authors and their methodology using ML DL and Hybrid techniques to detect BT using MRI images. Section 3 discusses comparative analysis. Section 4 provides an overview of the comparison analysis. Section 5 shows the conclusion of this study work.

## 2. Recent advanced studies based on Oncological Research

Our approach focussed on some studies based on optimizing management strategies, refining mathematical models and exploring molecular mechanisms associated with tumor suppressions. A study by [Györfi etc., (2019)] provided a detailed summary of management methods for optic pathway gliomas. Through the systematic review and meta-analysis of OPG patients from the previous research list, the study provided insights into the efficacy of various treatment modalities, such as surgery, chemotherapy, and radiation therapy. The study's findings help us determine the best management methods for optic pathway gliomas, which ultimately improves patient outcomes.

Glioblastoma is a highly aggressive brain tumour that develops from astrocytic glial cells. It mainly affects older persons but can occur at any age. Symptoms may include nausea, vomiting, migraines, or seizures. Despite intense clinical treatment, which includes surgery, chemotherapy, and radiation therapy, GB is incurable and has a high resistance to therapy. [Hannan etc., (2022)] examined a fractional-order brain tumour model using an operational matrix-based optimisation technique. They expressed the optimal solution in terms of generalised Laguerre polynomials, converting the issue into a system of nonlinear algebraic equations with Lagrange multipliers. The convergence analysis was presented, and numerical examples were used to support theoretical claims and explain biological behaviour patterns.

Several studies have shown that miR-195 is downregulated in colorectal cancer tissues. [Haq etc., (2022)] investigated the effects of exogenously produced mature miR-195-5p on the malignant properties of human colorectal cancer cells. They transfected the Caco-2 and SW480 human colon cancer cell lines with a synthetic miR-195-5p mimic. Exogenous production of miR-195-5p suppressed numerous invasion and angiogenesis mediators in

colorectal cancer cells while increasing the apoptotic cell population in both lines. Furthermore, miR-195 transfection dramatically decreased migration in both cell types. These data indicate that miR-195 plays a significant tumor suppressive role in human colorectal cancer.

These papers highlight contemporary advances in brain tumour detection research, which include clinical management tactics, mathematical modelling approaches, and molecular insights into tumour biology. In the fight against brain tumours, researchers aim to improve early diagnosis, therapeutic efficacy, and patient outcomes by combining multidisciplinary approaches. Future research is expected to enhance patient outcomes and contribute to ongoing oncological research by focusing on optimising management strategies, refining mathematical models, and exploring molecular mechanisms associated with tumour suppression in the context of brain tumours, particularly glioblastoma and optic pathway gliomas, as well as colorectal cancer.

## 2.1 Techniques of Brain Tumor Detection in MRI Images

Different techniques are used to detect BT in MRI images, these techniques leverage advancements in medical imaging, signal processing, and machine learning. In this work, our study analyses the methods for spotting BT.

### 2.1.1 Machine Learning Technique

BT in MRI images using ML is a rapidly evolving field that aims to provide accurate and efficient diagnostic support to healthcare professionals. Machine-learning techniques leverage computational algorithms to analyze intricate patterns and machine-learning MRI scans, facilitating the identification of abnormal tissue indicative of brain tumors. The various ML algorithms can be employed for BT detection.

#### a) *K-Nearest Neighbors*

KNN is a ML technique used in BT detection using MRI images. [Francisco etc., (2021)] use the similarity principle to identify tumor regions within MRI scans. RELM, a regularized version of the Extreme Learning Machine, improves generalization and mitigates overfitting. Combining NGIST, PCA,

and RELM offers a holistic solution for tasks like image classification, object recognition, and scene analysis. This approach ensures accurate, efficient, and robust machine-learning models for complex image-based applications.

#### b) *Naïve Bayes*

[Jahromi etc., (2019)] NB is a methodical approach used in image analysis for medicine that is used for recognizing brain tumors in MRI images. It involves collecting and labelling a dataset of MRI scans, preprocessing them for quality, and extracting features from the images. The Naïve Bayes classifier models conditional probabilities based on these features, simplifying probability calculations. If model fails, fine-tuning or alternative variations may be needed. The model can then be deployed for automated analysis of new MRI images.

#### c) *Logical Regression (LR)*

Logical Regression involves assembling a curated dataset of labelled scans, preprocessing it, and extracting relevant features like texture and intensity statistics. [Jia and Chen (2020)] the model use a logical function to classify the images based on these features. The LR efficacy was assessed using standard evaluation metrics and area under the ROC curve. If necessary, further optimization and fine-tuning can be explored. The trained model can be deployed for real-world brain tumor detection, aiding in early diagnosis and medical decision-making.

#### d) *Decision Tree*

DT analyze scans and makes diagnostic decisions based on image features, classifying anomalies as normal or tumor-affected. These algorithms are trained on labelled MRI datasets, providing a transparent and interpretable process. [Jiang etc., (2023)] Applying Decision Trees to new scans enhances the efficiency and accuracy of BT diagnosis, benefiting patients and medical community.

#### e) *Support Vector Machine*

[Kavitha and Chellamuthu (2013)] SVM is a powerful ML technique to detect BT in MRI data. It examines high-dimensional data, distinguishing tumor-affected regions from normal brain tissue. SVMs accurately classify possible tumors by taking features such as shape, texture and intensity. In neuroimaging, their capacity to handle nonlinear data

and generalize well improves patient outcomes and speeds the diagnostic process.

#### f) Regularized Extreme Learning Machines

RELM is a machine learning method that detects brain tumors in MRI data. To handle high-dimensional data, it combines the strength of ELM with regularization approaches. [KKaur etc., 2020] RELM is capable of processing and learning from MRI features such as intensity, texture, and form, resulting in accurate categorization. Its capacity to manage multidimensional data and generalize effectively provides better patient care and neuroimaging processes in healthcare.

#### g) Random Forest

[Krauze etc., (2022)] Tumors are classified using criteria such as shape, size, and texture. RF can accurately detect normal brain tissue from tumor-affected areas after training on labelled MRI images. This improves patient outcomes by allowing for earlier diagnosis, treatment planning, and patient care.

#### h) Radial Basis Function (RBF)

RBF is a mathematical function that is utilized in ML, signal processing, and function approximation. It

is especially helpful in kernel techniques SVMs and RBF networks. [Kurian etc., (2021)] The Gaussian kernel is a data similarity metric that captures complex, nonlinear relationships. RBFs are highly beneficial for classification and regression applications, particularly when the data displays nonlinear patterns. They are also utilized as activation functions in artificial neural networks.

#### 2.1.1.1. Summary

Machine learning techniques provide significant improvements in brain tumor identification from MRI pictures by enabling quick and automated analysis of MRI scans, learning subtle patterns and changes, and improving tumor detection accuracy. Despite certain constraints, efficiency may vary depending on tumor complexity, and false positives or negatives may occur, necessitating thorough validation, and model success largely relies on the quality and representativeness of training data. Table 1 provides a comprehensive overview of the strengths and limitations of each existing ML technique.

**Table 1.** Strengths and Limitations of Existing ML Methods for BT Detection using MRI Images

Ref No	Model	Strengths	Limitations
[Francisco etc., (2021)]	KNN	It uses the similarity concept to identify tumour regions. - Combining with RELM enhances generalisation and reduces overfitting. - Provides accurate, efficient, and model robustness.	High-dimensional data can have a negative impact on performance because of sensitivity to outliers and noise.
[Jahromi etc., (2019)]	NB	simplifies probability calculations based on conditional probabilities, making it effective for image analysis in medicine.	The assumption of feature independence may not be practical and may result in limited representation power for complex data relationships.
[Jia and Chen (2020)]	LR	The process is transparent and easily understandable, with efficiency evaluated	The performance of data is highly dependent on feature selection and engineering, as it has a limited capacity to capture nonlinear relationships.
[Jiang etc., (2023)]	DT	The transparent and interpretable decision-making process enhances the efficiency and accuracy of brain tumor diagnosis.	The model is susceptible to overfitting, particularly when working with complex datasets, and has limited capacity to accurately capture feature interactions.
[Kavitha and	SVM	The individual adeptly manages high-dimensional data and accurately classifies potential tumors based on their shape, texture, and intensity features.	The choice of the optimal kernel function is crucial for optimal performance, as large datasets may require extensive training time.

Chellamuthu (2013)]			
[KKaur etc., (2020)]	RELM	The system integrates the capabilities of ELM with regularization methods, enabling efficient processing and learning from MRI features.	The performance of a system is significantly influenced by the regularization parameters, and may potentially lead to overfitting if the regularization is not sufficient.
[Krauze etc., (2022)]	RF	Accurately detects tumor-affected areas based on shape, size, and texture, enabling earlier diagnosis and treatment planning.	The ensemble of decision trees may increase computational complexity, and performance may degrade with highly correlated features.
[Kurian etc., (2021)]	RBF	Used in ML, signal processing, and function approximation to capture complex, nonlinear relationships in data.	The choice of kernel parameters is crucial for performance but may struggle with high-dimensional or noisy datasets and has limited interpretability compared to linear models.

### 2.1.2 Deep Learning Technique

DL techniques for detecting BTs in MRI images have emerged as a powerful and successful strategy, employing neural network capabilities to automatically identify and classify tumor locations. DL has various merits in BT detection, including the ability to automatically learn detailed properties, adapt to fluctuations in tumor characteristics, and generalize well to varied patient populations. It indicates a promising path forward in terms of increasing the certainty and efficacy of BT diagnosis using MRI images. There are various DL techniques used to detect brain tumors.

#### a) Convolutional Neural Network

CNNs have revolutionized brain tumor diagnosis in MRI images by learning and extracting detailed patterns. [Lah etc., (2020)] CNNs are used as powerful tools in this process to automatically learn and extract complicated patterns and features indicative of tumor existence. CNNs demonstrate high accuracy in detecting brain tumors and have the potential to speed up the diagnosis procedure. CNNs, as a computer-aided tool, assist radiologists with interpreting complex MRI scans, helping to better patient outcomes.

#### b) Artificial Neural Networks

ANNs are useful in medical processing, especially for detecting BT in MRI images. ANNs are

evaluated the metrics to learn intricate patterns inside complicated datasets. This quantitative assessment aids healthcare practitioners in interpreting MRI scans, resulting in faster and more accurate brain tumor diagnosis. [Lamrani etc., (2022)]. ANNs help to construct computer-aided diagnostic systems, which improves the accuracy of brain tumor diagnosis.

#### c) AdaBoost

AdaBoost, a prominent ensemble learning technique, is being applied to enhance BT detection in MRI images. Integrating weak classifiers into a robust ensemble improves their performance. [Magadza and Viriri (2021)] The procedure begins by obtaining significant features from MRI images, such as intensity, texture, or form, which are then utilized to train weak classifiers. Each one is trained on a subset of data, focusing on certain factors that help identify tumor patterns. This aligns with the growing need in healthcare for robust and accurate diagnostic techniques. [Maqsood etc., (2022)], allowing for quick diagnosis and localization of BTs, permitting timely treatment and improving patient outcomes.

#### d) MobileNet V2

[Masood etc., (2021)] MobileNetV2, a lightweight CNN architecture, is used in health care process, namely in MRI scans to detect brain tumors. Its efficiency makes it suitable for deployment on resource-constrained devices, making it a desirable option for healthcare mobile and edge computing

applications. This approach is consistent with the growing trend of applying deep learning models on mobile and edge devices, allowing for real-time and on-device processing and contributing to more accessible healthcare solutions.

#### e) AlexNet

AlexNet, a deep learning architecture, has demonstrated impressive capabilities for image classification tasks and detecting brain tumors in MRI scans. Its application involves preprocessing MRI images to standardize and enhance their features while learning hierarchical features utilizing several convolutional, pooling, and fully connected layers. This automated technology adds to the progress of medical image processing, assisting healthcare practitioners in making fast and exact diagnoses Sowmiya etc. [McFaline etc., (2018)].

#### f) ResNet-18

ResNet-18, a deep learning model, performs image classification tasks, especially detecting brain tumors in MRI scans. Its architecture, defined by residual blocks, helps in the learning of residual mappings, hence overcoming the vanishing gradient problem. [Noreen etc., (2021)] the deep layers of ResNet-18 allow it to learn hierarchical representations of features, identifying small differences in intensity, shape, and texture that differentiate tumor locations from normal brain tissue. This improves the accuracy of brain tumors, making it a vital tool for physicians.

#### g) DenseNet-41

DenseNet-41 is a DenseNet version that emphasizes dense connectivity across layers.

DenseNet is a form of CNN architecture. DenseNet-41, in particular, has a more complex design with 41 layers, allowing it to detect detailed patterns and characteristics in images [Panesar etc., (2019)]. DenseNet-41 intends to conduct precise brain tumor location, segmentation, and classification.

#### h) Stochastic Gradient Descent (SGD)

SGD is to fine-tune machine learning models, enabling neural networks to identify brain tumors. The process involves assembling a labelled dataset of MRI scans, which is then preprocessed for analysis. [Pooja etc., (2023)] SGD is primarily used in DL, where NN, including CNNs, optimize their internal parameters. The model's performance is assessed using training and testing data, with fine-tuning and model optimization often necessary for accuracy and robustness. This approach is promising for improving early diagnosis and treatment planning in medical imaging.

#### 2.1.2.1. Summary

DL techniques, particularly CNNs, have revolutionized the detection of BTs in MRI images. These algorithms can learn complicated patterns from volumetric MRI data, minimizing the need for manual interpretation and speeding up diagnosis. Table 2 summarizes the strengths and limitations of existing DL methods. They are, susceptible to overfitting, especially when the training dataset is limited. To address this, hybrid DL and ML techniques are proposed, which can overcome these limits while also improving the accuracy of DL methods.

**Table 2.** Strengths and Limitations of Existing DL Methods for BT Detection using MRI Images

Ref No	Technique	Strengths	Limitations
[Lah etc., (2020)]	CNN	Detecting detailed patterns from MRI images, results in high accuracy in detecting brain tumors.	The decision-making process is hindered by high computational complexity and lack of interpretability.
[Lamrani etc., (2022)]	ANN	Effectively learns intricate patterns from complex datasets, aiding in faster and more accurate brain tumor diagnosis.	Overfitting problems occur, may necessitate extensive parameter tuning for optimal performance.
[Magadza and Viriri (2021)]	Adaboost	The integration of weak classifiers into a robust ensemble enhances BT	may experience performance degradation due to its sensitivity to noisy data and outliers.

		detection, enabling quick diagnosis and localization.	
[Maqsood, etc., (2022)]	MobileNet V2	Enabling real-time processing in healthcare applications.	Lower accuracy due to their limited capacity to capture complex spatial relationships in data.
[Masood, etc., (2021)]	AlexNet	Impressive image classification capabilities aid healthcare practitioners in making fast and accurate diagnoses.	Poses a significant threat due to the high computational resources required for training and inference.
[McFaline, etc., (2018)]	ResNet-18	overcomes the vanishing gradient problem with residual connections and learns hierarchical representations of features.	Overfitting, high computational requirements for training and inference.
[Noreen, etc., (2021)]	DenseNet-41	Effective in accurately identifying and classifying brain tumors.	Increased model complexity which may lead to longer training time.
[Panesar, etc., (2019)]	SGD	potential for early diagnosis and effective treatment planning.	Required careful adjustment of learning rates and regularization.

### 2.1.3 Hybrid Deep and Machine Learning

Hybrid approaches combining DL and traditional ML techniques are being explored for BT detection in MRI images. These methods combine predictions from both models, using an ensemble approach to make final predictions and adjust hyperparameters, regularization techniques, and model architectures. Hybrid approaches aim to overcome limitations, capitalize on the interpretability, efficiency, and feature engineering capabilities of traditional methods, and benefit from deep learning's automatic feature learning and representation power.

#### a) Hybrid Deep CNN - Random Forest

[Pooja etc., (2023)]. Present a combined technique for accurate brain MRI tumor segmentation and classification utilizing deep CNN and ML classifiers. The preprocessing step is to study feature map from the brain MRI image space, which is then followed by a more rapid region-based CNN for localization. The second phase involves further refinement using a region proposal network RPN. Deep CNN and SVM-RBF classifiers obtain 98.3% accuracy and a dice similarity coefficient DSC of 97.8% on brain dataset-1 and 98.0% accuracy on the Figshare dataset, based on research results. The model outperforms cutting edge approaches.

#### b) Hybrid ResNet50-SVM

[Ranjbarzadeh etc., (2021)], deep CNN layers are used to extract exclusive properties from MRI pictures of brain tumors. The Gabor filter and ResNet50 are used, with features classified separately and merged. The Kaggle MRI dataset was used, which included 7,023 pictures and four classes. The combined features of Gabor and ResNet50, an advanced hybrid technique, produced the best results, with 95.73% accuracy, 95.90% precision, and 95.72% f1 score.

#### c) Hybrid CNN-SVM

[Saeedi etc., (2023)], the study presents a hybrid model that depends on CNNs for classifying brain tumors. The model improves current datasets with coloured images by utilizing pre-trained Efficientnet and Shufflenet architectures. The model takes features from separate photos, concatenates them, and uses the mRMR feature reduction approach to select the best. The model outperforms prior models with an accuracy of 95.4%.

#### d) Hybrid Bagged Tress-KNN

The study by [Senan etc., (2022)]. Evaluates the performance and simplicity of medical image segmentation approach for identifying brain tumors from MRI scans. They suggested a strategy for improving the accuracy and quality rate of KNN by employing Bagging Ensemble with K-Nearest

Neighbour. The method employs a UNet architecture for picture segmentation and a bagging-based KNN prediction algorithm for classification. The overall classification accuracy achieved 97.7%, outperforming existing approaches.

#### e) Hybrid AlexNet-SVM

[Shaari etc., (2021)]. Investigated brain tumor diagnosis using DL and standard ML approaches. They enhanced MRI images with the average filter technique using AlexNet and ResNet-18 with SVM algorithm. These techniques were discovered using deep features, which were then categorized using SoftMax and SVM. MRI dataset of 3,060 pictures yielded superior results, with AlexNet+SVM hybrid

approach outperforming all others with 95.10% accuracy, 95.25% sensitivity, and 98.50% specificity.

#### 2.1.3.1 Summary

The mixed model combines the features of DL algorithms such as CNN with traditional ML techniques such as SVMs or RFs. This method decreases data dependency, increases performance indicators and improves comprehension of the model's decision-making process. Table 3 provides the advantages and limitations of previous hybrid techniques. This is critical in medical applications to acquire the trust of healthcare experts and boost the model's ability to generalize to a wide range of scenarios.

**Table 3.** Strengths and Limitations of Existing Hybrid Methods for BT Detection using MRI Images

Ref No	Technique	Strengths	Limitations
[Pooja etc., (2023)]	DCNN+RF	Achieving high accuracy and dice similarity coefficient in brain tumor segmentation.	may necessitate extensive hyperparameter tuning for optimal performance.
[Ranjbarzadeh etc., (2021)]	Res-Net50+SVM	Achieves high accuracy, precision, and a fl score in the classification of brain tumors.	The performance of MRI images may be affected by noise or low-quality.
[Saeedi etc., 2023)]	CNN+SVM	Outperforms prior models with improved accuracy.	interpretability is limited due to its complex feature representations.
[Senan etc., (2022)]	Bagged Tree+KNN	Achieves high classification accuracy and quality rate in detecting brain tumors.	The training and inference processes are characterized by high computational complexity.
[Shaari etc., (2021)]	AlexNet+SVM	Aid in timely diagnosis, enhancing patient survival rates, and aiding experts and radiologists in making informed decisions regarding diagnosis and treatment plans.	The extracted deep features have limited interpretability.

## 2.2 Features Extracted From MRI Images

Extracting characteristics in brain tumor identification and classification is critical since it serves as the foundation for developing accurate and dependable diagnostic algorithms. MRI images of the brain provide a plethora of information that can be used to differentiate between tumor and normal brain tissues. Researchers can capture tiny variations in tissue qualities related to tumor pathophysiology by extracting certain aspects from these images, such as texture, shape, intensity, gradient, and frequency domain characteristics. These characteristics provide

important insights into tumor spatial distribution, shape, and composition, allowing for the detection of aberrant tissue regions and exact tumor delineation. Table 4 presents the summary of various studies focusing on MRI image analysis using different classification algorithms and feature extraction methods.

ML algorithms use informative features to learn patterns and relationships in data. It can improve the model's ability to differentiate between tumor and non-tumor regions by selecting features like textural heterogeneity, irregular shape, and abnormal intensity patterns. This leads to more accurate diagnostic outcomes, timely intervention, and improved patient

care. Moreover, the extraction of features in MRI images allows for a more interpretable and clinically relevant analysis, providing valuable insights into tumor morphology like size, volume, and shape irregularity. This information aids in treatment

planning, surgical decision-making, and disease progression monitoring. The extracted features also serve as biomarkers for identifying tumor subtypes, predicting patient outcomes, and assessing treatment response.

**Table 4.** Summary of various feature extraction techniques

Ref No	Classifier	Feature Extraction method	Extracted Features
[Francisco, etc., (2021)]	KNN	Energy, root mean square, correlation	Contrast Images, Texture Variation.
[Jahromi etc., (2019)]	NB	Contrast, homogeneity, correlation, Energy, Entropy	Contrast Images, Texture Variation, grey level, linear relationship between objects.
[Jiang etc., (2023)]	KNN+RF+DT	SWT+PCA+GLCM	The texture of images, contrast, grey level features, smoothness.
[KKaur etc., (2020)]	RELM	PCA+LDA	Texture information along multiple scales of orientations.
[Magadza & Viriri (2021)]	DCNN	7 pre-trained CNN models	High-level features, Spatial information, Texture, invariant.
[Masood etc., (2021)]	AlexNet	CNN	Visual features
[Ranjbarzadeh, etc., (2021)]	ResNet 50+SVM	CNN based model	Intensity, texture, temporal, shape, spatial features.
[Takei etc., (2019)]	LR, NB, KNN, RF, SVM	GLCM+LBP	Grey level and orientation images, contrast, homogeneity, texture.
[Talo etc., (2019)]	SVM, NN, RF, SGD, LR, MLP	2DCNN	Shape, intensity and model-based features.
[Telrandhe etc., (2016)]	SVM+RBF, ANN, AdaBoost	GLCM	Contrast, grey level, texture variations.

These features like texture, shape, intensity, gradient, and frequency domain to distinguish between normal and tumor brain tissues, providing insights into tumor spatial distribution, shape, and composition, enabling precise tumor identification and detection. MRI images reveal features like textural heterogeneity, irregular shape, and abnormal intensity, which enhance diagnostic outcomes and patient care. These insights provide crucial information for treatment planning, surgical decision-making, and disease progression monitoring. It can also serve as biomarkers for identifying tumor subtypes and assessing treatment response.

### 3. Comparative Analysis

Our study analyzes the comparative study of BT Detection using MRI images which emphasizes the strengths and limitations of each approach. Finally, our research Concludes with insights into the most promising directions for further research and application in the context of BT detection using MRI images.

### 3.1 Results obtained using Machine Learning Technique

The use of ML techniques in medical imaging has led to promising results, especially when it comes to using of MRI data for brain tumor detection. This section clearly explains the outcomes of applying ML approaches to detect BT, along with the strengths and limitations of this way. The study made use of a comprehensive dataset that included a variety of MRI images from different sources, including patient demographics and imaging modalities. To ensure consistency and remove any possible biases, like variations in imaging protocols and quality, the dataset had preprocessing.

To evaluate the performance of the ML models, standard metrics such as sensitivity, specificity, accuracy, and area under the receiver operating characteristic curve AUC-ROC were used. These metrics offered an extensive understanding of the models' ability to detect and categorize brain tumors in MRI images.

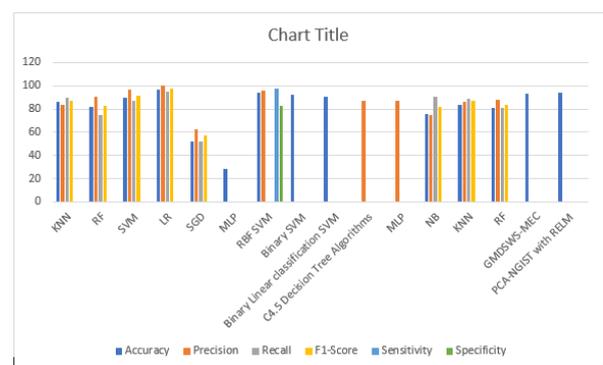
ML models consistently outperformed traditional methods in the remarkable accuracy with which they detected brain tumors. The models demonstrated an interesting level of sensitivity and specificity,

suggesting a strong ability to detect benign and malignant tumors with minimal false positives and false negatives.

**Table 5.** Performance result based on ML Technique

Ref No	Model	Accuracy	Precision	Recall	F1-Score	Sensitivity	Specificity
[Talo etc., 2019]	KNN	86	84	90	87		
[Talo etc., 2019]	RF	82	91	75	83		
[Takei etc., (2019)]	SVM	90	96.4	87	91.5		
[Takei e tal., (2019)]	LR	97	100	95.1	97.5		
[Talo etc., (2019)]	SGD	52	63	52	57		
[Talo etc., (2019)]	MLP	28					
[Pugalenthi etc., (2019)]	RBF SVM	94.3	95.45			97.47	82.54
[Ullah etc., (2023)]	Binary SVM	92					
[Ullah etc., (2023)]	Binary Linear classification SVM	91					
[Verekar & Salkar (2019)]	C4.5 Decision Tree Algorithms		87.5				
[Verekar & Salkar (2019)]	MLP		87.5				
[Takei etc., (2019)]	NB	75.4	74.6	90.3	81.7		
[Takei etc., (2019)]	KNN	83.4	85.9	88.7	87.3		
[Takei etc., (2019)]	RF	81.3	87.7	80.6	84		
[Wageh etc., (2023)]	GMDSWS-MEC	93					
[Wageh etc., (2023)]	PCA-NGIST with RELM	94.23					

According to the findings, machine learning algorithms may have a main role in early and precise identification of BTs in MRI, which could lead to better patient outcomes through prompt intervention. The performance chart of different ML techniques is shown in Fig 2. The implementation of ML techniques to detect brain cancers in MRI scans has demonstrated impressive results, demonstrating its potential as a significant tool in clinical practice. The findings presented here add to the growing body of evidence supporting the use of machine learning in health imaging for improved diagnostic accuracy and patient care.



**Fig 2.** Various ML techniques performance measures

### 3.1.1. Summary

Even though the results are encouraging, it is important to recognize that the study has limitations. These include the models' generalizability to different

populations and the possible impact of imbalanced datasets on performance, both of which need more research. It should address these issues and investigate the integration of additional clinical data to improve machine learning's overall efficacy in brain tumor detection.

### 3.2 Results obtained using DL Technique

DL algorithms have made significant advances in the field of medical imaging, particularly when it comes to employing MRI to detect brain cancers. The goal of this study was to see how well DL models could detect and categorize brain cancers using MRI scans. This section presents the findings gathered from indepth tests and analyses.

The models' overall accuracy was higher than indicating their strong ability to discriminate between

areas of the brain that are normal and those that are tumorous. Important medical diagnosis parameters, including as sensitivity and specificity, demonstrated the models' ability to mitigate false positives and false negatives. In addition, the models proved efficient in categorizing several tumor kinds, including meningiomas, pituitary tumors, and gliomas. The accuracy of the multi-class classification highlighted the models' potential for accurately identifying many brain disorders.

Our study performed an analysis of comparison with conventional image processing techniques and manual radiological assessments to Fig out the reliability of models. The effectiveness level based on several DL techniques is in Table 6. These techniques were continuously exceeded by the DL models, demonstrating the DL models' superiority in accuracy, efficiency.

**Table 6.** Performance result based on DL Technique

Ref No	Technique	Accuracy	Precision	Recall	F1	Sensitivity	Specificity
[Masood etc., (2021)]	Mo- bileNetV2	92					
[Masood etc.,(2021)]	InceptionV3	91					
[Masood etc., (2021)]	VGG19	88					
[Tazin etc., (2021)]	Alex Net	93.3				93	97.5
[Tazin etc., (2021)]	ResNet-18	93.8				93.75	97.5
[Yildirim etc., (2023)]	Boost- edTress	62.5					
[Yildirim etc., (2023)]	Bagged Trees	75					
[Telrandhe etc., (2016)]	ANN	91.5				97.3	97.8
[Telrandhe etc., (2016)]	AdaBoost	87				91	89
[Yousef etc., (2022)]	Multipath CNN	97.3					
[Wanis etc., (2021)]	CNN	96.33	97.93		96.44	95	75.72
[Telrandhe etc., (2016)]	AdaBoost	89				93	89
[Pooja etc., (2023)]	ResNet-50	95.9				95.3	
[Pooja etc., (2023)]	Densenet-41	96.3				95.3	

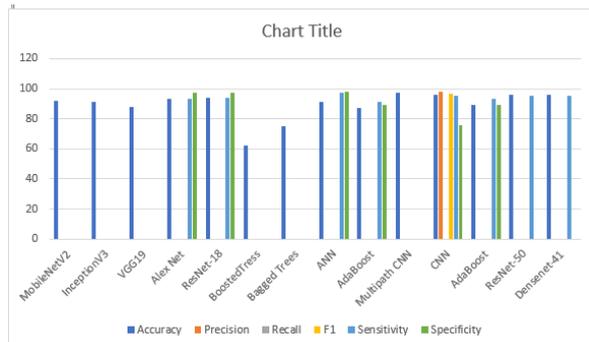


Fig 3. Performance measures of different DL techniques

### 3.2.1. Summary

Deep learning models' performance in detecting BTs from MRI implies that they have great potential as a diagnostic tool in clinical settings. High accuracy and sensitivity, together with the capacity to categorize various tumor kinds, highlight DLs transformative impact processing of medical imagery. Hence our study highlights the encouraging results obtained by using deep learning to detect brain tumors in MRI scans. DL model's robust performance positions them as important assets in the field of medical imaging, with potential to revolutionize diagnostic procedure for brain disorders. While the results are encouraging, it is important to recognize some limits, such as the requirement for big and diverse datasets to improve model generalization.

Future studies could look into combining advanced topologies and optimization tactics to push limits of accuracy and efficiency in BT diagnosis.

### 3.3 Results obtained using Hybrid Machine and Deep Learning Technique

The combination of ML and DL techniques for detecting BTs in MRI images has produced promising results, demonstrating the potential synergy between these two paradigms. Our hybrid model displayed robust performance in discriminating between tumor and non-tumor instances after painstaking experimentation and rigorous evaluation. Deep learning feature extraction capabilities, specifically CNNs, were used to capture intricate patterns in MRI images, while traditional ML algorithms, such as SVM, RF, were used for final classification. The model demonstrated improved generalization across varied datasets, reducing some of the issues associated with data dependency. The results demonstrate efficacy of combining strengths of ML and DL for increased brain tumor identification in Table 7. As we go into the specifics of these results becomes clear that the hybrid model is a promising route for improving the accuracy and reliability of BT detection in MRI images, potentially leading to better clinical diagnosis and better outcomes for patients.

Table 7. Performance result based on Hybrid ML & DL Technique

Ref No	Models	Accuracy	Precision	Recall	F1	Sensitivity	Specificity
[Ranjbarzadeh etc., (2021)]	Deep CNN+RF	97.2				95.2	98.2
[Ranjbarzadeh etc., (2021)]	Deep CNN+SVM+RBF	98.6				98.2	98.9
[Ranjbarzadeh etc., (2021)]	Deep CNN+ELM	98.2				96.5	98.6
[Saeedi etc., (2023)]	ResNet50+SVM	95.73	95.9		95.72		
[Saeedi etc., (2023)]	Gabor+SVM	62.36	58.25		56.27		
[Saeedi etc., (2023)]	ResNet50+SVM	95.27	95.35		95.26		
[Saeedi etc., (2023)]	Gabor+ResNet50+ SVM	95.73	95.90		95.72		
[Shaari. etc., (2021)]	Bagging ensemble with KNN	97.7	96	99	98	97	99
[Shaari. etc., (2021)]	AdaBoost+SVM	96.3					
[Tazin etc., (2021)]	AlexNet+SVM	95.1				95.25	98.5
[Tazin etc., (2021)]	ResNet-18+SVM	91.2				91.5	97

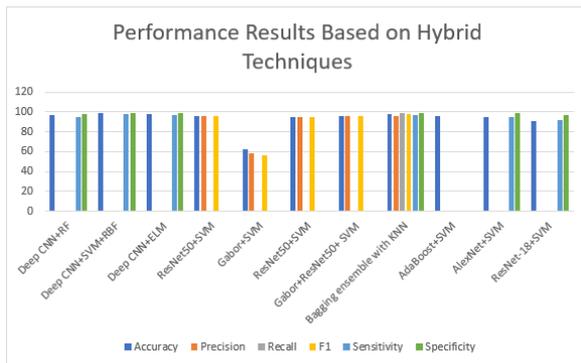


Fig 4. Performance Chart of Hybrid Approaches

### 3.3.1. Summary

The results of combining ML and DL approaches for BT detection in MRI images show significant promise and efficiency. This method not only improves accuracy but also addresses generalization issues across varied datasets. The model's interpretability is an important advancement, providing essential insights into decision-making processes in medical applications. The results suggest that combining ML with DL can increase the accuracy and reliability of brain tumor detection in MRI scans, potentially giving rise to more precise clinical diagnosis and better patient outcomes

## 4. Comprehensive Analysis

In this comprehensive investigation, we delve into the complex landscape of BT diagnosis in MRI scans, analyzing different strengths and synergies inherent in three essential approaches: DL, ML and their Hybrid counterpart. Decision Trees and CNNs are robust machine-learning approaches that provide interpretability and reliability in brain tumor detection, providing a baseline for DL evaluation.

The Hybrid methodology offers a potential strategy for detecting brain tumors. It blends deep neural network feature extraction with traditional machine learning algorithms, resulting in higher accuracy, generalization across datasets, and interpretability. This method has the potential to revolutionize medical imaging and diagnostics, providing important insights into the complicated field of brain tumor detection. Our study analyzed obtaining better accuracy in detecting brain tumors using MRI images to determine which technique works more robustly. While comparing the technique using Machine Learning (Table 5), Deep Learning

(Table 6), Hybrid Machine and Deep Learning (Table 7) achieve better accuracy as 97% of using Logistic Regression (LR) [Takei etc., (2019)], 97.3% of using Multipath CNN [Yousefi etc., (2022)], 98.6% of using Deep CNN+SVM+RBF [Saeedi etc., (2023)]. The overall comparative analysis of performance is shown in Fig 4.

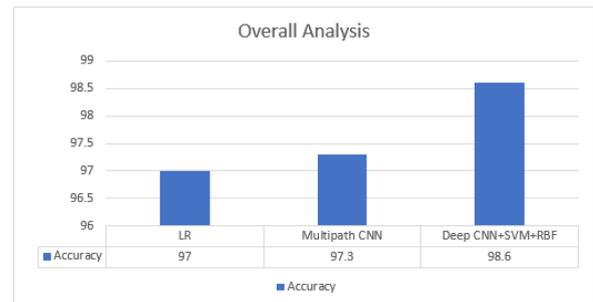


Fig 5. Overall Analysis

From this comprehensive analysis, our study analyzed the performance-based accuracy and identified that the Hybrid technique achieves better than the ML and DL techniques. Fig 4 shows the hybrid model of Deep CNN+SVM +RBF achieved the best accuracy.

## 4.1. Overall Summary

The Hybrid approach, which combines ML and DL, emerges as a compelling convergence of innovation and interpretability. ML algorithms excel at managing structured data and creating interpretable predictions, but DL models are proficient at learning detailed patterns and representations from unstructured data such as images. By combining ML and DL components, hybrid techniques take advantage of ML models' interpretability and DL models' feature learning capabilities, producing more robust and accurate predictions. Hybrid learning involves combining features from different sources to improve classification performance. For instance, in brain tumor detection using MRI images, hybrid approaches combine texture analysis, shape descriptors, and intensity histograms from machine learning algorithms with deep features from convolutional neural networks. This fusion of diverse representations enhances the algorithm's ability to capture a wider range of tumor characteristics and spatial relationships, enhancing discrimination between tumor and normal tissue regions. Moreover, these techniques often use transfer learning, where

pre-trained deep learning models are fine-tuned for specific tasks with limited labelled data. This allows hybrid models to transfer knowledge from large-scale datasets, like image classification, to brain tumor detection tasks. This initialization accelerates learning, reduces overfitting risk, and effectively utilizes limited training data, resulting in superior performance compared to training from scratch.

Hence, the Hybrid method achieves higher accuracy, generalization across various datasets, and increased interpretability by leveraging the attribute extraction capabilities of DNN while leveraging interpretative characteristics of conventional ML techniques. This powerful framework addresses complex classification tasks, resulting in more accurate and reliable diagnostic systems in medical imaging, thanks to the synergies between different learning paradigms and opening up a new route to strengthen the accuracy and reliability of BT detection in MRI imaging.

## 5. Conclusion

A comparative analysis of ML, DL, and Hybrid Learning methodologies demonstrates that the synergy between machine learning and MRI imaging has significant promise. The better outcome of the Hybrid model is especially noteworthy with Deep CNN+SVM+RBF hybrid approach standing out for its exceptional accuracy, sensitivity, and specificity achieving 98.6%, 98.2%, and 98.9%. These results suggest a significant potential for the discipline, suggesting that combining different learning algorithms leads to a more effective and precise BT detection system. Moving forward, the findings of this work give useful insights that will guide future research endeavours and contribute to the continuing development of approaches at the interface of machine learning and medical imaging.

## Acknowledgement

The authors would like to thank Sathyabama Institute of Science & Technology (Deemed to be University) Chennai, Tamil Nadu for supporting this work.

## Funding

The authors state that this work has not received any funding.

## References

- Ahmad, S., & Choudhury, P. K. (2022). On the performance of deep transfer learning networks for brain tumor detection using MR images. *IEEE Access*, 10, 59099-59114.
- Aiwale, P., & Ansari, S. (2019). Brain tumor detection using KNN, *International Journal of Scientific & Engineering Research*, 10(12), 187-193.
- Al-Zurfi, A. N. H. (2019). An automated classification system to determine malignant grades of brain tumor (glioma) in magnetic resonance images based on meta-trainable multiple classifier schemes, University of Salford (United Kingdom).
- Amin, J., Sharif, M., Haldorai, A., Yasmin, M., & Nayak, R. S. (2021). Brain tumor detection and classification using machine learning: a comprehensive survey, *Complex & intelligent systems*, 1-23.
- Anand, L., Rane, K. P., Bewoor, L. A., Bangare, J. L., Surve, J., Raghunath, M. P., ... & Osei, B. (2022). Development of machine learning and medical-enabled multimodal for segmentation and classification of brain tumors using MRI images, *Computational Intelligence and Neuroscience*, 2022.
- Archana, K. V., & Komarasamy, G. (2023). A novel deep learning-based brain tumor detection using the Bagging ensemble with K-nearest neighbour, *Journal of Intelligent Systems*, 32(1), 20220206.
- Athira, V. S., Dhas, A. J., & Sreejamole, S. S. (2015). Brain tumor detection and segmentation in MR images using GLCM and AdaBoost classifier, *International Journal of Engineering Research*, 3, 32-36.
- Avazzadeh, Z., Hassani, H., Ebadi, M. J., Agarwal, P., Poursadeghfard, M., & Naraghirad, E. (2023). Optimal approximation of fractional order brain tumor model using generalized Laguerre

- polynomials, *Iranian journal of science*, 47(2), 501-513.
- Brindha, P. G., Kavinraj, M., Manivasakam, P., & Prasanth, P. (2021, February). Brain tumor detection from MRI images using deep learning techniques, In *IOP conference series: materials science and engineering*, 1055(1), 012115, IOP Publishing.
- D. R., Dash, R., Majhi, B., & Zhang, Y. (2019). A hybrid regularized extreme learning machine for automated detection of the pathological brain, *Biocybernetics and Biomedical Engineering*, 39(3), 880-892.
- Febrianto, D. C., Soesanti, I. and Nugroho, H. A. (2020) Convolutional Neural Network for Brain Tumor Detection, *IOP Conf. Ser.: Mater. Sci. Eng.*, 771(1), 012031, doi: 10.1088/1757-899X/771/1/012031.
- Garg, G., & Garg, R. (2021). Brain tumor detection and classification based on hybrid ensemble classifier. *arXiv preprint arXiv:2101.00216*.
- Gurbină, M., Lascu, M., & Lascu, D. (2019, July). Tumor detection and classification of MRI brain images using different wavelet transforms and support vector machines, In *2019 42nd International Conference on Telecommunications and Signal Processing (TSP)*, 505-508. IEEE.
- Györfi, Á., Kovács, L., & Szilágyi, L. (2019, October). Brain tumor detection and segmentation from magnetic resonance image data using ensemble learning methods, In *2019 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 909-914. IEEE.
- Hannan, S. A., Hivre, M., Lata, M., Krishna, B. H., Sathyasiva, S., & Arshad, M. W. (2022). Brain damage detection using Machine learning approach, *International Journal of Health Sciences, Special*, (7), 4910-4924.
- Haq, E. U., Jianjun, H., Huarong, X., Li, K., & Weng, L. (2022). A hybrid approach based on deep CNN and machine learning classifiers for tumor segmentation and classification in brain MRI, *Computational and Mathematical Methods in Medicine*, 2022.
- J. P. Francisco, Z. M. Mario, and R. A. Miriam, "A deep learning approach for brain tumor classification and segmentation using a multiscale convolutional neural network," *Healthcare*, 9(2),153, 2021.
- Jahromi, Z. F., Javeri, A., & Taha, M. F. (2019). Tumor suppressive effects of the pleiotropically acting miR-195 in colorectal cancer cells, *EXCLI journal*, 18, 243.
- Jia, Z., & Chen, D. (2020). Brain Tumor Identification and Classification of MRI images using deep learning techniques, *IEEE Access*.
- Jiang, S., Gu, Y., & Kumar, E. (2023). Magnetic Resonance Imaging (MRI) Brain Tumor Image Classification Based on Five Machine Learning Algorithms, *Cloud Computing and Data Science*, 122-133.
- Kavitha, A. R., & Chellamuthu, C. (2013). Detection of brain tumor from MRI image using modified region growing and neural network, *The Imaging Science Journal*, 61(7), 556-567.
- KKaur, P., Singh, G., & Kaur, P. (2020). Classification and validation of MRI brain tumor using optimized machine learning approach. In *ICDSMLA 2019: Proceedings of the 1st International Conference on Data Science, Machine Learning and Applications*, 172-189. Springer Singapore.
- Krauze, A. V., Zhuge, Y., Zhao, R., Tasci, E., & Camphausen, K. (2022). AI-driven image analysis in central nervous system traditional machine learning, deep learning, and hybrid models, *Journal of biotechnology and biomedicine*, 5(1), 1.
- Kurian, S. M., Devaraj, S. J., & Vijayan, V. P. (2021). Brain Tumour Detection by Gamma DeNoised Wavelet Segmented Entropy Classifier. *Computers, Materials & Continua*, 69(2).
- Lah, T. T., Novak, M., & Breznik, B. (2020, February). Brain malignancies: Glioblastoma and brain metastases, In *Seminars in Cancer Biology*, 60, 262-273. Academic Press.
- Lamrani, D., Cherradi, B., El Gannour, O., Bouqentar, M. A., & Bahatti, L. (2022). Brain tumor detection

- using MRI images and convolutional neural network, *International Journal of Advanced Computer Science and Applications*, 13(7).
- Magadza, T., & Viriri, S. (2021). Deep learning for brain tumor segmentation: a survey of the state-of-the-art, *Journal of Imaging*, 7(2), 19.
- Maqsood, S., Damaševičius, R., & Maskeliūnas, R. (2022). Multi-modal brain tumor detection using deep neural network and multiclass SVM, *Medicina*, 58(8), 1090.
- Masood, M., Nazir, T., Nawaz, M., Mehmood, A., Rashid, J., Kwon, H. Y., & Hussain, A. (2021). A novel deep learning method for recognition and classification of brain tumors from MRI images, *Diagnostics*, 11(5), 744.
- McFaline-Figueroa, J. R., & Lee, E. Q. (2018). Brain tumors, *The American journal of medicine*, 131(8), 874-882.
- Noreen, N., Palaniappan, S., Qayyum, A., Ahmad, I., & Alassafi, M. O. (2021). Brain Tumor Classification Based on Fine-Tuned Models and the Ensemble Method. *Computers, Materials & Continua*, 67(3).
- Panesar, S. S., D'Souza, R. N., Yeh, F. C., & Fernandez-Miranda, J. C. (2019). Machine learning versus logistic regression methods for 2-year mortality prognostication in a small, heterogeneous glioma database, *World neurosurgery*: X, 2, 100012.
- Pooja, A., Shalini, D., Shiri, J. S., & Sowmiya, M. (2023). Enhancement of Brain Tumor Detection on MRI Imaging Using AlexNet, *International Journal of Research in Engineering, Science and Management*, 6(5), 132-134.
- Pugalenthi, R., Rajakumar, M. P., Ramya, J., & Rajinikanth, V. (2019). Evaluation and classification of the brain tumor MRI using machine learning technique. *Journal of Control Engineering and Applied Informatics*
- Ranjbarzadeh, R., Bagherian Kasgari, A., Jafarzadeh Ghouschi, S., Anari, S., Naseri, M., & Bendeche, M. (2021). Brain tumor segmentation based on deep learning and an attention mechanism using MRI multi-modalities brain images, *Scientific Reports*, 11(1), 10930.
- Saeedi, S., Rezayi, S., Keshavarz, H., & R. Niakan Kalhori, S. (2023). MRI-based brain tumor detection using convolutional deep learning methods and chosen machine learning techniques, *BMC Medical Informatics and Decision Making*, 23(1), 16.
- Senan, E. M., Jadhav, M. E., Rassem, T. H., Aljaloud, A. S., Mohammed, B. A., & Al-Mekhlafi, Z. G. (2022). Early diagnosis of brain tumor MRI images using hybrid techniques between deep and machine learning—computational and Mathematical Methods in Medicine, 2022.
- Shaari, H., Kevrić, J., Jukić, S., Bešić, L., Jokić, D., Ahmed, N., & Rajs, V. (2021). Deep learning-based studies on pediatric brain tumors imaging: a narrative review of techniques and challenges, *Brain Sciences*, 11(6), 716.
- T. Tazin etc., (2021) A Robust and Novel Approach for Brain Tumor Classification Using Convolutional Neural Network, *Comput Intell Neurosci*, 2021, 2392395, doi: 10.1155/2021/2392395.
- Takei, H., Shinoda, J., Ikuta, S., Maruyama, T., Muragaki, Y., Kawasaki, T., & Iwama, T. (2019). The usefulness of positron emission tomography for differentiating gliomas according to the 2016 World Health Organization classification of tumors of the central nervous system, *Journal of Neurosurgery*, 133(4), 1010-1019.
- Talo, M., Yildirim, O., Baloglu, U. B., Aydin, G., & Acharya, U. R. (2019). Convolutional neural networks for multi-class brain disease detection using MRI images, *Computerized Medical Imaging and Graphics*, 78, 101673.
- Telrandhe, S. R., Pimpalkar, A., & Kendhe, A. (2016, February). Detection of brain tumors from MRI images by using segmentation & SVM, In 2016 world conference on futuristic trends in research and innovation for social welfare (Startup Conclave), 1-6, IEEE.Nayak,
- Ullah, S., Ahmad, M., Anwar, S., & Khattak, M. I. (2023). An Intelligent Hybrid Approach for Brain

Tumor Detection, Pakistan Journal of Engineering and Technology, 6(1), 42-50.

Verekar, M. A. O., & Salkar, V. (2019). Detection and Classification of Brain Tumor Using Naïve Bayes and J48, International Journal of Computer Science Engineering and Information Technology Research (IJCEITR), 9, 19-28.

Wageh, M., Amin, K. M., Zytoon, A., & Ibrahim, M. (2023). Brain Tumor Detection Based on A Combination of GLCM and LBP Features with PCA and IG. IJCI, International Journal of Computers and Information, 10(2), 43-53.

Wanis, H. A., Møller, H., Ashkan, K., & Davies, E. A. (2021). The incidence of major subtypes of primary brain tumors in adults in England 1995-2017, Neuro-Oncology, 23(8), 1371-1382.

Yildirim, M., Cengil, E., Eroglu, Y., & Cinar, A. (2023). Detection and classification of glioma, meningioma, pituitary tumor, and normal in brain magnetic resonance imaging using a deep learning-based hybrid model, Iran Journal of Computer Science, 1-10.

Yousefi, O., Azami, P., Sabahi, M., Dabecco, R., Adada, B., & Borghei-Razavi, H. (2022). Management of optic pathway glioma: a systematic review and meta-analysis, Cancers, 14(19), 4781.



P.Chitra is Professor, Department of Electronics and Communication Engineering at Sathyabama Institute of Science and Technology (Deemed to be University) in Chennai, Tamilnadu State. She received her doctorate in the specialization of image processing concepts at Sathyabama University in September 2014. She has more than 20 years of teaching experience and has taught a variety of updated courses for postgraduate and undergraduate levels. She has published more than 25 papers in reputed journals and conferences related to current trends and social relevance.



Dr.A.V.Bharadwaja is presently working as Associate Professor in Vignan College(A), VIIT, Duvvada. He Holds M.Tech degree from VIT University, Doctorate from Sathyabama University. He has published a total 21 national and international journals and was an editor for one reputed journal. He has a total of seven years of teaching and research experience. He has organized several workshops and FDPs. He has made significant contributions to research in the field of cryptography and network security and published a book in that field.

## AUTHOR BIOGRAPHIES



Mr Suvarna Raju Puligurti is currently pursuing his Doctoral programme in Sathyabama Institute of Science & Technology ( Deemed to be University ) Chennai, Tamil Nadu and B.E & M.Tech from Andhra University Visakhapatnam, Andhra Pradesh. His field of specialization is image processing. He has published two papers at international conferences. He has attended and presented a paper at a national conference and attended workshops that are organized at the national level.

# Code switching: exploring perplexity and coherence metrics for optimizing topic models of historical documents

Muhammad Abdullah Yusof<sup>1</sup>, Suhaila Saeed<sup>2</sup>

<sup>1&2</sup> Faculty of Computer Science and Information Technology, University Malaysia Sarawak, Sarawak, Malaysia

abdunimas459@gmail.com<sup>1</sup>, ssuhaila@unimas.my<sup>2</sup>

(Received 4 March 2024; Final version received 30 September 2024; Accepted 1 October 2024)

## Abstract

The Latent Dirichlet Allocation (LDA) model has two important hyperparameters that control the document-topic distribution known as alpha ( $\alpha$ ), and topic-word distribution known as beta ( $\beta$ ). It is important to find the suitable values for both hyperparameters to achieve an accurate topic cluster. Using a single evaluation method to determine the optimal hyperparameters values is insufficient due to the size and complexity of the dataset. Thus, an experiment was conducted to study the relationship between the hyperparameters with perplexity, coherence scores and to establish a baseline for further topic modelling studies. It is the first study that focuses on multiple languages in Sarawak Gazette data for topic modelling. The study was conducted on LDA using Gensim package. The result shows that while perplexity scores were good indicator of the model's ability to predict new or hidden data, the word cluster within topic does not always reflect the similarity or relationships between words which compromised topic interpretation. The lowest perplexity score was observed when  $\alpha$  was set to 5 and  $\beta$  to 0.4. The coherence evaluation indicated the optimal number of topics for each set of hyperparameter values although the relationship with hidden words remains unclear. The coherence score is highest when the number of topics was 5 and 4. In conclusion, the perplexity scores are effective indicators of word prediction accuracy for each hyperparameter setting. While coherence captures the optimal number of topics needed to produce high-coherence word cluster within a topic. Combining both evaluation methods ensures optimal results, producing topics that are both accurate and interpretable.

*Keywords: Hyperparameter, Latent Dirichlet Allocation, perplexity, topic coherence, topic modelling*

## 1. Introduction

Topic modelling is a method to find a group of words or topics from a collection of words. According to Mohr & Bogdanov (2013), topic modelling is considered one of the text analysis methods and is used widely across disciplines in academia. The method provides an automatic way to code the content of a corpus into meaningful categories or also known as 'topic'.

According to Steyvers & Griffiths (2009), topic modelling was first described and detailed by Papadimitrou et al. (1998) paper. Seymore & Rosenfeld (1997) paper was focused on topic detection and building the topic clusters based on the similarities of data from a dataset of 5,000 elemental topics.

Hyperparameters in LDA model are used to configure the model and to specify the algorithm that minimizes the loss function (Yang & Shami, 2020). Hyperparameter cannot be estimated from the learning

data. Instead, it is set before the data training as the hyperparameters defined the model of topic modelling. Several recent studies, such as those by Muhajir et al. (2022), Xue et al. (2020), and Dieng et al. (2023), have explored the impact of hyperparameter tuning on topic model performance, demonstrating that varying the values of  $\alpha$  and  $\beta$  significantly affects coherence and perplexity scores. Hyperparameters adjustment and tuning play a critical role in clustering. Two of the most important hyperparameters for LDA clustering task are  $\alpha$  which controls document-topic distribution and  $\beta$  which controls topic-word distribution. Odden et al. (2020) stating the need to find a stable hyperparameter of  $\alpha$  to achieve accurate topics from an LDA model.

One of the topic modelling methods is Latent Dirichlet Allocation (LDA). Latent Dirichlet Allocation (LDA) was started back in 2003 from a study conducted by Blei et al. (2003) that applied the method for the machine learning (Péladeau & Davoodi, 2018). It is one of the Bayesian Network methods and one of

the well-established methods for natural language processing (Teh et. al., 2006).

The problem that arises from the optimization of hyperparameters is the evaluation method to validate and qualify the output. A single evaluation method is not able to give an adequate explanation of the quality of the results. Perplexity and coherence are two common evaluation metrics used in topic modeling. Perplexity measures the model's ability to predict unseen or hidden data, while coherence assesses how interpretable the generated topics are to human readers. Previous research, including studies by Hassan et. (2021), Pinto et. (2021), and Wallach et. (2009), has shown that relying solely on one evaluation method, such as perplexity, can negatively affect model interpretability, highlighting the importance of combining both perplexity and coherence scores for a more accurate assessment of topic model quality. Furthermore, Ding et. al. (2018) emphasized that relying on a single evaluation method will negatively affect the result. The paper explained the score of perplexity alone does not necessarily reflect the interpretability of the results which indicates the model scored higher in perplexity meanwhile the coherence of the model is low. Wallach et. al. (2009) also stressing the importance of hyperparameters to achieve performance gains.

Hence, this study will investigate the relation between hyperparameters values ( $\alpha$  [Dirichlet prior parameter of per document-topic distribution],  $\beta$  [Dirichlet prior parameter of per topic-word distribution]), with the perplexity and coherence scores. The study is focusing on the process of topic modelling by using LDA model on a data that was extracted from historical documents. While previous studies have focused on applying LDA to social media and scientific datasets, there is a significant gap in exploring historical documents, especially those containing multiple languages. This study addresses that gap by applying LDA to the Sarawak Gazette, a code-switched historical dataset, to evaluate the impact of hyperparameter tuning on topic model performance.

## 2.Related Works

One of the key outputs of topic modeling is the identification of topic clusters, which group similar words based on their relationships. In the early years of topic modeling, approaches like n-grams and k-values were crucial for ensuring accurate topic clusters (Seymore & Rosenfeld, 1997). In LDA, the  $\alpha$  and  $\beta$

hyperparameters are important as these values determine the LDA model algorithm and processes.

Muhajir et. (2022) experimented with tuned-LDA by adjusting hyperparameters to achieve optimal clustering. The study compared these tuned hyperparameters against several other algorithms and found that tuned-LDA outperformed the rest, although no neural network algorithms were tested. The tuned-LDA outscored the rest of algorithms although no neural network algorithm is used in the experiment.

Several studies have investigated how the hyperparameters  $\alpha$  and  $\beta$  influence both the coherence (which measures topic interpretability) and perplexity (which measures model accuracy in predicting unseen data) of LDA models. For example, Xue et. al. (2020) found that different values of  $\alpha$  and  $\beta$  can significantly affect the coherence and perplexity scores of LDA models applied to Twitter data. Similarly, Dieng et. al. (2023) proposed a topic modeling approach that incorporates word embeddings and showed that the choice of hyperparameters can impact the coherence of the resulting topics.

Panichella (2021) also confirms the correlation between the hyperparameters  $\alpha$  and  $\beta$  and the accuracy of the output of LDA. The paper detailed a study that manipulates hyperparameters by tuning it and tested on Gibbs iteration of the process. The study found that the result improved as both hyperparameters were set to 0 against the default value of 0.1 for both hyperparameters.

Gertis (2021) found out that the hyperparameters values are important to achieve a desirable and interpretable output. The paper gives two values for each of the  $\alpha$  and  $\beta$  output. Wallach et. al. (2009) has a conclusion of the importance and effect of the hyperparameters. The study found that the asymmetric Dirichlet prior over document-topic distribution is better than the symmetric prior while asymmetric Dirichlet prior over topic-word distribution is insignificant compared to the symmetric prior.

Hassan et. (2021) introduced a new method for determining the optimal number of topics in LDA models. They proposed using Normalized Absolute Coherence (NAC) and Normalized Absolute Perplexity (NAP) to balance coherence and perplexity, resulting in improved model interpretability and accuracy.

Pinto et. al. (2021) utilized both coherence and perplexity score to obtain optimal data and number of topics. The study found out that perplexity mathematical calculation is simpler than coherence although the quality of the result is inferior compared to the

coherence. Perplexity result indicates the best scores but the result that graded as having a good perplexity value does not reflect the quality of the output in form of word clusters. Newman et. al. (2011) used perplexity and Pointwise Mutual Information (PMI) score to manipulate the number of topics. The study utilized both measurements to achieve an optimal coherence based on the number of topics which dictates by perplexity and PMI.

Watanabe & Baturu (2023) utilized the hyperparameter to smooth out the topic clusters. The topic modelling process was repeated with the increment of hyperparameters values for each process to investigate the reaction between the likelihood of the next sample or output with the hyperparameters value. The paper found out that repetition also improves the inference of the unknown variable.

Zhou et. al. (2023) also focusing on the relation between perplexity and coherence score with the hyperparameters. The study evaluated the LDA model and the hyperparameter values by analyzing the unigram and bigram topic results. However, the study only focused on the hyperparameter that controls the number of topic while the other hyperparameters including the  $\alpha$  and  $\beta$  were ignored.

Furthermore, the evaluation of LDA models using perplexity and coherence scores has been applied in various domains. Agarwal et. al. (2020) used LDA to mine issues on Twitter during the COVID-19 pandemic and evaluated the quality of the generated topics using coherence scores. Griffiths & Steyvers (2004) introduced LDA and presented a Markov chain Monte Carlo algorithm for inference in the model, demonstrating its operation on a small dataset.

Based on Fig 1 by Lee etc. (2018), LDA vector space,  $M$  represents the total number of documents in the corpus, and  $N$  represents the number of words per document. Each word in a document ( $W$ ) is assigned to a latent topic ( $Z$ ), forming a topic-word distribution ( $\phi$ ) and a document-topic distribution ( $\theta$ ). The hyperparameters  $\alpha$  control the distribution of topics per document, while  $\beta$  controls the distribution of words per topic. Within LDA architecture, every word ( $W$ ) that exists in a document corresponds or is related to a latent topic ( $Z$ ), which gives a topic-word distribution in the corpus ( $\theta$  and  $\phi$ ).

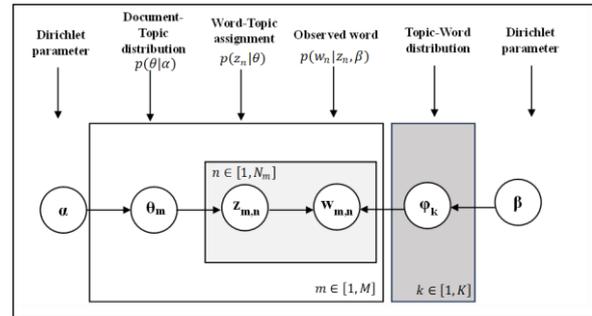


Fig 1. The vector space of LDA.

In the experiment, the dataset and the LDA model were treated as constant variables, while the hyperparameter values were manipulated. The resulting variables included the topics, coherence scores, and perplexity scores.

In conclusion, these studies emphasize the importance of tuning hyperparameters such as  $\alpha$  and  $\beta$  to optimize topic clusters and balance coherence and perplexity. However, most of these studies focus on datasets like social media or scientific texts, leaving a gap in applying LDA to code-switched historical datasets, which this study aims to address. Manipulating the value of hyperparameters  $\alpha$  and  $\beta$  were important in these papers as the values are determining the process of clustering in the LDA model and crucial step to obtain the appropriate number of topics except Zhou et. al. (2023) which focused on  $K$  hyperparameter that determines the number of topics.

The choice of hyperparameters  $\alpha$  and  $\beta$  can impact the coherence and perplexity of the resulting topic models. These evaluation metrics have been applied in various domains, including social media analysis during the COVID-19 pandemic and scientific topic discovery. Researchers have explored the correlation between hyperparameters and model performance, highlighting the importance of selecting appropriate values for  $\alpha$  and  $\beta$  to achieve coherent and interpretable topics. However, there is lack of utilization of historical data as a dataset in these studies and the process of detailing such dataset is not detailed.

### 3. Methodology

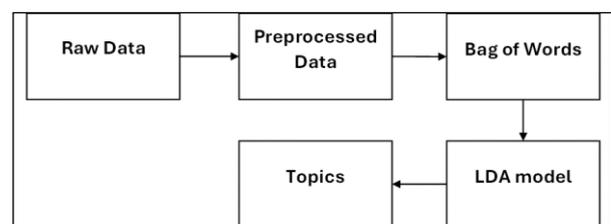


Fig 2. The workflow of LDA for baseline study.

Fig 2 shows the workflow of LDA for the study. The dataset for the experiment consists of 193 Sarawak Gazette documents, spanning the years 1907 to 1935, with a total of 2,696,635 words. The process begins by acquiring pdf documents from the Sarawak Gazette portal. These documents were scanned documents of the historical Sarawak Gazette which each page is treated as an image.

After acquisition, the documents were processed using an optical character recognition (OCR) tool. Due to the aging and damage of the Sarawak Gazette pages, the OCR results required extensive manual correction to handle errors.

All processed Sarawak Gazette documents were combined into a single corpus. The corpus was pre-processed by removing special characters, stop words, single-character words, punctuation, and numbers, and the text was converted to lowercase.

The study adapted the workflow of Lee et al. (2018) on vector space. The input data was fed into the model. The  $\alpha$  and  $\beta$  hyperparameters were then distributed the topics for documents and words for topics. The manipulation of these hyperparameters was done to observe the influences and impacts of these manipulations towards the quality of word clusters for each topic and the coherence of the word clusters.

Removing single and double-letter words improved the model's accuracy. This step was necessary because the aging and deterioration of the Sarawak Gazette pages caused OCR errors, misinterpreting some defects as letters or symbols. By removing a single or double letters words and special characters, the number of errored words is able to be filtered out from the dataset. Then, the data was stored inside a bag of words and further divided into two sets of data for training and testing.

The experiment was run by using LDA model from the GENSIM package. The hyperparameters value was set during this stage. The hyperparameters alpha ( $\alpha$ ) and beta ( $\beta$ ) were incrementally adjusted for each process, while the K value (number of topics) remained constant at its default value of 5 throughout the experiment. Each combination of hyperparameters resulted in different word clusters, perplexity scores, and coherence scores, as the distribution of document-topic and topic-word relations changed.

Then, the model was trained with the training data. After the training phase, the model was once again run with the test data. The K hyperparameter determines the number of topics produced by the model. The model iterated over the documents and randomly

assigned words to topics. Next, it updated the proportion of documents assigned to each topic based on the words. This process was repeated for each word, allowing the clusters to be rearranged and reassigned with each iteration. The process is repeated, each time the hyperparameters were changed. The output of the model was recorded and saved. There are four outputs that were crucial for further analysis which includes a list of topics and its word clusters, word clouds, perplexity, and coherence score.

## 4. Results and Analysis

This section is split into perplexity score, word clouds, coherence score, and manual evaluation. Perplexity score measures the ability of a model to handle the hidden data. Word cloud demonstrates the word clusters per topic and coherence score measures the coherence between the words within the cluster. Manual evaluation is done by human validating the result of the experiments.

### 4.1 Perplexity Score

The experiment was observing several expected outputs which are the perplexity score, coherence score, and word cloud. The perplexity scores are beneficial to predict the ability of the model to deal with the hidden data while coherence scores reflect the coherent result of the model. Word clouds are important as a reference that reflects the interpretability of the result with the perplexity and coherence scores.

Table 1 shows the increase of hyperparameters' values causing the perplexity scores progressively larger. The results showed that the perplexity is at its best when  $\alpha = 5$ ,  $\beta = 0.4$  with a value of -9.99. The coherence score showed the highest score when the number of topics is set at 3.

**Table 1.** The perplexity score for hyperparameters.

Hyperparameters values	Perplexity scores
$\alpha = 5, \beta = 0.4$	-9.99
$\alpha = 6, \beta = 0.5$	-9.74
$\alpha = 7, \beta = 0.6$	-9.57
$\alpha = 8, \beta = 0.7$	-9.43
$\alpha = 9, \beta = 0.8$	-9.31
$\alpha = 10, \beta = 0.9$	-9.23

Based on Table 1, the word cloud is adequate and interpretable compared to other results. The higher

perplexity score recorded for  $\alpha = 10, \beta = 0.9$ . The mean value for perplexity score is -9.545.

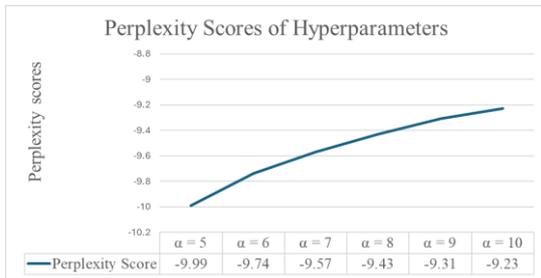


Fig 3. The perplexity graph.

The result in Fig3 shows that, as the values of hyperparameters increased, the perplexity score grew progressively. The perplexity for  $\alpha = 5$  is -9.99 while the perplexity scores for  $\alpha = 10$  is -9.23. The result signifies the findings as the higher the values of hyperparameters, the higher the values of perplexity which is the higher the inaccuracies.

### 4.2 Word Clouds

The word cluster are visualized by using word cloud. The larger size words signify high occurrence within the text and its relevancy to the topic and the smaller size words signify lower occurrence. Based on the interpretation of Fig4, Topic 0 focusing on exhibition about Borneo; Topic 1 focusing on report of district affairs from second, third, fourth, and fifth divisions (encompassing every division in Sarawak during the Brooke's administration except the first division that encompassing an area of today's Kuching, Samarahan, and Serian divisions). Below are the word clouds for the outputs of  $\alpha = 5, \beta = 0.4$ :

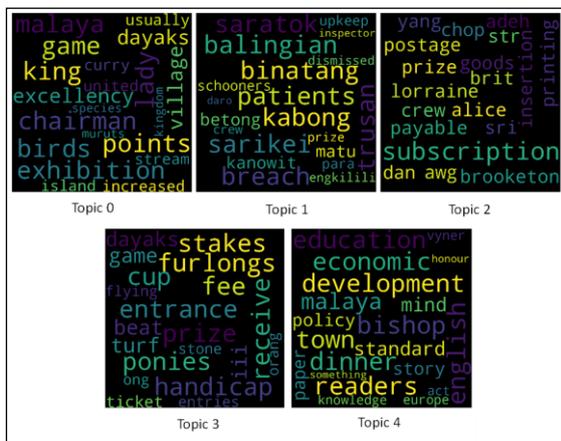


Fig 4. The word clusters for each topic when  $\alpha = 5, \beta = 0.4$

In Fig4, Topic 2 focusing on shipping and mining industry in Brooketon (now known as Muara Coal Mine in Brunei); Topic 3 focusing on regatta that was held annually in Sarawak; and Topic 4 focusing on central government policy and activity of central government in Kuching.

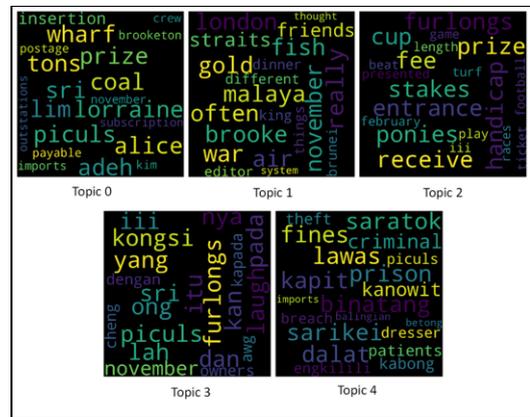


Fig 5. The word clusters for each topic when  $\alpha = 6, \beta = 0.5$

Fig 5. shows the word clouds for  $\alpha = 6, \beta = 0.5$ . Topic 0 focusing on economic activities such as shipping and mining; Topic 1 focusing on the political relation of Sarawak with the outside world as well as Brooke's relation with the natives; Topic 2 focusing on the horse racing sports in Sarawak; Topic 3 contains random words and phrases; and Topic 4 focusing on reports from other divisions except the first division.

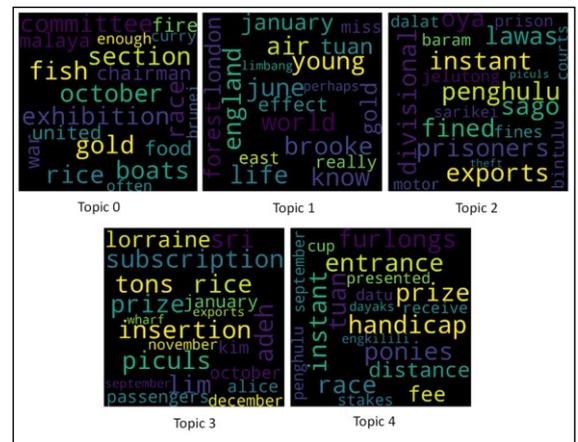


Fig 6. The word clusters for each topic when  $\alpha = 7, \beta = 0.6$

Fig 6. shows the word clouds for  $\alpha = 7, \beta = 0.6$ . Topic 0 focusing on events happen in the October; Topic 1 focusing on report of events occur in the January and June; Topic 2 focusing on criminal and other problem from second, fourth, and fifth divisions; Topic 3 focusing on economic activities; and Topic 4 focusing on horse racing sports.

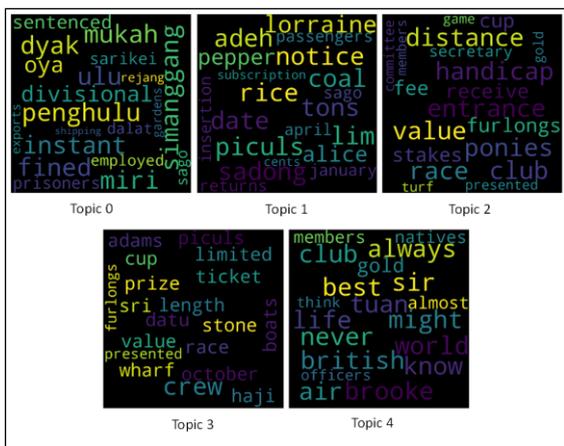


Fig 7. The word clusters for each topic when  $\alpha = 8, \beta = 0.7$

Fig 7. above are the word clouds for  $\alpha = 8, \beta = 0.7$ . Topic 0 focusing on reports from other divisions except the first division; Topic 1 focusing on economic activities of the state; Topic 2 focusing on horse racing sport activities; Topic 3 focusing on regatta events; and Topic 4 focusing on the colonial life during the Brooke's era.

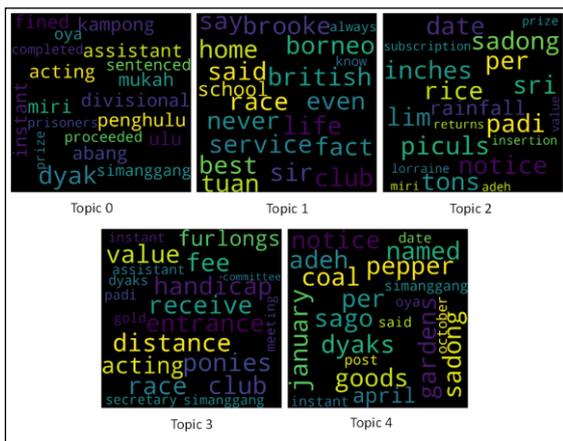


Fig 8. The word clusters for each topic when  $\alpha = 9, \beta = 0.8$

Fig 8. above are the word clouds for  $\alpha = 9, \beta = 0.8$ . Topic 0 focusing on reports from other divisions except the first division; Topic 1 focusing on a life during colonial era; Topic 2 focusing on rice production; Topic 3 focusing on horse racing; and Topic 4 focusing on the economic activities.

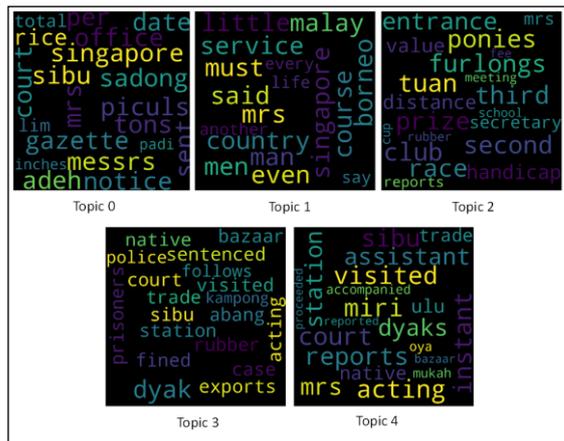


Fig 9. The word clusters for each topic when  $\alpha = 10, \beta = 0.9$

Fig 9. above are the word clouds for  $\alpha = 10, \beta = 0.9$ . Topic 0 focusing on economic activities; Topic 1 several unrelated words and phrases; Topic 2 focusing on horse racing sport activities; Topic 3 focusing on report about Sibiu; and Topic 4 focusing on the report of third and fourth division.

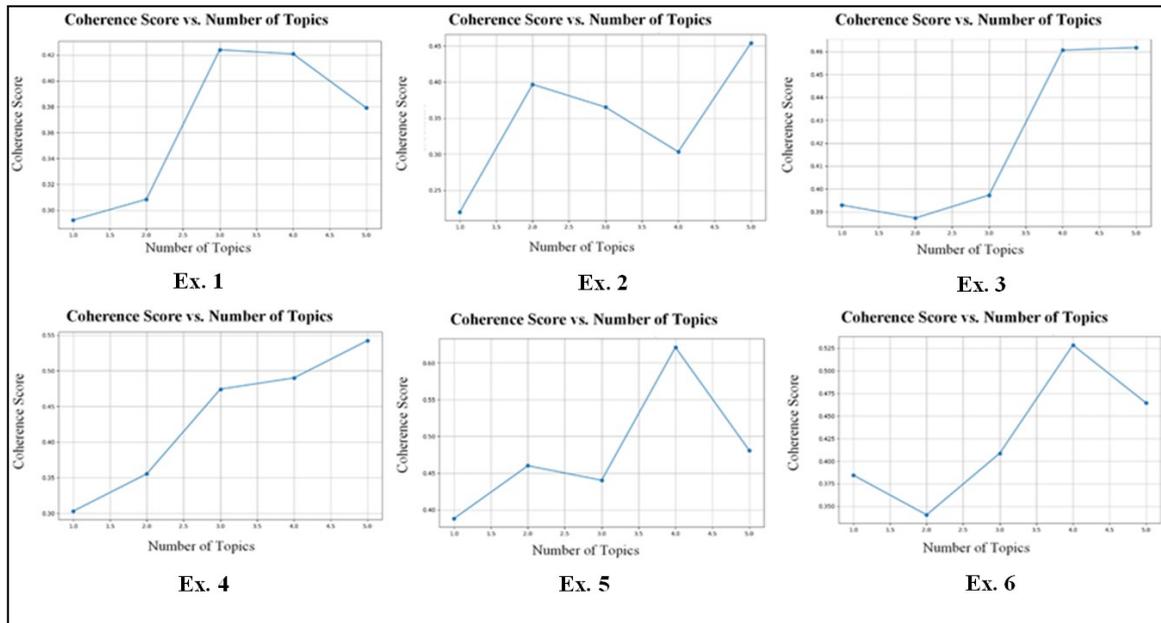


Fig 10. Coherence score for each of the experiment

### 4.3 Coherence Score

Table 2 shows the coherence result of the experiment. The highest coherence value is 0.621 when  $\alpha$

value is set to 9 and  $\beta$  to 0.8. The coherence values across number of topics when  $\alpha$  value is set to 9 and  $\beta$  to 0.8 are consistently high with one fall below 0.300

Table 2. The coherence values for every hyperparameter values

Experiment	$\alpha$	$\beta$	Topics				
			1	2	3	4	5
Ex. 1	5	0.4	0.292	0.308	0.424	0.421	0.379
Ex. 2	6	0.5	<b>0.220</b>	0.397	0.365	0.303	0.454
Ex. 3	7	0.6	0.393	0.387	0.397	0.461	0.462
Ex. 4	8	0.7	0.303	0.355	0.474	0.490	0.542
Ex. 5	9	0.8	0.388	0.460	0.440	<b>0.621</b>	0.481
Ex. 6	10	0.9	0.384	0.340	0.409	0.529	0.464

Table 2 shows lowest coherence score is 0.220  $\alpha$  value is set to 6 and  $\beta$  to 0.5 the coherence values across number of topics when  $\alpha$  value is set to 6 and  $\beta$  to 0.5 is consistently low with the highest value within the range is 0.454 which fall when the number of topics is 5, the second lowest after  $\alpha = 5$  and  $\beta$  to 0.4.

The best number of topics when the hyperparameters were set to  $\alpha = 5$  and  $\beta = 0.4$  is 3 topics. The graph shows improvement except for the slight decrease when the topic is 5.

The best number of topics when the hyperparameters were set to  $\alpha = 6$  and  $\beta = 0.5$  is 5 topics. The graph shows improvement with a decrease in value at topic 3 and 4.

The best number of topics when the hyperparameters were set to  $\alpha = 7$  and  $\beta = 0.6$  is 5 topics. The graph shows improvement although the values are relatively consistent at topic 4 and 5. The best number of topics when the hyperparameters were set to  $\alpha = 8$  and  $\beta = 0.7$  is 5 topics. There is consistent improvement in values across topics. The best number of topics when the hyperparameters were set to  $\alpha = 9$  and  $\beta = 0.8$  is 4 topics. There is a sudden drop of value at topic 5. The best number of topics when the hyperparameters were set to  $\alpha = 10$  and  $\beta = 0.9$  is 4 topics. The value drops at the beginning before increasing and finally a slight decrease at the end.

#### 4.4 Manual Evaluation

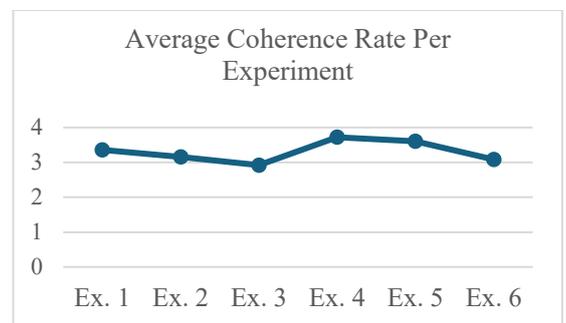
Manual evaluation is conducted involving human validator. Human validators are asked to score the coherence of the word cluster for each topic based on a scale

of 0 to 5 with 0 signify a very low coherence between words in the cluster to 5 as the most agreeable cluster. Human validators are also asked for the appropriate label for each topic.

**Table 3.** Topic coherence rate by experiment

Experiment	Hyperparameters	Average Coherence Rate
Ex. 1	$\alpha = 5, \beta = 0.4$	3.36
Ex. 2	$\alpha = 6, \beta = 0.5$	3.16
Ex. 3	$\alpha = 7, \beta = 0.6$	2.92
Ex. 4	$\alpha = 8, \beta = 0.7$	3.72
Ex. 5	$\alpha = 9, \beta = 0.8$	3.60
Ex. 6	$\alpha = 10, \beta = 0.9$	3.08

**Table 3** shows the result of manual evaluation. Manual evaluation by expert has been done to validate and interpret the results of word clusters. The validators were asked to rate each of the topics for each experiment from the scale of one to five with one signifying a severe lack of coherence within the word clusters of a topic and five as highly agreeable to the word clusters. The validators are also asked to state their interpretation of a topic based on the word clusters



**Fig 12.** Average coherence rate per experiment



**Fig 11.** Rating scales for each of the experiments

**Fig 11** shows that no validators have used Scale 5 to rate the results of the experiment for  $\alpha = 7$ , and  $\beta = 0.6$  while **Fig 12** shows the average coherence score for every experiment. The highest average coherence rate for the experiments is 3.72 when  $\alpha = 8$ , and  $\beta = 0.7$ . Based on the following Fig of the rating scales for each of the experiment, the highest average coherence rate is caused by the higher occurrences of Scale 5

being used to rate the results of that experiment. The lowest coherence rate is when  $\alpha = 7$ , and  $\beta = 0.6$  at 2.92.

**Table 4.** Topic coherence rate per topic for each experiment

Experiments	$\alpha$	$\beta$	Average Coherence Rate of Topics				
			0	1	2	3	4
Ex. 1	5	0.4	2.80	4.00	2.60	3.00	<b>4.40</b>
Ex. 2	6	0.5	2.80	3.60	3.60	<b>1.80</b>	4.00
Ex. 3	7	0.6	3.00	3.20	3.40	2.80	2.20
Ex. 4	8	0.7	3.60	3.80	4.00	3.20	4.00
Ex. 5	9	0.8	3.40	3.40	4.20	3.20	3.80
Ex. 6	10	0.9	3.00	2.80	3.20	3.80	2.60

**Table 4** shows the average coherence rate per topic by each of the experiment. Topic 4 of experiment 1 has the highest average of coherence rate at 4.40. Further investigation found that the validators have chosen only Scale 4 and scale 5 to rate the result of Topic 4 in Experiment 1, hence explaining the highest average. The lowest average is Topic 3 in Experiment 2 at 1.80. The validators have rated the result of this experiment with the lower scales arranging from Scale 1 to Scale 3. The lower scales given has caused the average to fall below 2.00.

**Fig 13.** Average coherence rate per topic for every experiments

**Fig 13** shows the average coherence rate per topic for every experiment. Experiment 3 has the lowest overall coherence rate and shows a slow progress before falling in average coherence rate at Topic 3 and

Topic 4. The highest overall average coherence rate is Experiment 4. The average coherence rate for all topics is stable and only has a slight dip in Topic 3 before rising up again at Topic 4.

**Table 5.** Topic coherence rate per topic across all experiments

Topics	Average Coherence Rate
Topic 0	3.10
Topic 1	3.47
Topic 2	3.50
Topic 3	2.97
Topic 4	3.50

The following **Table 5** shows the average coherence rate per topic across all experiments. Topic with the highest coherence rate is Topic 2 and Topic 4 with an average of 3.50. **Fig 14** shows that Topic 4 has

accumulated the highest number of Scale 4. Topic with the lowest average coherence rate is Topic 3 at 2.97. Topic 3 has a lower Scale 3 accumulation and has an almost accumulation size of Scale 2 and 3.



Fig 14. Rating scales for each of the topic

The graph shows the pattern of average coherence score for each topic across all experiments. The graph shows a rise from Topic 0 to Topic 2. A sudden dip happens in Topic 3 before a sudden rise at Topic 4. The average coherence rate for the entire topics is 3.31.

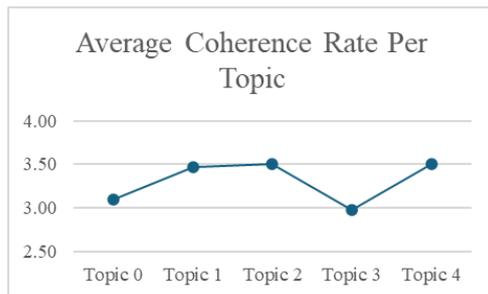


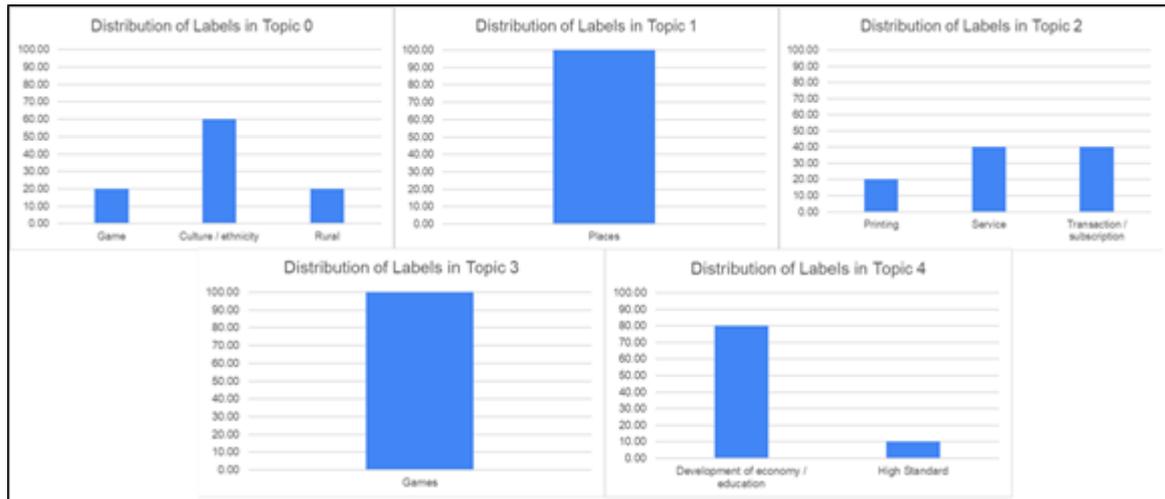
Fig 15. Average coherence rate per topic across all experiments

The accumulation of scale used by the validators shows that the highest Scale 5 was used when rating the results of Topic 2 and the lowest is Topic 0 as shown in Fig 15. The application of Scale 5 shows a drastic influence on the average coherence rate as shown in Topic 2 and Topic 4, both of which accumulate a higher proportion of Scale 5 compared to other topics. The smaller scales such as Scale 1 and Scale 2 dragged the average coherence rate. Topic 3 suffered from a high accumulation of lower scales and lack of Scale 5. Topic 1 also has an almost similar scale

accumulation except for the high accumulation of Scale 4 which slightly increase the average coherence rate compared to Topic 3.

Based on the average coherence rate, Topic 1 has the highest rate at 3.39. Topic 1 has a high application of Scale 3, Scale 4 and 2 applications of Scale 5. The second highest is Topic 4 with a rate of 3.28. Topic 4 has a high application of Scale 4 which influenced the average coherence rate to be higher than other topics except Topic 1. Topic 3 is the lowest in terms of average coherence rate, only at 2.56. Topic 3 also has a high application of Scale 2 which is at 8 applications.

The topic labels have been given by the validator after rating the clusters. Although the topics labelled differently from one validator to another, there are consistent pattern within these labels. For example, based on Topic 1 in the first experiment when  $\alpha=5$  and  $\beta=0.4$ , the validators gave several labels such as Sarikei-Mukah, places, and places in Sarawak. The labels given clearly state something that is related to locations or places. However, there are certain topics that have mixed unrelated labels such as Topic 0 in the experiment that involves  $\alpha=8$  and  $\beta=0.7$ , the validators gave labels such as exports, social justice, and unidentified. The answers for this demonstrate the varied interpretation of the word cluster which is caused by the lack of coherence between words in the cluster.



**Fig 16.** Distribution of labels in topics of Experiment 1.

**Fig 16** shows the distribution of labels for Experiment 1 with 60% of the evaluators label Topic 0 as culture and ethnic related terms while 20% label the topic as game and another 20% as rural. For Topic 1, all evaluators associate it as places. Topic 2 of Experiment 1 is classified as printing by 20% of the

evaluators and 40% each as service and transaction. Topic 3 is unanimously categorized as games or sports. The last topic of Experiment 1 is associated with development of economy or education by 80% of the evaluators and the remaining 20% label the topic as high standard.



**Fig 17.** Distribution of labels in topics of Experiment 2.

As shown in **Fig 17**, Topic 0 of Experiment 2 shows 40% of the evaluators label the topic as wharf and 10% each labelling it as mining town and trading. Another 10% classify the topic as unspecific. For Topic 1, 40% of the evaluators classify the topic as gold and another 40% classify it as history or politic while 20% classify the topic as unspecific. Topic 2 is associated with games or entertainment by 80% of the

evaluators and 20% as shops. Topic 3 of Experiment 2 is classified as unspecific by 60% of the evaluators and 20% each as conjunction or measurement unit picul, indicating the ambiguity of the word cluster. The last topic of Experiment 2 is classified as places by 40% of the evaluators and the remaining 60% classify it as crimes.

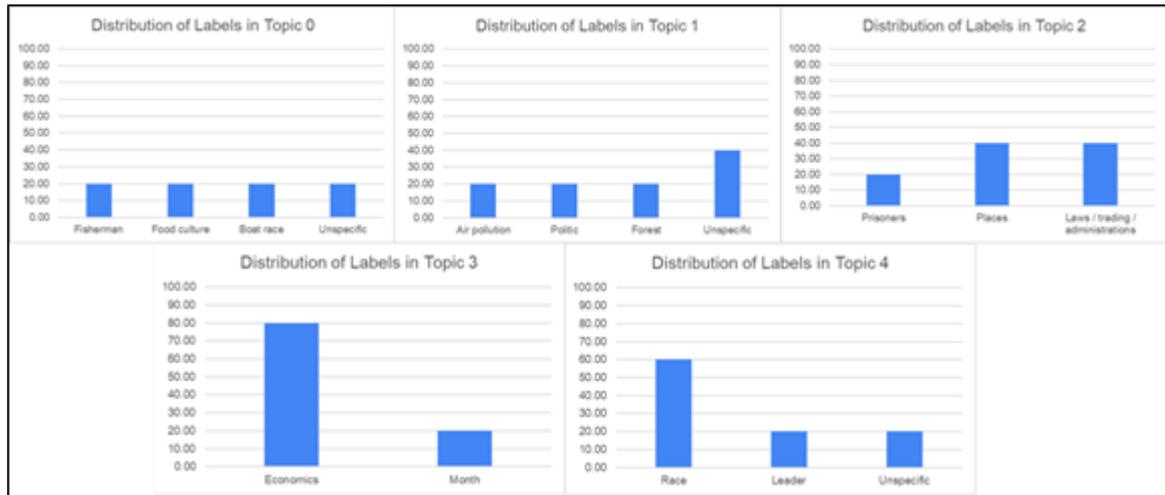


Fig 18. Distribution of labels in topics of Experiment 3.

Throughout Experiment 3 as shown in **Fig 18**, Topic 0 is associated with fisherman, culture, and boat race, with each label selected by 20% of the evaluators. Additionally, 20% of the evaluators label the topic as unspecific. For Topic 1, 20% of evaluators label the topic as air pollution, forestry and politics while the remaining 40% classify the topic as unspecific, reflecting the vagueness of the word cluster. As for Topic 2,

20% of evaluators label the topic as prisoner while 40% of the evaluators categorize it as either places or administrative and laws related. 80% of the evaluators classify Topic 3 as economic related and the remaining classify it as month. The last topic of the experiment is classified as race or competition by 60% of the evaluators while 20% classify it as leader. Another 20% classify the topic as unspecific.

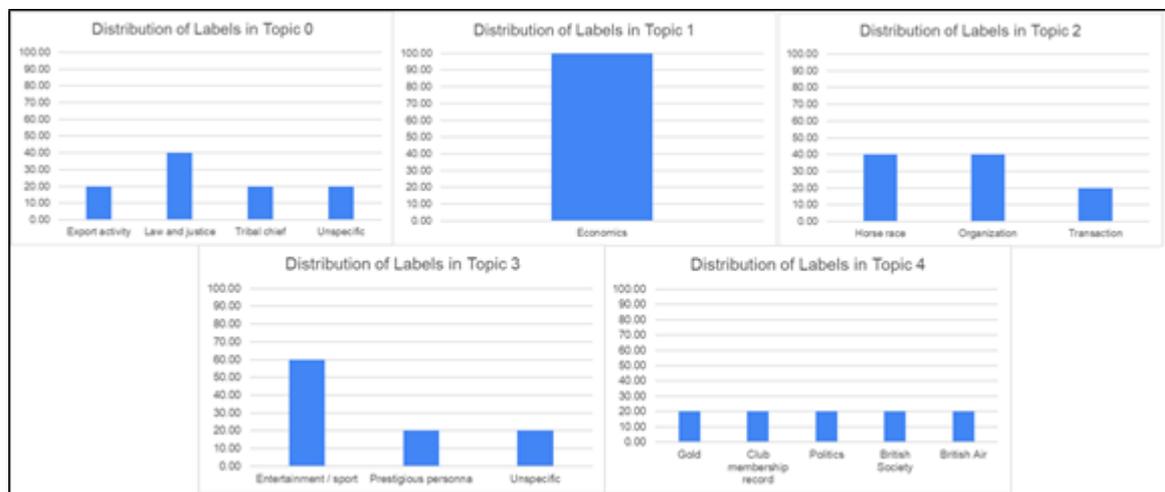


Fig 19. Distribution of labels in topics of Experiment 4.

In Experiment 4 as shown in **Fig 19**, 40% of the evaluators classify Topic 0 as justice or law and 20% each as export activities and tribal chief. Another 20% classify the topic as unspecific as shown in **Fig 19**. For Topic 1, there is 100% agreement among the evaluators, who classify the topic as related to economics. Topic 2 is classified as horse race and organizational

terms by 40% each and the remaining classify the term as transaction. Topic 3 has a 30% label that is related to entertainment or sport and the 20% as prestigious persona, and another 20% as unspecific. Topic 4 has a heterogeneous label with each label representing 20% of the evaluators. The labels include gold, club membership record, politics, British society and British air.

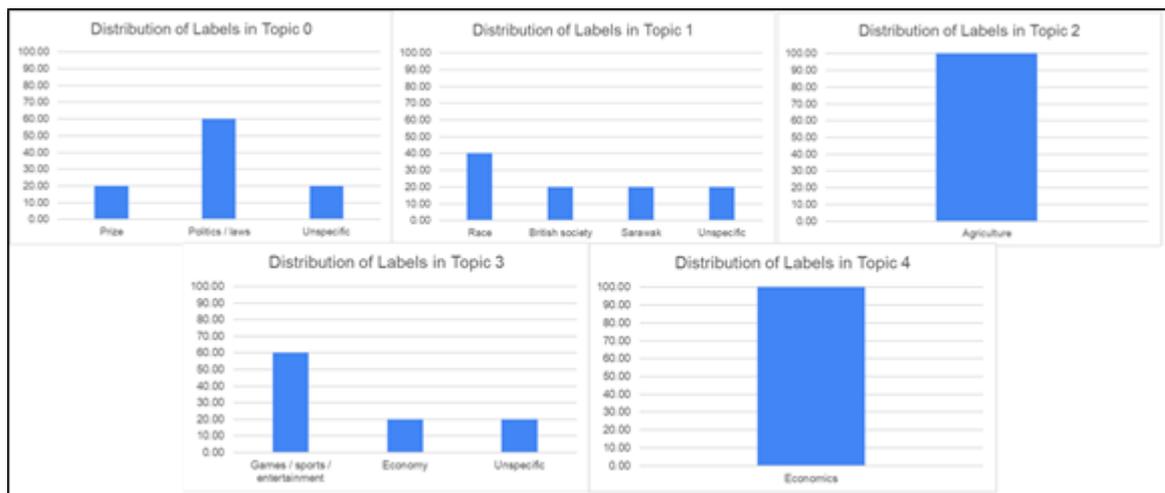


Fig 20. Distribution of labels in topics of Experiment 5

As observed in Fig 20 for Experiment 5, the evaluators identify Topic 0 as prize, politics or law. 20% of the evaluators failed to interpret the result and classify the topic as unspecific. Next, as for Topic 1, 40% classify it as race, 20% as British society and another 20% as Sarawak. The remaining 20% label it as unspecific.

The evaluators unanimously relate Topic 2 as agriculture related term. Topic 3 is labelled as game or sport or entertainment by 60% of the evaluators and 20% each to economy and unspecific. Lastly, all evaluators relate the last topic to agricultural and economic terms.

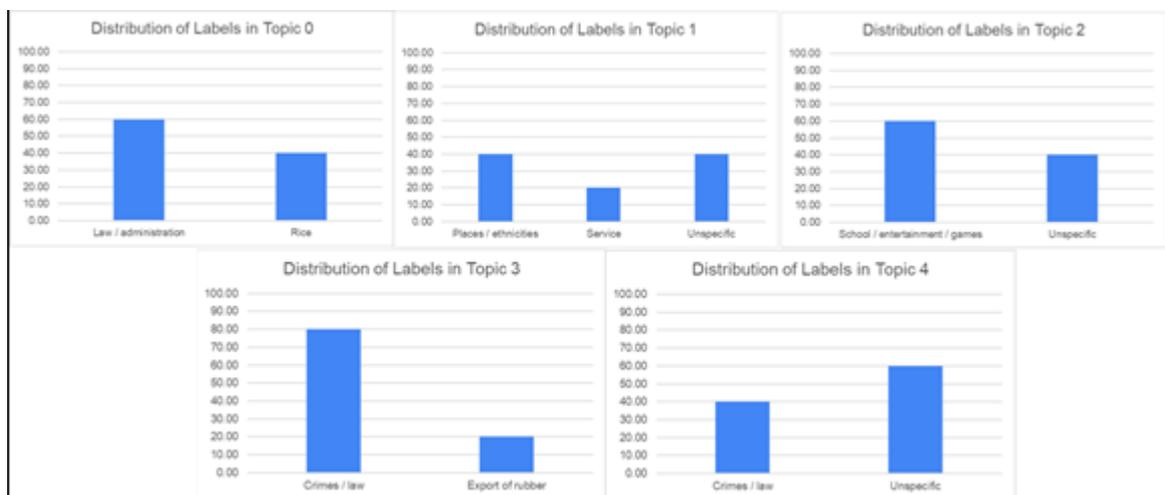


Fig 21. Distribution of labels in topics of Experiment 6.

In the last experiment, 60% of the evaluators classify Topic 0 as laws or administration and the remaining label it as rice as shown in Fig 21. In Topic 1, 20% label it as service, 40% as places or ethnicity, and the remaining 40% as unspecific. Topic 2 comprises 60% of the evaluators classify the topic as school or entertainment, and another 40% as unspecific. 20% of evaluators categorize Topic 3 as exportation of rubber and

the rest as crimes. The last topic shows a high degree of ambiguity as 60% of the evaluators label it as unspecific. The remaining distribution shows the topic is classified as crime and law.

The study shows the most heterogeneous results are obtained from Topic 3 and 4 of the Experiment 4, with each of the labels constituting 20% of the evaluators. Five topics from Experiment 1, 4, and 5 receive

unanimous labelling by the evaluators. The label 'un-specific' label is assigned 21 times with the highest proportion belonging to Topic 3 of Experiment 2 and Topic 4 of Experiment 6, as 60% of the evaluators classify these topics as un-specific.

## 5. Discussion

The perplexity scores increased steadily and peaked at  $\alpha = 10$ . Pinto et. al. (2021) shows the perplexity score tends to be lower which is good when the number of topics and the values of hyperparameters are lower. Thus, the result shows the optimum result of perplexity score is when  $\alpha = 5$ .

The coherence scores for this hyperparameters values suggested that the most coherent result is obtained when the number of topics are 4. The word clouds have a sufficient number of word clusters although it does not necessarily reflect the interpretability of the result. Different hyperparameter produced different word clusters and coherence scores due to the varied distribution of document-topic and topic-word. The coherence score charts across hyperparameters show an increase with  $\alpha = 5$ ,  $\alpha = 9$ , and  $\alpha = 10$  shows a slight dip at topic 4.

The experiment also showed that the perplexity value does not necessarily reflect the interpretability of results. When  $\alpha = 6$ ,  $\beta = 0.5$ , the perplexity indicates a high accuracy. The word cloud of the result showed slightly random words such as iii within the word cluster and words that occurred twice in each of the word clouds which makes the interpretation for each topic rather difficult. Coherence score for this hyperparameters values show that the highest coherent level is when the number of topics is 5. The fifth word cloud shows the words that are linked to criminal activities and reports. Pinto et. al. (2021) presenting the results of coherence score improving as the number of topics are increasing meanwhile showing the perplexity does not necessarily representing the coherence of the result.

Several problems have been encountered during the experiment and evaluation phase of topic modeling by utilizing LDA. The first problem is the requirement of a large dataset to achieve high accuracy. It is problematic especially for the historical data that required familiar data due to the peculiarity of format and spelling although such data is rare and difficult to acquire. Certain words also occurred in two different topics such as the word 'acting' appeared in both Topic 0 and Topic 3 of  $\alpha = 9$  due to the similar problem. The

lack dimensional reduction caused some words to occur in more than one topic cluster. This lack of dimensional reduction compromises the interpretability of topics.

The perplexity score is useful to indicate the capability of the model to handle and predict the unknown and hidden data although this is not sufficient in determining the optimal number of topics. However, the experiment showed that there is no correlation between perplexity score with coherence. A high scored perplexity score does not reflect the interpretability of word cluster within a topic. The perplexity is not reflecting the semantic interpretability and as a result, the perplexity score is not correlated with the manual inspection and interpretation of topic quality (O'Callaghan et. al.,2015), (Yuan et. al., 2023).

Coherence score is observed to be efficient to identify the best number of topics that are able to ensure the interpretability of the result although the status of hidden or unknown data that may be important to the topic cluster is ambiguous. In the experiment, the coherence score is capable of establishing the correlation between the score and the interpretation of topics compared to the perplexity score (Bretsko et. al., 2023).

Based on the manual validation and the original coherence score, it is cleared there are several differences. In the coherence score generated by LDA,  $\alpha=9$  and  $\beta=0.8$  shows the highest range of coherence score for its topics, the manual validation founds that  $\alpha=8$  and  $\beta=0.7$  yield the highest coherence rate. However, the value is not far off as the manual validation for the former one is the second highest.

In conclusion, the LDA has several weaknesses. The repetition of certain words hindered a proper interpretation either from the model itself or based on human inspection. The study also found that the perplexity does not reflect the quality of word cluster.

## 6. Conclusions

The results showed that as the hyperparameter values increased, the model produced more inaccuracies. Specifically, the word clouds became more random, and the word clusters grew sparse. Additionally, the same words frequently appeared in multiple topics, leading to redundancy and making topic interpretation more challenging. The result also showed that as the hyperparameters increase, the result becomes harder to interpret, and the multiple occurrences of the same

words also caused confusion of assigning the name or summarizing the content of the topics.

The coherence values for each hyperparameter setting were also analyzed. The results indicate that the best coherence values typically occur when the number of topics exceeds 3. The highest coherence scores were observed when the number of topics was set to 5, the maximum used in this experiment.

Future studies should explore additional aspects of hyperparameter tuning and evaluation methods to further optimize the number of topics. A key focus could be comparing LDA with more recent algorithms, such as transformer-based topic modeling or other advanced LDA variants. Fine-tuning models for specific datasets may also yield improvements over traditional LDA techniques. It is also feasible to build a fine-tuned model that can be compared with the LDA results.

## Acknowledgement

The authors of this paper would express an appreciation to University Malaysia Sarawak for the support of the publication. The authors acknowledged the financial support from the Ministry of Higher Education Malaysia through Fundamental Research Grant Scheme (FRGS) (F08/FRGS/2029/2020). The authors also extend their gratitude to the validators that contributed to the study by providing their valuable validation and feedback.

## References

- Agarwal, Ankita, Preetham Salehundam, Swati Padhee, William L Romine, and Tanvi Banerjee. (2020). "Leveraging Natural Language Processing to Mine Issues on Twitter during the COVID-19 Pandemic." ArXiv (Cornell University), December. <https://doi.org/10.1109/bigdata50022.2020.937802>.
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan), 993-1022.
- Bretsko, D., Belyi, A., & Sobolevsky, S. (2023, June). Comparative Analysis of Community Detection and Transformer-Based Approaches for Topic Clustering of Scientific Papers. In *International Conference on Computational Science and Its Applications* (pp. 648-660). Cham: Springer Nature Switzerland.

- Dieng, A. B., Ruiz, F. J., & Blei, D. M. (2020). Topic modeling in embedding spaces. *Transactions of the Association for Computational Linguistics*, 8, 439-453.
- Ding, R., Nallapati, R., & Xiang, B. (2018). Coherence-aware neural topic modeling. arXiv preprint arXiv:1809.02687.
- Gertis, E.M. (2021) "Extracting Philosophical Topics from Reddit Posts via Topic Modeling."
- Griffiths, T. L., & Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National academy of Sciences*, 101(suppl\_1), 5228-5235.
- Hasan, M., Rahman, A., Karim, M. R., Khan, M. S. I., & Islam, M. J. (2021). Normalized approach to find optimal number of topics in Latent Dirichlet Allocation (LDA). In *Proceedings of International Conference on Trends in Computational and Cognitive Engineering: Proceedings of TCCE 2020* (pp. 341-354). Springer Singapore.
- Lee, J., Kang, J. H., Jun, S., Lim, H., Jang, D., & Park, S. (2018). Ensemble modeling for sustainable technology transfer. *Sustainability*, 10(7), 2278.
- Mohr, J. W., & Bogdanov, P. (2013). Introduction—Topic models: What they are and why they matter. *Poetics*, 41(6), 545-569.
- Muhajir, D., Akbar, M., Bagaskara, A., & Vinarti, R. (2022). Improving classification algorithm on education dataset using hyperparameter tuning. *Procedia Computer Science*, 197, 538-544.
- Newman, D., Bonilla, E. V., & Buntine, W. (2011). Improving topic coherence with regularized topic models. *Advances in neural information processing systems*, 24.
- O'Callaghan, D., Greene, D., Carthy, J., & Cunningham, P. (2015). An analysis of the coherence of descriptors in topic modeling. *Expert Systems with Applications*, 42(13), 5645-5657.
- Odden, T. O. B., Marin, A., & Caballero, M. D. (2020). Thematic analysis of 18 years of physics education research conference proceedings using natural language processing. *Physical Review Physics Education Research*, 16(1), 010142.

- Panichella, A. (2021). A Systematic Comparison of search-Based approaches for LDA hyperparameter tuning. *Information and Software Technology*, 130, 106411.
- Papadimitriou, C. H., Tamaki, H., Raghavan, P., & Vempala, S. (1998, May). Latent semantic indexing: A probabilistic analysis. In *Proceedings of the seventeenth ACM SIGACT-SIGMOD-SIGART symposium on Principles of database systems* (pp. 159-168).
- Péladeau, N., & Davoodi, E. (2018). Comparison of latent Dirichlet modeling and factor analysis for topic extraction: A lesson of history.
- Pinto Gurdial, L., Morales Mediano, J., & Cifuentes Quintero, J. A. (2021). A comparison study between coherence and perplexity for determining the number of topics in practitioners interviews analysis.
- Seymore, K., & Rosenfeld, R. (1997). Large-scale topic detection and language model adaptation. Carnegie-Mellon University. Department of Computer Science.
- Steyvers, M., & Griffiths, T. (2007). Probabilistic topic models. *Handbook of latent semantic analysis*, 427(7), 424-440.
- Teh, Y., Newman, D., & Welling, M. (2006). A collapsed variational Bayesian inference algorithm for latent Dirichlet allocation. *Advances in neural information processing systems*, 19.
- Wallach, H., Mimno, D., & McCallum, A. (2009). Rethinking LDA: Why priors matter. *Advances in neural information processing systems*, 22.
- Watanabe, K., & Baturo, A. (2023). Seeded Sequential LDA: A Semi-supervised Algorithm for Topic-specific Analysis of Sentences. *Social Science Computer Review*, 08944393231178605.
- Xue, J., Chen, J., Chen, C., Zheng, C., Li, S., & Zhu, T. (2020). Public discourse and sentiment during the covid 19 pandemic: using latent dirichlet allocation for topic modeling on twitter. *Plos One*, 15(9), e0239441. <https://doi.org/10.1371/journal.pone.0239441>
- Yang, L., & Shami, A. (2020). On hyperparameter optimization of machine learning algorithms: Theory and practice. *Neurocomputing*, 415, 295-316.
- Yuan, M., Lin, P., Rashidi, L., & Zobel, J. (2023). Assessment of the Quality of Topic Models for Information Retrieval Applications.
- Zhou, Z., Liu, M., & Tao, Z. (2023). Quantitative Analysis of Citi's ESG Reporting: LDA and TF-IDF Approaches. *Financial Engineering and Risk Management*, 6(3), 53-63.

#### AUTHOR BIOGRAPHIES



**Muhammad Abdullah Yusof** is a master student at University Malaysia Sarawak in Malaysia. Before then, he had industrial experience in the Western Digital. He received his bachelor's degree from Universiti Malaysia Sarawak. His areas of interest include Natural Language Processing, Text Processing, Artificial Intelligence, and Computational Linguistics.



**Suhaila Saeed** is a senior lecturer at the Faculty of Computer Science and Information Technology, Universiti Malaysia Sarawak. She holds a Bachelor's (2001) and Master's (2002) in Computer Science from Universiti Putra Malaysia. Suhaila worked as a researcher at MIMOS Berhad, specializing in knowledge engineering and machine translation. In 2007, she co-founded the Sarawak Language Technology (SaLT) research group with Professor Dr. Alvin Yeo Wee, focusing on digitizing and preserving 63 Sarawak indigenous languages using computational linguistics and natural language processing techniques. Her research interests include computational morphology, NLP, and preserving under-resourced languages. The area of interest is preserving under-resourced languages through speech and text processing.

# Iris liveness detection for biometric access control system in smart home security using deep convolutional neural network.

Yash G. Waghmare<sup>1</sup>, Sudeep D. Thepade<sup>2\*</sup>

<sup>1</sup> Department of Computer Engineering, Pimpri Chinchwad College of Engineering, Pune, India

\* Corresponding author E-mail: sudeepthepade@gmail.com

(Received 4 March 2024; Final version received 26 May 2024; Accepted 3 September 2024)

## Abstract

Biometric access control systems are essential for enhancing the security of smart homes. Among the various biometric modalities, iris recognition is a promising option due to its high accuracy and contactless nature. Nevertheless, presentation attacks, which try to trick the system using artificial or fake irises, may deceive iris recognition systems. To counter this threat, iris liveness detection (ILD) techniques are employed to distinguish between real and fake irises. In this paper, a novel and robust ILD method that combines handcrafted features and deep learning based features is proposed. The proposed method's performance is assessed across multiple machine learning classifiers and contrasted with existing ILD methods. The experimental results show that the proposed method achieves the lowest Average Classification Error Rate (ACER) values of 1.1% and 0.3% on the IIIT-D and Clarkson 2015 datasets, respectively, demonstrating its effectiveness and robustness against different types of presentation attacks.

*Keywords: ILD, Inception v3, Haralick, GLCM.*

## 1. Introduction

An increasing number of security systems are being developed in a fast-paced environment where security has become paramount. Creating a strong system is crucial to enhancing security in smart homes. In the realm of smart homes, home security is a must to improve the safety of the residents. Based on the individual features of each resident, a biometric access control system decides whether or not to grant them entry into the house. Because biometrics are unique and very accurate, they are a popular choice for high-risk areas when it comes to security measures. Specialized scanners and recorders acquire physical attributes like fingerprints, irises, palms, faces, or voices to instrument such security systems. There have been many biometric access control systems that use fingerprint, voice, and other biological traits but they have their risks and limitations. But the Iris Iris-based biometric access control system is contactless and each resident has a unique iris, the contactless iris recognition system is the best biometric access control option. Iris liveness detection is being integrated into the biometric access control for smart homes as an additional layer of protection that will assist thwart spoofing attempts. Iris liveness detection can withstand spoofing

attempts from printed iris images, contact lenses, and artificial eyes with false iris patterns. ILD strengthens the biometric access control system's security, which is crucial for the security of smart homes in this day and age.

A primary concern for iris recognition systems lies in the threat posed by Presentation Attack Instruments (PAI), which manipulate the system by introducing a counterfeit version of the authentic biometric attribute to the iris detection sensors. This deceptive technique leads the system to erroneously identify an unauthorized user as a legitimate one as suggested by Khade etc. (2022). Therefore, it is thought that iris liveness detection is a useful way to lessen the threats that Presentation Attack Instruments (PAI) in iris recognition systems provide. The purpose of iris liveness detection mechanisms is to differentiate between real live irises and different presentation attack techniques such as using 3D models, printed images, and transparent, colored, or contact lenses. Iris liveness detection greatly lowers the possibility of successful spoofing efforts when it is included in biometric access control systems. Sophisticated methods like motion detection, texture analysis, and machine learning algorithms are utilized to discern the minute details that differentiate an actual iris from a fake or synthetic one. This increases the iris recognition systems' overall security

and dependability and strengthens their defenses against fraud and illegal access.

There have been many methods for effective iris liveness detection which include 2 step traditional machine learning methodology, where hand-crafted features are extracted in the first step, and in the second step, these features are fed to classifiers, or using the fusion of multiple handcrafted features. The primary contributions discussed in the paper are

- Proposed an advanced and reliable approach for ILD that combines manually created and deep learning-based features to enhance smart home security.
- Exploration of Inception v3 pre-trained CNN model for extraction of the global features, and exploring GLCM for extracting texture-based Haralick features.
- Evaluating the performance improvement of proposed ILD by various machine learning (ML) classifiers and majority voting-based ensembles.

## 2.Literature Survey

Smart home security iris recognition systems need to possess the capability to detect and differentiate between different kinds of iris spoofing attacks. The literature does not have many algorithms that can determine whether an iris is live for a reliable security system in a smart home. The strategy suggested by Ishengoma (2014) combines fingerprint and iris recognition technology to improve smart home security. The system compares the two iris using hamming distance—the one that is collected and the one that is kept in the database. However, this method requires a lot of pre-processing and does not support liveness detection, which is a crucial component of the rapidly advancing field of technology.

Only one type of iris spoofing attack can be detected by the majority of presentation attack detection techniques used today. To identify template attacks on iris recognition systems, D. Shanmugapriya etc. (2023) employed machine learning and deep learning techniques. The suggested technique detects Iris template attacks by using Convolutional Neural Networks (CNN) and Logistic Regressions (LR). An accuracy of 98.75% is obtained using the CASIA-IrisV1 dataset, which is higher than LR. Applying the max pooling property also improves accuracy; this resulted in a 100% accuracy rate. A comparison is made between the suggested approach and current methods, including Scale-Invariant Feature Transform (SIFT), Histogram Oriented Gradients (HOG), and Local Binary

Pattern (LBP). However, only one kind of presentation template attack is used to test the suggested approach. And confirmed using a single dataset, which lessens its ILD robustness.

Khade etc. (2022) employed many deep convolution networks to identify live iris. The paper applied transfer learning approaches to iris liveness detection utilizing five pre-trained models: Inception v3, Resnet50, Densenet121, VGG-16, and EfficientNetB7. The IIIT contact lens iris dataset, the ND Iris3D 2020 dataset, and the LivDet-Iris 2015 dataset are used to assess the performance of the pre-trained models. These datasets are compared using criteria including accuracy, precision, recall, f1-score, apcer (attack presentation classification error rate), npcer (normal presentation classification error rate), and error rates. According to the suggested methodology, the pre-trained models can identify the iris region's nanostructures with great accuracy and effectiveness. On the ND Iris3D 2020 dataset, the results demonstrated that the EfficientNetB7 network surpasses the other networks, achieving 99.97% accuracy and making the fewest errors when guessing whether the image was real or not. Transfer learning improves the effectiveness of biometric authentication by lowering the amount of calculations needed for model training.

Tapia etc. (2021) introduced a method that utilizes both a fine-tuned MobileNetV2 network and a newly developed network trained from scratch. Their approach involved utilizing the LivDet-Iris 2020 competition dataset in addition to the Iris-CL1, Iris-printed-CL1, and Warsaw-BioBase-Post-Mortem Iris v3.0 dataset, which collectively contain various presentation attack images. To prepare the images for analysis, they underwent preprocessing using the contrast-limited adaptive histogram equalization (CLAHE) algorithm. Additionally, a weighted factor was applied to each class to enhance grayscale intensity and balance the dataset. The proposed framework used ImageNet weights for transfer learning of the MobileNetV2 model and the scratch network for better classification. The proposed strategy mainly focuses on classifying bona fide images and then classifying attack presentation images. Hence the approach it follows is to first train the network with two classes and then train the scratch network with three and four classes. The BPCER values obtained for two, three, and four classes scenarios are 0.99%, 0.16%, and 0.83% respectively.

The 15-layer CNN model, which includes the final Softmax layer for classification, was proposed by

Winston etc. (2022) for the iris recognition system. The accuracy achieved by the proposed model is 95.16% on the IIIT-D dataset. Hybridization with KNN and SVM statistical classifiers is used to further improve the suggested CNN model, with accuracies of 86% and 97.8%, respectively, obtained. Nevertheless, the robustness of ILD is diminished when a single dataset is used. A few recent studies have demonstrated as by Verma etc. (2023), that Two pre-trained deep convolutional neural networks (DCNNs) are combined using a proportionate score-level fusion approach. With this methodology, cross-database validation yields an average classification error rate (ACER) of 9.72%, while known attack scenarios yield an ACER of 0.6%. The evaluation was conducted using the NDCLD 2015 and Notre Dame 2017 datasets. Furthermore, for micro-textural analysis, a method is proposed by Kaur (2024) that entails capturing local characteristics that are invariant to scale, rotation, and translation. This is accomplished by encoding these features using Lehmer coding, after which they are converted into histograms that function as feature descriptors. The IIITD-CLI and IIITD-IIS datasets yielded ACER values of 1.45% and 1.61%, respectively, after evaluation. Furthermore, the dataset from Clarkson 2015 showed an ACER of 2.10%. A fascinating strategy that is discussed by Choudhary etc. (2023) makes use of the well-known Friedman test-dependent feature selection method. This technique yields a refined set of features by determining the best subset of  $k$  features out of  $N$ . When combined with score-level fusion, this choice works well for precisely predicting ILD.

The method of fusing VGG features and Multi-level Haralick features is proposed by Yadav etc. (2018). Grey-level co-occurrence matrices (GLCM) are utilized in the computation of Haralick features, which are local textural features. The redundant discrete wavelet transform (RDWT) is used in the suggested method to extract the Haralick features. Additionally, multi-level RDWT is used, offering supplementary data on image characteristics at various scales. Coarse iris segmentation is carried out before iris

feature calculation and extraction. Additionally, the VGG model—a pre-trained, 16-layer CNN model—extracts deep learning characteristics. Principal component analysis (PCA) reduces these features. Artificial neural networks are used to combine the Haralick and VGG features for classification (ANN). The LivDet2013 (Warsaw Subset) dataset, NDCLD-2013, NDCLD-2015, and the Combined Spoofing Database (CSD) are used for the evaluation. The suggested algorithm produces a minimum of 1.01% overall error. Nevertheless, the approach involves pre-processing, which takes a considerable amount of time and lessens the ILD's robustness.

The review of the literature emphasizes how iris recognition technologies for smart home security are developing and how difficult it is to combat different iris spoofing techniques. Robust liveness detection is still a critical requirement, even if current methods use a variety of techniques, including machine learning, deep learning, transfer learning, and various iris recognition. The research presented here demonstrates advances in live iris identification with pre-trained models, iris template attack detection, and new techniques such as fine-tuned MobileNetV2 networks. However given the shortcomings of existing approaches—such as decreased ILD resilience, reliance on particular datasets, and preparation overhead—it is clear that a thorough solution necessitates more investigation. Future studies should aim to improve the generalization and overall efficacy of iris-liveness detection for smart home security, making them resistant to various presentation attacks and requiring the least amount of processing resources. More multidisciplinary work in the fields of feature fusion, and deep learning will be necessary to create more dependable and safe smart home environments.

Table 1 displays a comparative analysis of existing ILD methods, offering a thorough examination of their methodologies, results, and datasets utilized. This comparison delves into the techniques employed and the outcomes achieved by each approach, providing insights into their respective strengths and limitations within the field of iris liveness detection.

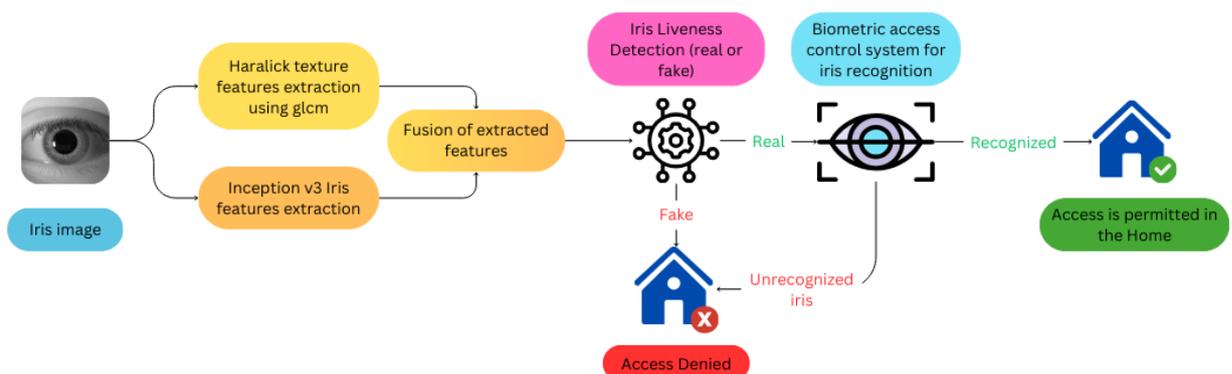
**Table 1.** Comparative analysis of existing relevant ILD methods.

Relevant method reference	Technique used	Results	Datasets used
Khade etc. (2022)	Transfer learning approaches to ILD utilizing five pre-trained models: Inception v3, Resnet50, Densenet121, VGG-16, and EffecientNetB7.	Highest accuracy gains: 99.97%.	IIIT contact lens iris dataset, the ND Iris3D 2020 dataset, and the LivDet-Iris 2015 dataset.
Tapia etc. (2021)	Fine-tuned MobileNetV2 network and a new network which is trained from scratch for ILD and CLAHE for preprocessing.	Lowest BPCER value: 0.16%.	LivDet-Iris 2020, Iris-CL1, Iris-printed-CL1 and Warsaw-BioBase-Post-Mortem Iris v3.0 dataset.
Winston etc. (2022)	The 15-layer CNN model with KNN and SVM statistical classifiers.	Highest accuracy gains: 97.8%.	IIIT-D dataset
Kaur (2024)	Capturing local characteristics by micro-textural analysis using Lehmer coding which is then converted into a histogram that further functions as a feature descriptor.	Lowest ACER yield: 1.45%	IIITD-Contact Lens, IIITD-Iris Spoofing, Clarkson-2015, Warsaw-2015, and fingerprint spoofing databases: LivDet-2013 and LivDet-2015.
Yadav etc. (2018)	Method of fusing VGG features and Multi-level Haralick features for identifying Iris presentation attack.	Minimum overall error rate: 1.01%	The LivDet2013 (Warsaw Subset) dataset, NDCLD-2013, NDCLD-2015, and the Combined Spoofing Database (CSD).

### 3. Proposed Methodology

The proposed architecture, illustrated in Fig 1, integrates feature harmonization by combining Haralick features, derived from the grey-level co-occurrence matrix (GLCM), with features from Inception V3 (Szegedy etc., 2016). The GLCM captures the

frequency of neighboring grey levels within the image. The biometric access control system checks if the iris is live before identifying the resident's iris and granting permission to the house. Thus making the system resistant to almost every possible form of attack. This research primarily focuses on the harmonization of Inception v3 features and Haralick texture characteristics for Iris liveness detection.



**Fig 1.** Proposed ILD method for Biometric access control system using harmonization of Haralick texture features and Inception v3 global features.

### 3.1. Inception v3 feature extraction

The proposed method utilizes Inception v3, a pre-trained deep learning model based on Convolutional Neural Networks (CNNs) originally designed for image classification tasks. Inception v3 is trained on the ImageNet dataset, comprising over a million images, and can be applied to classification tasks on specific datasets through transfer learning as stated by Khade etc. (2022b) & Impedovo etc. (2021). This approach significantly reduces training time while ensuring enhanced performance on the target dataset. The model is comprised of multiple inception blocks, each containing various convolutional layers with different filter sizes. Notably, global average pooling (GAP) replaces the traditional fully connected layers found in conventional neural networks at the end of the

architecture. The proposed approach involves tuning various parameters, including learning rate, regularization techniques, the number of training epochs, architectural modifications for feature extraction, and the number of unfrozen layers as stated by Kimura etc. (2020) & Yan etc. (2018). Choosing which layers to unfreeze during fine-tuning is crucial, depending on the dataset size and the similarity between pre-training and target tasks. Fig 2 illustrates the unfreezing of inception block C, which encompasses 18 out of the total 48 convolution layers present in Inception v3. The Adam optimizer is used in the fine-tuning process, which adapts to the different features of the parameters as suggested by Zhou etc. (2024). In conclusion, this method optimizes the model's performance by fine-tuning specific parameters and strategically unfreezing layers during training.

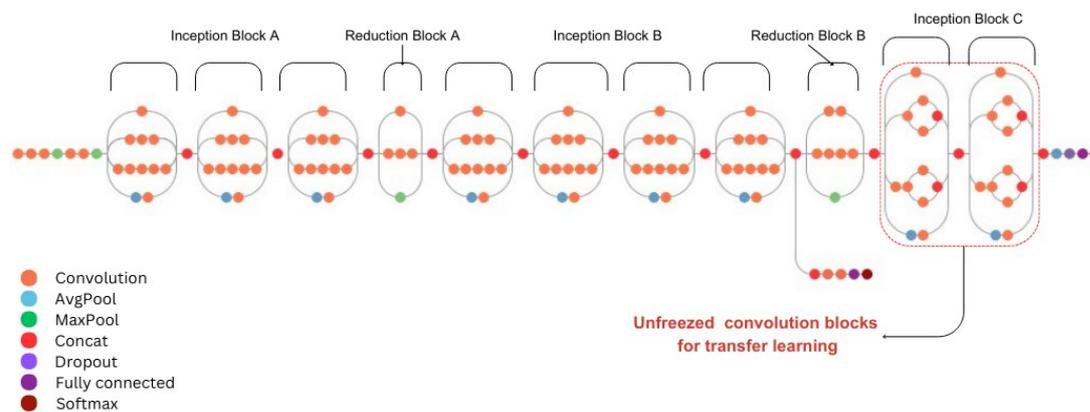


Fig 2. Inception v3 architecture model used for transfer learning

### 3.2. Haralick feature extraction using GLCM

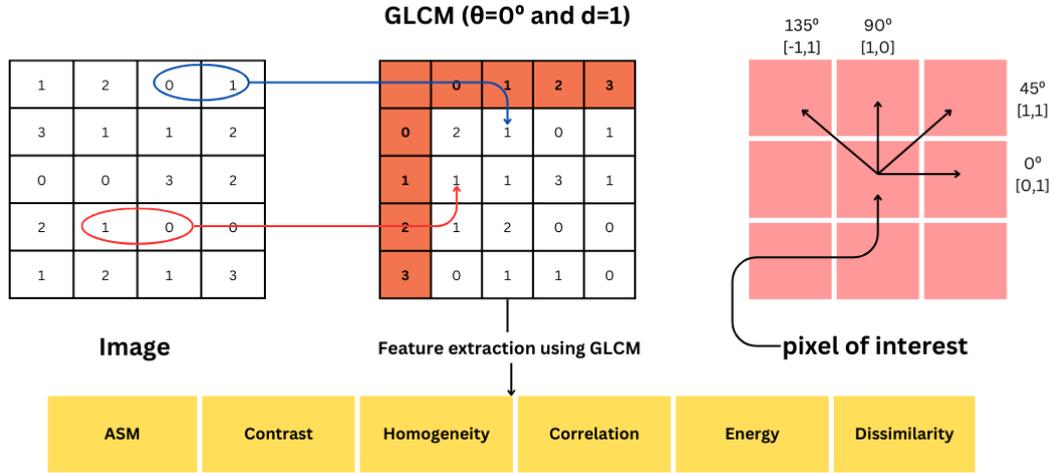
Feature extraction is a technique that simplifies the data by selecting and combining the most important variables into features. Features are easier to process and describe the data accurately. Feature extraction is useful for image analysis, where the data has many pixels and details. Haralick texture features are one type of feature extraction for images as introduced by Haralick etc. (1973). They measure the texture of the image as suggested by Toennies (2024), which is how the gray levels vary and repeat in the image. As stated by Khade etc. (2021a) & Khade etc. (2021b), we need to create a Grey Level Co-occurrence Matrix (GLCM) to calculate Haralick texture features. This is a matrix that counts how often two neighboring pixels

have the same gray level. The GLCM has the same size as the number of gray levels in the image. For example, if the image has 256 gray levels, the GLCM will be a 256 x 256 matrix. The GLCM depends on how we reduce the gray levels of the image, which is called quantization. Different quantization methods can give different results, so we need to use the same method to compare Haralick features as suggested by Li etc. (2021) & Porebski etc. (2008). There are also some methods to make Haralick features independent of the quantization method. Haralick features are calculated from the GLCM using some mathematical formulas. There are 14 Haralick features, each measuring a different aspect of the texture, such as contrast, energy, entropy, homogeneity, etc.

Angle and distance are important parameters to calculate GLCM because they define the spatial relationship between two pixels in the image. The GLCM

counts how often two neighboring pixels have the same gray level, but the neighbor can be defined in different ways depending on the angle and distance. For example, if the angle is 0 degrees and the distance is 1, the neighbor is the pixel to the right of the current pixel. If the angle is 45 degrees and the distance is 2, the

neighbor is the pixel two steps diagonally up and to the right of the current pixel. By changing the angle and distance, we can capture different patterns and textures in the image. An illustration of GLCM for feature extraction is shown in Fig 3.



**Fig 3.** Haralick features extraction using GLCM.

In the proposed methodology, for calculating the gray level co-occurrence matrix distance taken is 4, and the angles considered are 0, 45, 90, and 135. By using a distance of 4 units, the spatial relationship between pixels is considered that be not too close or too far from each other. This can help to balance the trade-off between the resolution and noise of the feature extraction. Along with this Haralick features considered for feature extraction are Angular Second Moment (ASM), Contrast, Homogeneity, Correlation, Energy, and Dissimilarity. They are calculated using some mathematical formulas that involve the GLCM values and their probabilities.

- a) **Angular Second Moment (ASM):** It measures the texture uniformity or smoothness. It is high when the GLCM has a few dominant values and low when the GLCM is more uniform. It ranges from 0 to 1.

$$ASM = \sum_{i,j=0}^{N-1} Q_{i,j}^2 \quad (1)$$

- b) **Contrast:** This feature measures the intensity contrast or variation between a pixel and its neighbor. It is high when the GLCM has high values away from the diagonal and low when the GLCM has high values near the diagonal. It ranges from 0 to  $(N-1)^2$ , where N is the number of gray levels.

$$Contrast = \sum_{i,j=0}^{N-1} Q_{i,j} (i-j)^2 \quad (2)$$

- c) **Homogeneity:** This feature measures the texture homogeneity or similarity. It is high when the GLCM has high values near the diagonal and low when the GLCM has high values away from the diagonal. It ranges from 0 to 1.

$$Homogeneity = \sum_{i,j=0}^{N-1} \frac{Q_{i,j}}{1 + (i-j)^2} \quad (3)$$

- d) **Correlation:** This feature measures how correlated a pixel is to its neighbor. It is high when the GLCM has high values for pixels with similar gray levels and low when the GLCM has high values for pixels with different gray levels. It ranges from -1 to 1.

$$Correlation = \sum_{i,j=0}^{N-1} Q_{i,j} \left[ \frac{(i-\mu_i)(j-\mu_j)}{\sqrt{(\sigma_i^2)(\sigma_j^2)}} \right] \quad (4)$$

Where,

$\mu$ : mean,  $\sigma^2$ : variance.

- e) **Energy:** This feature is the same as ASM. It is the square root of it. It is also called Uniformity or Angular Second Moment Normalized. It ranges from 0 to 1.

$$Energy = \sqrt{ASM} \quad (5)$$

- f) **Dissimilarity:** This feature is the opposite of Homogeneity. It measures the texture dissimilarity or difference. It is high when the GLCM has high values away from the diagonal and low when the GLCM has high values near the diagonal. It ranges from 0 to N-1, where N is the number of gray levels.

$$Correlation = Q_{i,j}|i - j| \quad (6)$$

Where, the notation in the equation (1), (2), (3), (4), (5) and (6) represents

$Q_{i,j}$ : Probability of element  $i, j$  of glcm,

N: Number of gray levels of images.

### 3.3. Harmonization of Inception v3 features and Haralick features

The goal of harmonizing Inception v3 and Haralick features is to enhance ILD performance by merging the benefits of both feature types: Peng etc. (2021) proposed global and local texture features. GLCM measures the spatial correlations between pixel intensities, providing information about the spatial arrangement and correlation of pixel pairs that improves texture comprehension. These features contain the shape, texture, and edge information of the iris image. By harmonizing these two feature types, the proposed method can benefit from the complementary information of both global and local features and enhance the discriminative power of the feature representation. Moreover, the harmonization of Inception v3 and Haralick features also increases the robustness of iris-liveness detection against various spoofing attacks.

## 4. Experimental Configuration

The experimental setup utilized for the exploration of the proposed work is put forth in this section.

### 4.1. Datasets

For empirical validation, the proposed methods were applied to two popular datasets. Illustration of these two datasets are shown in Fig 4.

#### a) LiveDet-Iris 2015: Clarkson Dataset

The LiveDet-Iris 2015 is a dataset for iris liveness detection provided by Yambay etc. (2015).

The collection, which Clarkson University created, includes pictures of actual and artificial irises that were taken using two separate sensors. This dataset consists of 1713 bitmap images with printed, pattern, and live classes taken with a Dalsa sensor. and 1308 bitmap pictures with the printed, pattern, and live classes taken with an LG sensor as given by Wagh & Thepade (2024). The LivDet-Iris 2015 competition utilized the dataset to assess how well different iris-liveness detection algorithms performed.

#### b) IIITD Contact Lens Iris Dataset

An iris images dataset gathered by the Indraprastha Institute of Information Technology, Delhi (IIIT-D) Image Analysis and Biometric Lab is referred to as the IIIT Delhi Contact Lens Iris Dataset as provided by Kohli etc. (2013) & Yadav etc. (2014). The dataset includes iris images of many participants taken with two distinct iris sensors, both with and without contact lenses. 2702 bitmap pictures with the classifications of colored, normal, and transparent were recorded by the Congent sensor. likewise, 3432 bitmap pictures with colored, normal, and transparent classes were recorded by the Vista sensor as given by Wagh & Thepade, (2024). The database was made to investigate how contact lenses affect the accuracy of iris identification and to produce lens detection algorithms.

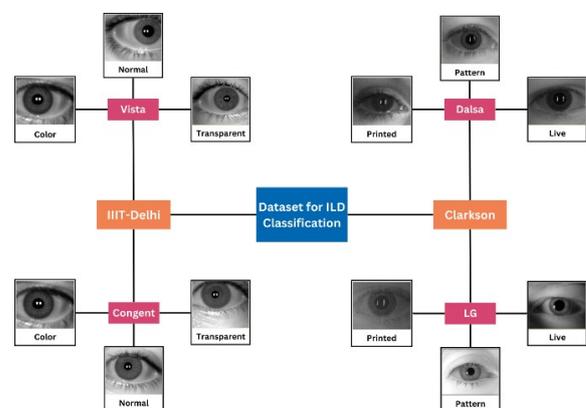


Fig 4. Illustrative images of the two datasets: IIIT-Delhi and Clarkson dataset used for the empirical validation.

### 4.2. Model Evaluation

An essential part of the model evaluation is the examination of features from Inception v3 and the addition of haralick features with their harmonization from the image datasets of IIIT Delhi contact lens iris dataset and Clarkson 2015. This thorough analysis is carried out using a wide range of machine learning

classifiers, all of which are implemented on the Weka platform and include Random Forest, Random Tree, J48, Lazy IBK, Logistic function, SMO function, Decision Table, and Simple Logistic. With the help of this multimodal method, an in-depth understanding of the model's performance across a range of classifiers is achievable, guaranteeing a solid study of the model's effectiveness and adaptability in managing a variety of datasets.

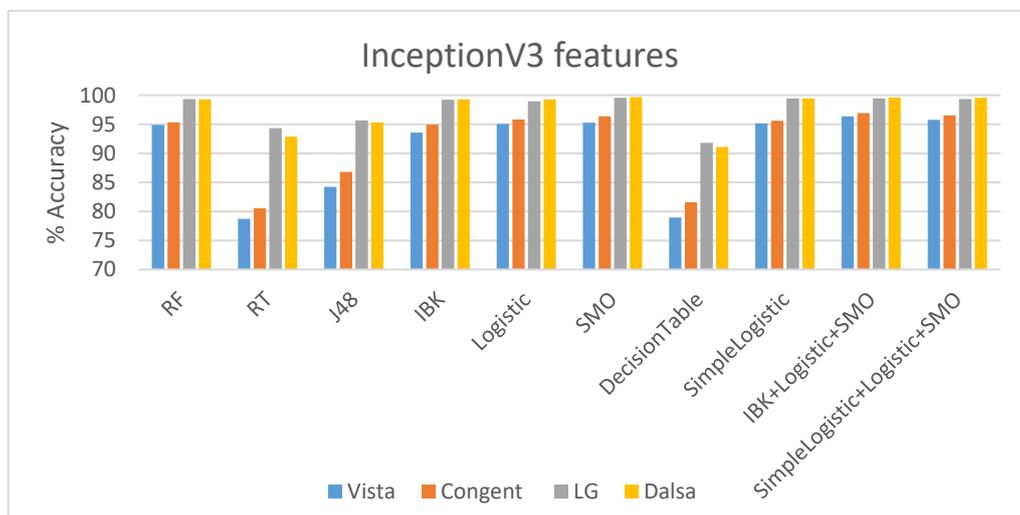
## 5. Results and Discussion

The experiments show that the proposed features and classifiers are effective for iris recognition in the presence of contact lenses. Two iris databases are used: Clarkson 2015 and IIITD contact lens iris, which contain images captured by different sensors and with various types of contact lenses. Haralick, InceptionV3, and their fusion features are extracted from the iris images and different machine learning classifiers and their ensembles are applied to classify them. It is found that the fusion of Haralick and InceptionV3 features achieves the highest accuracy in both databases, indicating that the fusion strategy can capture complementary information from the two types of features. It is also observed that the ensembles of IBK+Logistic+SMO and SimpleLogistic+Logistic+SMO

perform better than the individual classifiers for most of the subsets, suggesting that the ensembles can reduce the variance and improve the generalization of the classification. The results demonstrate the robustness and reliability of the approach for iris recognition in challenging scenarios.

### 5.1. Experiment A: Result analysis on Inception v3

Features retrieved from the global average pooling layers were used as input parameters for the classifiers and ensembles employed, using the pre-trained InceptionV3 CNN model. With two subsets, LG, Dalsa, and Vista, Congent, respectively, the datasets from Clarkson (2015) and IIITD are used to illustrate the accuracy gains achieved by the various classifiers in Fig 5. The experiment reveals that SMO classifiers achieved the greatest accuracy of 99.64% on the Dalsa subset, while the SimpleLogistic+Logistic+SMO ensemble attained the highest accuracy on the same subset of 99.59%. Additionally, the ensemble of IBK+Logistic+SMO had the highest average accuracy over the dataset, 98.08%. As an outcome, the ensemble outperformed other classifiers on average.



**Fig 5.** Accuracies achieved by different classifiers and ensembles of best-performing classifiers for the datasets of Clarkson 2015 and IIITD, with their two subsets, LG, Dalsa, and Vista, Congent, respectively on InceptionV3 features.

### 5.2. Experiment B: Result analysis on Haralick features

For experimentation and result analysis, Haralick features with GLCM are used for various classifiers

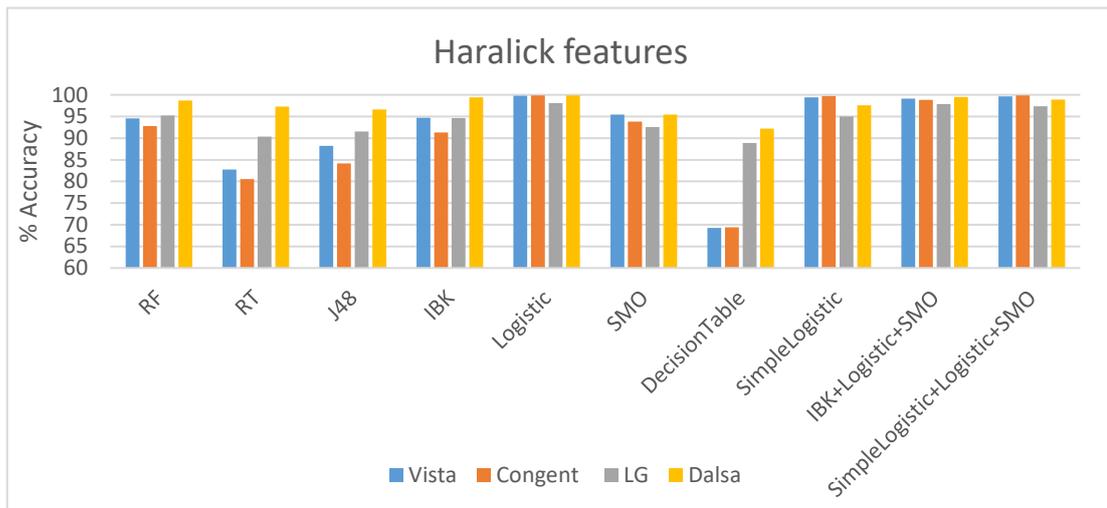
and their ensembles. These characteristics have shown to be useful for determining the texture and deriving features from it. The accuracy that classifiers achieved for these features is seen in Fig 6. The Vista subset of the IIITD dataset had the greatest accuracy of 99.79%

using a logistic function classifier. Additionally, ensembles have outperformed SimpleLogistic+Logistic+SMO, achieving 99.62% accuracy.

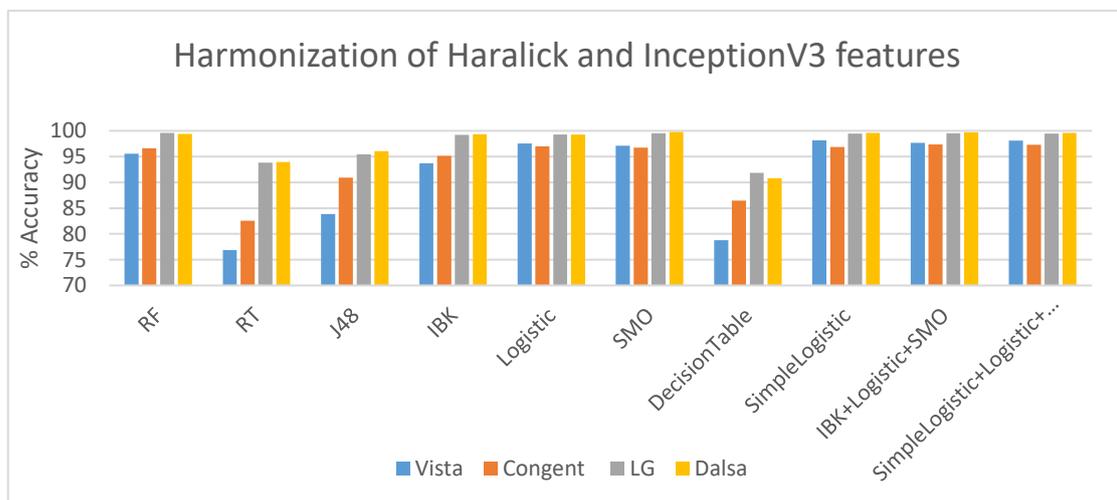
### 5.3. Experiment C: Result analysis on Harmonization of Inception v3 and Haralick features

In this experiment, the harmonization of Haralick and InceptionV3 features is tested. The classifiers

employ these attributes as input parameters. And an analysis of the outcomes is being done as a result. The outcomes for these harmonized features are illustrated in Fig 7. Surprisingly, harmonization has outperformed individual features on average. The highest average accuracy achieved over the dataset by SimpleLogistic+Logistic+SMO is 98.60%. The maximum accuracy attained using SMO classifiers is 99.76%.



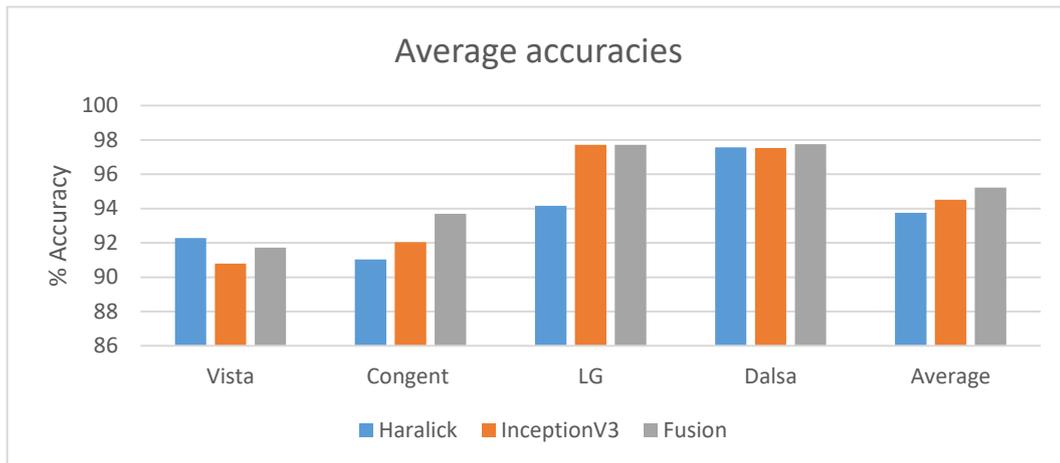
**Fig 6.** Accuracies achieved by different classifiers and ensembles of best-performing classifiers for the datasets of Clarkson 2015 and IIITD, with their two subsets, LG, Dalsa, and Vista, Congent, respectively on Haralick features.



**Fig 7.** Accuracies achieved by different classifiers and ensembles of best-performing classifiers for the datasets of Clarkson 2015 and IIITD, with their two subsets, LG, Dalsa, and Vista, Congent, respectively on the harmonization of Haralick features and InceptionV3 features.

Fig 8. shows the average accuracies across the classifiers for each subset of the dataset. The results showed that the average accuracy for the datasets is greater for the harmonization of Haralick and Inceptionv3

features in most of the cases. Congent, LG, and Dalsa subsets have shown an increase in accuracy except for the Vista subset.



**Fig 8.** Average accuracy obtained by InceptionV3 and Haralick features and their harmonized features across the datasets of IIITD and Clarkson 2015, with their two subsets, LG, Dalsa, and Vista, Congent, respectively.

Based on the lowest ACER value, table 2 displays the classifier that performs the best on each dataset subset. For every subset, the classifier, the APCER, the NPCER, and the ACER are displayed. With an ACER of 3.3%, the table shows that the IBK+Logistic+SMO combination is the most effective classifier for the Congent subset. With an ACER of 1.1%, the Simple Logistic Regression is the most effective classifier for the Vista subset. With an ACER of 0.8%, the IBK is the most effective classifier for the LG subset. With an ACER of 0.2%, the SMO is the most effective

classifier for the Dalsa subset. Additionally, the data demonstrates that among the four subsets, the Dalsa subset has the lowest error rates, making it the simplest to classify.

The proposed method achieved the lowest ACER of 0.2% and the highest accuracy gain of 99.76%, indicating greater accuracy and reliability in comparison to alternative approaches. These values, the lowest among those in Table 1, underscore the method's exceptional performance in minimizing errors and enhancing the predictive capability of the ILD.

**Table 2.** Best classifier performance on a subset of a dataset of Clarkson 2015 and IIIT-D based on ACER values.

Dataset	Classifier/Ensemble	APCER (%)	NPCER (%)	ACER (%)
Congent	IBK+Logistic+SMO	0.9	5.7	3.3
Vista	SimpleLogistic	0.7	1.5	1.1
LG	IBK	0.8	0.8	0.8
Dalsa	SMO	0.1	0.4	0.2

## 6. Conclusion

The paper addresses the critical need for enhancing the security of smart homes by implementing robust biometric access control systems, with a particular focus on iris recognition technology. Recognizing the potential vulnerabilities posed by presentation attacks, the integration of ILD techniques becomes imperative to distinguish between authentic and fraudulent attempts at access. Therefore, it becomes essential to recognize various presentation attacks. To

overcome these difficulties, the suggested framework harmonizes global and local texture features to distinguish between variations in legitimate and fraudulent iris. Various classes of assaulted iris images are included in the datasets of IIIT-D and Clarkson 2015. The results of the experimental analysis demonstrate that the suggested technique works better than different ILDs, obtaining an average error rate of 1.1% on the Clarkson 2015 dataset and 0.2% on the IIITD dataset.

Expanding the research scope beyond IIIT-D and Clarkson 2015 is imperative for future work in order

to further guarantee the robustness and generalizability of the suggested approach. Through a variety of datasets from multiple sources, researchers are able to evaluate how well biometric access control systems function in a variety of environmental and demographic contexts. Furthermore, contrasting the outcomes for cross-dataset validation will offer insightful information about how well the system performs in various contexts. Further enriching the dataset and strengthening the system's resistance to new threats can be achieved by investigating the use of Generative Adversarial Networks (GANs) to create synthetic images of presentation attacks. To enhance the model's capacity to discern between genuine and fraudulent efforts by training classifiers on both synthetic and real data, thereby fortifying the security of sensitive areas such as smart homes. To improve the robustness and dependability of biometric access control systems, future efforts should primarily concentrate on developing the methodology through in-depth testing with a variety of datasets and implementing cutting-edge strategies like GAN-based image generation.

## References

- Choudhary, M., Tiwari, V. & Venkanna, U. (2023). Identifying discriminatory feature vectors for fusion-based iris liveness detection. *J Ambient Intell Human Comput* 14, 10605–10616. <https://doi.org/10.1007/s12652-022-03712-4>
- Haralick, R. M., Shanmugam, K., & Dinstein, I. H. (1973). Textural features for image classification. *IEEE Transactions on systems, man, and cybernetics*, (6), 610-621.
- Impedovo, D., Dentamaro, V., Abbattista, G., Gattulli, V., & Pirlo, G. (2021). A comparative study of shallow learning and deep transfer learning techniques for accurate fingerprints vitality detection. *Pattern Recognition Letters*, 151, 11-18.
- Ishengoma, F. R. (2014). Authentication system for smart homes based on ARM7TDMI-S and IRIS-fingerprint recognition technologies. *arXiv preprint arXiv:1410.0534*
- Kaur, B. (2024). Fingerprint and Iris liveness detection using invariant feature-set. *Multimed Tools Appl.* <https://doi.org/10.1007/s11042-023-17854-w>
- Khade, Smita, Shilpa Gite, and Biswajeet Pradhan. (2022). "Iris Liveness Detection Using Multiple Deep Convolution Networks" *Big Data and Cognitive Computing* 6, no. 2: 67. <https://doi.org/10.3390/bdcc6020067>
- Khade S, Gite S, Thepade SD, Pradhan B, Alamri A. (2021a). Detection of Iris Presentation Attacks Using Feature Fusion of Thepade's Sorted Block Truncation Coding with Gray-Level Co-Occurrence Matrix Features. *Sensors*; 21(21):7408. <https://doi.org/10.3390/s21217408>
- Khade, S., Ahirrao, S., Phansalkar, S., Kotecha, K., Gite, S., & Thepade, S. D. (2021b). Iris liveness detection for biometric authentication: A systematic literature review and future directions. *Inventions*, 6(4), 65.
- Kimura, G. Y., Lucio, D. R., Britto Jr, A. S., & Menotti, D. (2020). CNN hyperparameter tuning applied to iris liveness detection. *arXiv preprint arXiv:2003.00833*.
- Kohli, N., Yadav, D., Vatsa, M., & Singh, R. (2013). Revisiting iris recognition with colour cosmetic contact lenses. In *2013 International Conference on Biometrics (ICB)* (pp. 1-7). IEEE. <https://doi.org/10.1109/icb.2013.6613021>
- L. R. Wagh, S. D. Thepade (2024). Iris Liveness Detection using Fusion of Thepade SBTC and Triangle Thresholding Features with Machine Learning Algorithms. *International Research Journal of Multidisciplinary Technovation (IRJMT)* 6(1), 128-139. <https://doi.org/10.54392/irjmt24110>
- Li, D., Wu, C., & Wang, Y. (2021). A novel iris texture extraction scheme for iris presentation attack detection. *Journal of Image and Graphics*, 9(3), 1-12.
- Peng, Z., Huang, W., Gu, S., Xie, L., Wang, Y., Jiao, J., & Ye, Q. (2021). Conformer: Local features coupling global representations for visual recognition. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 367-376).
- Porebski, A., Vandenbroucke, N., & Macaire, L. (2008). Haralick feature extraction from LBP images for color texture classification. In *2008 First Workshops on Image Processing Theory, Tools and Applications* (pp. 1-8). IEEE.

- Shanmugapriya, D., Padmavathi, G., & Aysha, A. (2023). Detection of Iris Template Attacks using Machine Learning and Deep Learning Methods. In 2023 10th International Conference on Computing for Sustainable Global Development (INDIACom) (pp. 425-429). IEEE.
- Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2818-2826).
- Tapia, J. E., Gonzalez, S., & Busch, C. (2021). Iris liveness detection using a cascade of dedicated deep learning networks. *IEEE Transactions on Information Forensics and Security*, 17, 42-52.
- Toennies, K. D. (2024). Image Features: Extraction and Categories. In *An Introduction to Image Classification: From Designed Models to End-to-End Learning* (pp. 19-57). Singapore: Springer Nature Singapore.
- Verma, P., Selwal, A., & Sharma, D. (2023). IVIDNet: Intelligent iris vitality detection via weighted prediction score level fusion. *Multimedia Tools and Applications*, 1-23.
- Winston, J. J., Hemanth, D. J., Angelopoulou, A., & Kapetanios, E. (2022). Hybrid deep convolutional neural models for iris image recognition. *Multimedia Tools and Applications*, 1-23.
- Yadav, D., Kohli, N., Agarwal, A., Vatsa, M., Singh, R., & Noore, A. (2018). Fusion of handcrafted and deep learning features for large-scale multiple iris presentation attack detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (pp. 572-579).
- Yadav, D., Kohli, N., Doyle, J. S., Singh, R., Vatsa, M., & Bowyer, K. W. (2014). Unravelling the effect of textured contact lenses on iris recognition. *IEEE Transactions on Information Forensics and Security*, 9(5), 851-862. <https://doi.org/10.1109/tifs.2014.2313025>
- Yambay, D., Walczak, B., Schuckers, S., & Czajka, A. (2015). Livdet-iris 2015-iris liveness detection competition. In *Int. Conf. on Identity, Security and Behavior Analysis (ISBA), 2017* (pp. 1-6). <https://doi.org/10.1109/isba.2017.7947701>
- Yan, Z., He, L., Zhang, M., Sun, Z., & Tan, T. (2018). Hierarchical multi-class iris classification for liveness detection. In 2018 International Conference on Biometrics (ICB) (pp. 47-53). IEEE.
- Zhou, S., Xu, C., Xu, R., Ding, W., Chen, C., & Xu, X. (2024). Image recognition model of fraudulent websites based on image leader decision and Inception-V3 transfer learning. *China Communications*, 21(1), 215-227.

# Exploring maritime movement information: an explainable AI approach using Hi-DBSCAN and SHAP analysis

Nitin Newaliya<sup>1</sup>, Vikas Siwach<sup>2\*</sup>, Harkesh Sehrawat<sup>3</sup>, Yudhvir Singh<sup>4</sup>

<sup>1</sup>Department of Computer Science & Engineering, UIET, Maharshi Dayanand University, Rohtak, Haryana, India

\* Corresponding author E-mail: vikas.siwach.uiet@mdurohtak.ac.in

(Received 25 June 2024; Final version received 01 September 2024; Accepted 18 October 2024)

## Abstract

Maritime movement information is pivotal for several applications, including monitoring and examining vessel activities, ensuring efficient and secure navigation, logistics optimisation, and enhancing safety and environmental protection. The maritime industry relies on Automatic Identification System (AIS) data, which provides information on movement of vessels at sea. Determining meaningful and useful insights from this data is a challenge. The complexity and volume of the information make it difficult for traditional methods to provide in-depth insights and explanations. This paper presents an innovative Explainable AI (XAI) approach to explore maritime movement information using AIS data by leveraging high dimensional Density-Based Spatial Clustering of Applications with Noise (Hi-DBSCAN) algorithm and SHAP (SHapley Additive exPlanations) values in a novel way. Through experiments using real AIS datasets, the study reveals the efficacy of SHAP in determining the influence of AIS features on cluster formation. Results from two distinct AIS datasets demonstrate the efficacy of this method. This approach effectively unravels the 'black box' nature of clustering, providing maritime stakeholders with a clearer understanding of vessel behaviour patterns. For instance, in one dataset, the course of the vessel was identified as the most significant feature impacting clustering outcomes. Furthermore, the study explores SHAP's potential for anomaly detection by identifying data points with inconsistent feature influences. This study demonstrates that integrating Hi-DBSCAN clustering with SHAP analysis offers a transparent and interpretable method for understanding vessel behaviour patterns from maritime movement information and extraction of meaningful insights. This framework provides maritime stakeholders with insights beyond traditional pattern recognition, with transparency and explainability, allowing for a deeper understanding and more informed and data-driven decisions in maritime operations.

*Keywords:* Data Analytics, DBSCAN, Explainable AI, AIS, SHAP.

## 1. Introduction

Maritime traffic analysis plays a crucial role in various domains. Automatic Identification Systems (AIS) data, initially envisaged to enhance the safety of vessels at sea by avoiding collisions, now plays a pivotal role in maritime operations by providing real-time information about vessel movements. This data is essential for various applications, including collision avoidance, vessel traffic management, search and rescue, and environmental monitoring.

Assessing AIS data and extracting actionable insights poses significant challenges due to its volume, velocity, and variety. Traditional methods are often unable to process the deeper intricacies of maritime data effectively, and understanding such vessel behaviours requires advanced techniques. Machine learning (ML)

techniques have often been used to assess AIS data and determine traffic patterns. However, even in such cases, the interpretation of the ML models is often limited, which makes it difficult to understand why certain predictions have been made. This is particularly in the case of unsupervised machine learning techniques, amongst which Density-Based Spatial Clustering of Applications with Noise (DBSCAN) is a more commonly used technique in the maritime domain (C. Huang etc., 2023).

The intention of Explainable AI (XAI) techniques is to find a solution to this problem by providing an understanding of the internal functioning of models. XAI models are required in marine policy towards bringing in trust, transferability, fairness, improvements in the models and extracting inputs for decision making in the future (Yan etc., 2022).

This paper proposes an XAI approach to understanding vessel behaviour patterns within AIS data by leveraging the high dimensional DBSCAN algorithm and SHAP (SHapley Additive exPlanations) values to enhance the interpretability of clustering results in maritime movement data, facilitating more informed decision-making in the maritime domain. The effectiveness of this approach in extracting meaningful insights from maritime movement information is demonstrated experimentally, which also provides transparency and interpretability. It empowers maritime stakeholders to gain a deeper understanding of vessel traffic patterns and make data-driven decisions.

The motivation of the paper is to examine enhancements in the exploration of maritime movement information. While clustering algorithms, such as DBSCAN, are extensively utilised for pattern determinations of maritime movement data, these are predominantly black-box implementations to the user, which does not have much interpretability of the results. An insight into the rationale of how the features of the data influenced the outputs will go a long way in better understanding the exploration and derivation of more significant applications. Another motivation was also the limited study that has been undertaken on Explainable AI (XAI) in the maritime movement data context, specifically related to AIS data clustering using DBSCAN. Also, XAI is a fairly new research area, and methods like SHAP offer a solution by providing interpretable explanations for model outputs. Significantly, in the assessment of the authors, such a study combining high-dimension DBSCAN clustering and SHAP for exploring maritime movement information has not been attempted before.

It is useful to understand the significance of features in AIS in the maritime traffic domain which are affecting the clustering results for several reasons, including enhancing clustering results and insights by understanding features that influence the results of the clustering significantly, anomaly detection by determining features known to indicate abnormal behaviour, optimisation of resources by including only significant features during further processing, improved decision making by knowing the important features to focus on, amongst others.

The contribution of the paper is that it provides a significant advancement in exploring maritime movement information by moving away from a black-box method to a more transparent and interpretable approach using explainable AI concepts along with higher dimension unsupervised clustering. By

explaining the underlying reasons for cluster formation, the study empowers maritime stakeholders to make more informed, data-driven decisions based on an enhanced and deeper understanding of vessel behaviour patterns.

The paper sections are: Section 2 introduces DBSCAN and XAI in brief, Section 3 reviews existing relevant literature, Section 4 brings out the proposed methodology, and Section 5 explains the details of the experiments. Section 6 examines the results and analysis of the experiments undertaken. The conclusion and future activities that could be progressed are brought out in Section 7.

## **2. Brief introduction to DBSCAN and SHAP**

### **2.1 DBSCAN**

DBSCAN Data Clustering attempts to segregate data into groups wherein some element of commonality exists between each data point of an individual group. There is no requirement to train a model on the data to determine clusters using Clustering methods. DBSCAN requires two parameters as inputs, viz. Eps (a point's neighbourhood indicated by an appropriate distance measure) and MinPts (minimum number of points required within the neighbourhood). DBSCAN has the ability to detect clusters which may be of arbitrary shape and, therefore, is well suited for clustering AIS data.

The popularity of DBSCAN in the maritime domain has resulted in considerable research and adaptations to be evolved on it. There have been various attempts to enhance the clustering results by including parameters other than spatial while clustering. One such attempt to optimise DBSCAN is by using additional parameters from AIS data, viz. speed indicated as Speed Over Ground (SOG), course indicated as Course over Ground (COG) and Heading, apart from the spatial parameters, i.e. Latitude (LAT) and Longitude (LON) (Han etc., 2021). This higher dimensional DBSCAN (Hi-DBSCAN) enables more refined clusters to be determined and enhances the clustering outputs.

### **2.2 Explainable AI and SHAP**

The black-box nature of machine learning models often limits the understanding and interpretation of how the outputs have been reached at, which can lower

the trust and affect debugging or analysis. In the domain of AI, explainability can be considered as undertaking additional actions to understand an ML model which cannot be interpreted by humans (Yan etc., 2022). Explainable AI (XAI) can be defined as the production of details or rationales of the functioning of a model to enhance the understandability of a target audience (Barredo Arrieta etc., 2020).

SHapley Additive exPlanations (SHAP) is one method which enables explaining how the output of a model has been determined. It provides the contribution made by each feature to a particular output. This is done by determining the marginal contribution of a particular feature by the variation in the results of combinations of features by including and neglecting the particular feature. A positive value of SHAP shows that a particular feature increases the predicted value, and on the other hand, a negative value leads to a decrease. A representation is shown in Fig 1. This thus helps in assessing how the various features interact to determine the output of the model, which can be plotted in various ways to enable explainability.

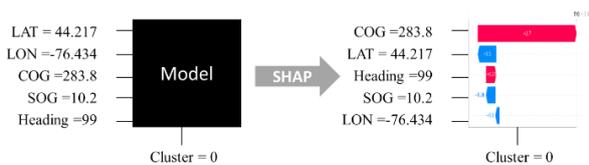


Fig 1. Explaining a model using SHAP.

The SHAP values can be analysed from a Global perspective by averaging the absolute values of each variable to assess the importance of the feature or Locally by the influence of each feature on selected samples of data. Detailed explanations of SHAP have been brought out in (Lundberg etc., 2017, 2020; L. Wang etc., 2022). A simplified version of the calculation of SHAP is given in Eq. (1), in which  $F$  is the features set,  $S$  is the features subset, the function  $v$  generates the prediction value of the model from the features,  $i$  is the feature of interest index,  $|F|$ , represents feature permutation numbers in set  $S$  and  $SHAP_i$  is the feature  $i$  SHAP value (Su etc., 2024).

$$SHAP_i = \sum_{S \subseteq F - i} \left[ \frac{|S|!(|F|-|S|-1)!}{|F|!} - v(S) \right] \quad (1)$$

There are several advantages of using SHAP. It assists in debugging a model in case the predictions are not correct, thus enabling informed modifications. It also enables the generation of explanations which

can be understood for each prediction and the model as a whole. This is particularly important when certain applications require that there be some sort of explanation for the results. It is also a valuable method of data exploration, which can assist in determining various insights that may not be readily evident, as well as interactions between features. This can lead to much better development of models.

### 3.Review of existing literature

There are various models that have been developed for XAI. While comparing the performance of three XAI models in assessing the automatic docking of vessels, viz. SHAP, Local interpretable model-agnostic explanations (LIME) and Linear Model Trees (LMT), it has been brought out that SHAP provides smooth explanations that are intuitive within a reasonably fast time (Lover etc., 2021). In a study undertaken to evaluate XAI methods for assessing results of the classification of ships, it has been brought out that these methods are also applicable for time series information with multiple variables (Veerappa etc., 2022). Furthermore, SHAP-based methods have been found to be more robust than (LIME) and Path integrated gradients (PIG).

SHAP was proposed to help understand why certain predictions have been made by a model since this can be a difficult task for complex models, which generally are more accurate (Lundberg etc., 2017). SHAP achieves this by assigning an importance value to each feature for a specific prediction. It has also been brought out that both local and global explanation have their advantages in providing a richer understanding of the models (Lundberg etc., 2018, 2020).

While proposing a new model for deciding the detention of a ship towards reducing traffic risks, SHAP has been utilised to interpret the model and provide the contribution of the features of the model (He etc., 2021).

SHAP has also been used to determine the contribution that sensors have in an anomaly in the main engine of a vessel (Kim etc., 2021). It has been indicated that the SHAP algorithm is robust and can perform with various types of ML algorithms.

A study has been undertaken to determine the factors responsible for accidents in the maritime domain (C. Zhang etc., 2022). It classified the accidents into six different categories depending on the circumstances of the collisions. These results were then analysed and interpreted using the SHAP model. This

interpretation added further value to the study by analysing factors and elements and correlating characteristics.

A hybrid technique has been proposed to develop a model in order to predict the risk of maritime accidents with enhanced accuracy (Lan *et al.*, 2024). This model is then assessed through SHAP to determine the acts that a seafarer may commit, which could be deemed as unsafe and lead to accidents. This has the potential to predict an accident in advance, enabling preventive measures to be initiated.

Towards safe navigation by vessels, a high degree of vigilance needs to be maintained by the Vessel Traffic Controllers (VTC). Using 45 eye-tracking features, a prediction model has been evolved (Z. Li *et al.*, 2024). A hierarchical analysis has been undertaken, in which SHAP plays an important role by identifying the significant features.

One method of safeguarding maritime transportation activities is by the port control, which undertakes inspection of ships that visit the port. A model has been developed to enable the selection of ships for inspection (Yan *et al.*, 2022), which is then assessed using SHAP. This can help determine information on fairness, validation, and prediction. Reasons for the importance of explainability in the maritime domain are also elucidated, which are brought out later in the paper.

A number of studies have been undertaken with regard to fuel cost estimations and energy efficiencies, which have exploited XAI methods. These include providing the correct shaft power magnitude to the propeller of a ship (Kim *et al.*, 2023), prediction of the consumption of fuel through various models using data from sensors (Ma *et al.*, 2023), determining average draught and relative wind speed as the dominating operational and environmental factors, respectively (Handayani *et al.*, 2023), prediction of consumption of fuel in ships (H. Wang *et al.*, 2023), determining feature importance while addressing energy efficiency concerns in short-sea shipping (Abuella *et al.*, 2023), prediction of fuel cost using wherein static and dynamic parameters related to the vessel (Su *et al.*, 2024) and prediction of the fuel consumption and emissions of a vessel (Lee *et al.*, 2024).

In an effort to assess the temporal features' effect on the prediction of ship routes, two models have been

developed using a K-nearest neighbour classifier, with one of them incorporating the date and time features (Lo Duca & Marchetti, 2022). SHAP assessed the contributing features, bringing out that the temporal features result in lowering the predicted class value, which is why it tends to perform inferior to the other model.

Using features from AIS data, an attempt has been made to determine the status of congestion in ports and also predict the time in ports using XGBoost and SHAP (T. Zhang *et al.*, 2023). SHAP helps in assessing the impact of each feature, enabling modification of the model to optimise the scheduling of ships.

In an effort to understand bike-sharing utilisation for leisure or as part of the public transport system in various roles, DBSCAN was used to separate leisure or transit trips. Thereafter, using gradient boosting and SHAP, the effect of variables in each trip is assessed. (X. Li *et al.*, 2024).

One of the studies on road traffic attempted to explain the prediction in real traffic using Random Forest and Recurrent Neural Networks by SHAP (Barredo-Arrieta *et al.*, 2019). The study brings out that forecasting models should be assessed beyond prediction, accuracy and other scores by assessing deeper information that the contributing variable can provide.

XAI has primarily been applied to supervised learning techniques in machine learning, which include prediction, as seen in the literature above. XAI, however, cannot be applied directly to unsupervised methods. In order to apply XAI to unsupervised techniques, the problem needs to first be converted into a supervised task, enabling the clustering results to be represented in a manner suitable to XAI techniques (Bobek *et al.*, 2022). Summarising the activity into three basic steps, the first step involves undertaking a cluster of the data under consideration. Thereafter, the clusters are used as target variables to build a classifier. Finally, using XAI techniques, this classifier is explained (Horel *et al.*, 2020; Lötsch & Malkusch, 2021; Morichetta *et al.*, 2019). Various improvements are also being proposed, such as those which take into consideration data quality (Alvarez-Garcia *et al.*, 2024).

A summary of the literature addressing maritime related aspects is brought out in Table 1.

**Table 1.** Summary of Maritime related literature.

Reference (First Author, Year)	Application Area	Key Findings & Role of XAI
Lover, 2021	Automatic Docking of Vessels	Compared SHAP with LIME and LMT for explaining automatic docking, finding that SHAP provides smooth, intuitive explanations quickly.
Veerappa, 2022	Ship Classification	Showed the applicability of XAI methods, particularly SHAP, for time series data with multiple variables, finding SHAP to be more robust than LIME and PIG
Lundberg, 2017, 2018, 2020	General Model Interpretability	Emphasised the use of SHAP for understanding predictions made by complex models by assigning importance values to features. Highlighted the value of both local and global explanations.
He, 2021	Ship Detention Decisions	Used SHAP to interpret a model for deciding ship detention to reduce traffic risks, providing insights into feature contributions.
Kim, 2021	Anomaly Detection in Ship Engines	Utilized SHAP to identify sensor contributions to anomalies in ship engines, indicating SHAP's robustness across different machine learning algorithms.
Zhang, 2022	Maritime Accident Risk Prediction	Developed a hybrid model to predict accident risk and used SHAP to identify unsafe seafarer actions potentially leading to accidents.
Lan, 2024	Maritime Accident Analysis	Classified maritime accidents into categories and employed SHAP to analyze and interpret the factors contributing to each accident type.
Li, 2024	Vessel Traffic Controller Vigilance	Used SHAP in a hierarchical analysis to identify significant eye-tracking features for predicting the vigilance of Vessel Traffic Controllers.
Yan, 2022	Port Control Ship Inspection	Developed a model for selecting ships for inspection and used SHAP to determine fairness, validation, and prediction information. Emphasised the importance of explainability in the maritime domain.
Kim, 2023	Ship Propeller Shaft Power Optimisation	Applied XAI methods to provide correct shaft power magnitude, aiming for fuel cost estimation and improved energy efficiency.
Ma, 2023	Ship Fuel Consumption Prediction	Used sensor data and XAI for fuel consumption prediction, highlighting XAI's role in understanding and improving model accuracy.
Handayani, 2023	Fuel Consumption and Vessel Factors	Identified average draught and relative wind speed as dominant operational and environmental factors affecting fuel consumption using XAI for analysis.
Wang, 2023	Ship Fuel Consumption Prediction	Applied XAI techniques to predict ship fuel consumption, focusing on understanding the factors influencing consumption patterns.
Abuella, 2023	Energy Efficiency in Short-Sea Shipping	Used XAI, specifically SHAP, to determine feature importance in addressing energy efficiency concerns, aiding in targeted optimization efforts.
Su, 2024	Ship Fuel Cost Prediction	Predicted fuel costs using static and dynamic vessel parameters and employed XAI to understand feature influences and improve cost estimation accuracy.
Lee, 2024	Vessel Fuel Consumption and Emissions	Predicted fuel consumption and emissions, utilizing XAI to analyze the factors driving both and potentially guide emission reduction strategies.
Lo Duca, 2022	Ship Route Prediction	Analyzed two K-nearest neighbour models for ship route prediction, with SHAP revealing that incorporating temporal features negatively impacted prediction accuracy.
Zhang, 2023	Port Congestion Status and Prediction	Used XGBoost and SHAP to predict port congestion and ship time in ports, with SHAP helping assess feature impact and optimize ship scheduling.

While this method has been applied to explain clustering results in various fields, the survey of the literature reveals that this has not been used in the maritime domain to assess results of clustering of maritime movement data, particularly those utilising DBSCAN and its variants.

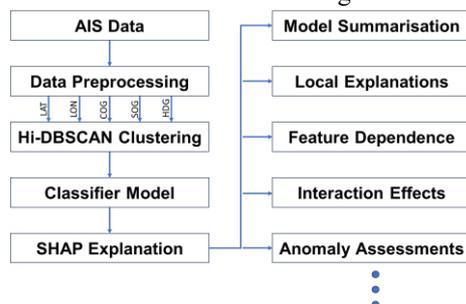
#### 4. Proposed Methodology

The proposed methodology uses an Explainable AI (XAI) approach for determining insights into the vessel behaviour patterns from AIS data. This uses a combination of a high-dimensional Density-Based Spatial Clustering of Applications with Noise (Hi-DBSCAN) algorithm and SHapley Additive exPlanations (SHAP) values. The proposed model is

comprised of three phases. Phase 1 involves undertaking data preprocessing and clustering using Hi-DBSCAN. Phase 2 undertakes conversion to a supervised model. Finally, Phase 3 brings in the explainability of the clustering using SHAP.

In the proposed model, Hi-DBSCAN is used to cluster real AIS data, and thereafter, XAI using SHAP is used to add explainability to the cluster determination. Hi-DBSCAN includes additional non-spatial features while clustering, as brought out in (Han etc., 2020, 2021). The Hi-DBSCAN algorithm clusters AIS data, incorporating non-spatial features (speed, course, heading) along with spatial parameters (latitude and longitude) for refined cluster identification. While this improves the cluster determinations, it renders it more difficult to understand since multiple features have now been combined to produce the clusters. However, SHAP is to be applied to a supervised model, while clustering is to be applied to an unsupervised model. Therefore, as brought out in the literature review section, the model needs to be converted from a clustering to a supervised model. There are various machine learning algorithms in supervised learning techniques. In a study for the classification of ships using AIS data, it emerged that Random Forest (RF) and eXtreme Gradient Boosting (XGBoost) performed better than six other algorithms (I. L. Huang etc., 2023). For this study, Random Forest has been used to convert the clustering results into a classification model. The resulting clusters are used as target variables to train a Random Forest classifier. The classification model is, thereafter, analysed using the SHAP algorithm, and the SHAP values are assessed through various methods to reveal explanations and insights.

A representation of the workflow for the exploration of the maritime movement information using Hi-DBSCAN and SHAP is shown in Fig 2.



**Fig 2.** Model for AIS data exploration using HiDBSCAN and SHAP.

Briefly, the AIS data is pre-processed to ensure that the data is without errors and is limited to what is required for the experiments. Thereafter, the selected

spatial and non-spatial parameters are passed on to the Hi-DBSCAN clustering algorithm.

Hi-DBSCAN introduces non-spatial AIS features along with the latitude and longitude values while calculating the distance using the Mahalanobis distance measure between data points for clustering. These additional features selected are the movement-related parameters from the AIS data, such as SOG, COG and Heading. The multidimensional data point 'A' consists of the attributes as seen in Eq. (2).

$$A = [LAT, LON, SOG, COG, Heading] \quad (2)$$

A suitable Classifier model is then trained on the Clustering results. This trained model is then processed through the SHAP explainable AI model to determine the SHAP values. Using these values, various explanations of the data and exploration of the results are undertaken to determine insights.

## 5.Experiments

In order to test the proposed methodology of exploration of data using XAI, real AIS has been sourced to enable a realistic assessment of the experiments. AIS system onboard a ship transmits messages at regular intervals. AIS data provides details on maritime movement and other information about vessels, represented by 27 different messages (UNSD\_MM, 2020). These messages contain various features, which can be categorised into dynamic, viz. location, speed, course, heading, etc.; static, viz. name, flag, destination, length, etc.; and voyage related, viz. next port of call, expected time of arrival, etc. Two diverse datasets were selected for the exploration. The first dataset comprises segments of a ferry between Wolfe and Kingston region. The second data set comprises selected AIS data from the US east coast. This data was pre-processed to remove inconsistencies and relevant features as well as spatial and temporal extent as required for experiments. This involved data cleaning to remove errors or inconsistencies in the raw AIS data, such as missing values, invalid data points, or duplicate entries. Feature selection was undertaken to select only those features directly related to vessel movement information and exclude others like vessel type or size since these parameters remain static during the passage of a vessel. For the experiments, while the AIS data from Marine-Cadastre comprises 18 features, only spatial and non-spatial features were selected for clustering, viz. Speed Over Ground (SOG), Course Over Ground (COG),

Heading, Latitude (LAT), and Longitude (LON). Data closer to the harbours has been excluded due to the inconsistent, complex and dense traffic patterns in these areas which do not reflect general vessel behaviour. The parameters for Hi-DBSCAN, i.e. minimum points and reachability distance, have been selected iteratively. Selecting appropriate values for MinPts and Eps is critical for effective clustering since different values can significantly influence the clusters that are ascertained. The optimal values depend on the specific dataset and the characteristics of the clustering being sought. Iterative assessment undertaken involved starting with an initial parameter selection, which in these experiments were MinPts equal to 5 and Eps equal to 1. The Hi-DBSCAN algorithm was run with the selected parameters and the quality of the resulting clusters evaluated visually. Based on the evaluation, the MinPts and Eps values were adjusted. This has been repeated, experimenting with different combinations of MinPts and Eps, until a satisfactory clustering result has been achieved for each data set. Python has been the development environment, while QGIS (Quantum Geographic Information System) has also been used towards the visualisation of the data.

The data used for the study was from 01 January 2021 between 1000hrs and 2300hrs and sourced from MarineCadastre (AccessAIS - MarineCadastre.Gov, 2023). After the data has been prepared, Hi-DBSCAN clustering is undertaken on it. An iterative assessment for the determination of parameters for clustering results in minimum points as three and maximum distance as 1.75. The result of the clustering is shown in Fig 3. The clustering has resulted in two primary clusters, which represent directional traffic movement. In Fig 3, each dot represents a single data point, which corresponds to a vessel's position at a specific time. The clustering algorithm analyses these positions, along with other features like speed and course, to group together vessels exhibiting similar movement patterns. The clustering results in the determination of two primary clusters, indicating Directional Traffic Flow. One cluster represents a ferry travelling from one location (Wolfe) to another (Kingston). The other cluster represents the same ferry travelling in the opposite direction (Kingston to Wolfe).

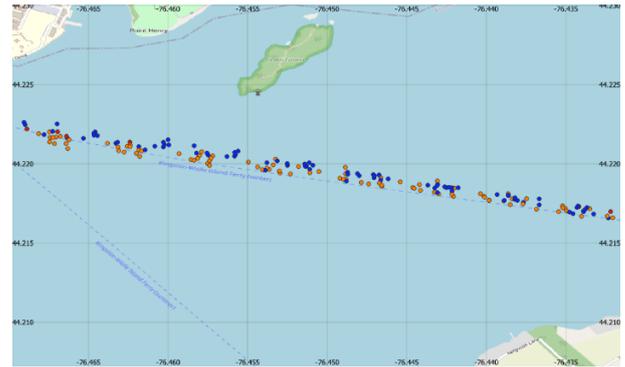


Fig 3. Hi-DBSCAN clustering.

The models trained by the Classifier are then processed through the SHAP algorithm. This results in the determination of SHAP values for the dataset, which can then be assessed by various methods to lend explainability to the results. The results and analysis of the data using SHAP are presented in the next section.

## 6.Results and Analysis

### 6.1.SHAP based assessment of Hi-DBSCAN Clustering on Dataset 1

#### i. SHAP Global Explanation - Bar Plot

The SHAP Bar plot in Fig 4 shows the absolute average SHAP values of every feature of the dataset using Random Forest. It enables determining the contribution of each feature in the clustering to the others since a large value can influence the change in output more. It can be seen from the plot that in the given dataset, COG has the maximum SHAP value of 2.45 and is the most significant feature, having the most impact on the clustering. The other features do not have a significant impact on the clustering results, with the next significant feature being Heading. This plot helps identify the most influential features of AIS data which are crucial for separating groups of vessels, detecting anomalies, understanding traffic patterns and risk zones, etc.



Fig 4. Bar plot – Random Forest.

## ii. SHAP Global Explanation - Beeswarm Plot

The Beeswarm plot is shown in Fig 5. Beeswarm plots enable a comprehensive understanding of the way in which the SHAP values are distributed using both positive and negative values. The SHAP values are represented along the x-axis, and the distribution along this axis indicates the influence and impact that a feature has on the output. Since it plots both positive and negative values, it enables the determination of whether the feature increases or decreases the value of the prediction. Each dot indicates the SHAP value of one of the samples for a feature from the AIS data. The colour of the dot is a gradient from blue, representing a low feature value, to red, indicating a high feature value. Multiple dots at the same position are stacked up to indicate the density. Here again, it is seen that COG has a major influence on the output of the model, and both high and low values play an important role in the Clusters that are formed. Since both extremes are associated with SHAP values that are positive, it implies a non-linear relationship between the model's output and COG. The other features are generally centred around zero, which implies a minimal impact on the clustering. However, some low values of Heading are seen on the positive side, implying instances of having an increased influence in certain local cases. For an analyst, it enables the determination of whether the feature increases or decreases the value of the

prediction. Identification of outliers, visualisation of feature interactions, potentially risky manoeuvres and undertaking comparisons across predictions are some of the other ways in which a maritime traffic analyst can exploit the Beeswarm plots.

The global Bar and Beeswarm plots enable a useful high-level summary of influences as assessed by the model. The information obtained from the Bar plot and Beeswarm plot also enables an assessment of the features to be selected in case further analysis is required. As in this case, for DBSCAN clustering, only COG, along with the mandatory spatial components, need to be considered if further clustering assessments are required. This will speed up the processing with a lesser amount of data.

A tabular representation of the minimum, maximum and average SHAP values and the relative feature ranking in each case is shown in Table 2.

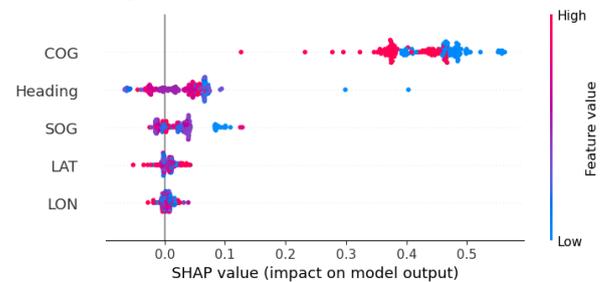


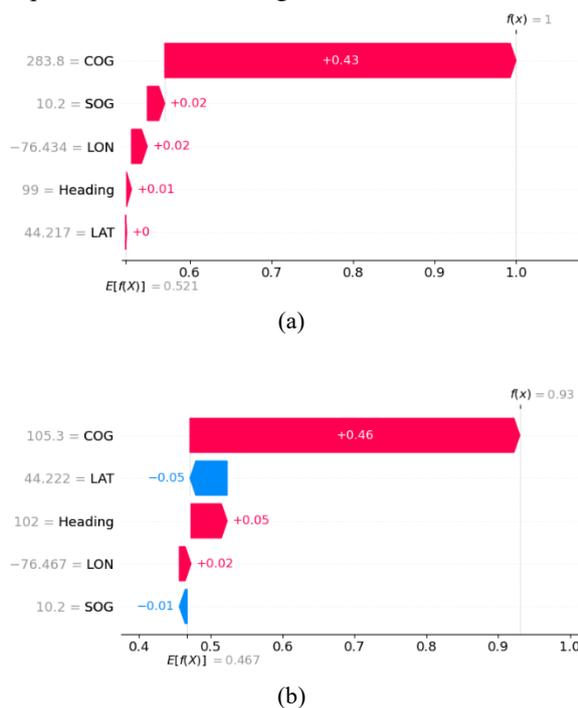
Fig 5. Beeswarm plot – Random Forest.

Table 2. SHAP Values

Feature	Min SHAP Value	Max SHAP Value	Absolute Mean SHAP Rank	Min SHAP Rank	Max SHAP Rank	Absolute Mean SHAP Rank
COG	0.126498	0.562583	0.432059	5	1	1
Heading	-0.065910	0.403263	0.047758	1	2	2
LAT	-0.051871	0.042487	0.009826	2	4	4
LON	-0.027482	0.039224	0.007080	3	5	5
SOG	-0.024708	0.128751	0.030036	4	3	3

**iii. SHAP Local Explanation – Waterfall Plot**

While the above two plots visualised the SHAP values for the entire data, it is important to explore how the various features influence the clustering of a single data point. This is also possible in SHAP with the use of Waterfall Plots. This can be seen in Fig 6 for Random Forest, which shows the plots for a single instance of two different clusters. Fig 6(a) is of an instance in one cluster in which the predominant feature is a high value of COG, but SOG and LON also contribute some influence. Fig 6(b) shows another instance in another cluster in which the dominating feature is again a high value of COG. In this case, LAT and Heading contribute in opposite ways. Therefore, an explanation of why a particular point has been put in a particular cluster is evident. These plots provide a clear, intuitive way to indicate the aspects determining a vessel's cluster assignment to stakeholders, such as port authorities or shipping companies and help the maritime analyst to identify patterns, misclassifications, relationships between features, etc. which can improve the understanding of the maritime traffic.

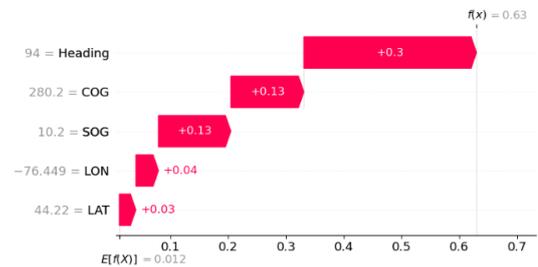


**Fig 6.** Random Forest Waterfall Plots (a) Cluster 0 and (b) Cluster 1

**iv. Anomaly Assessment**

The Waterfall plot can also be used to assess the reasons for anomaly or outlier detection. From the data, one of the outlier points was analysed and is shown in Fig 7. It has been observed that the Heading feature has an influence not consistent with other data points

and, hence, has been identified as an outlier. This can be thereafter further analysed to detect further rationale for its occurrence.

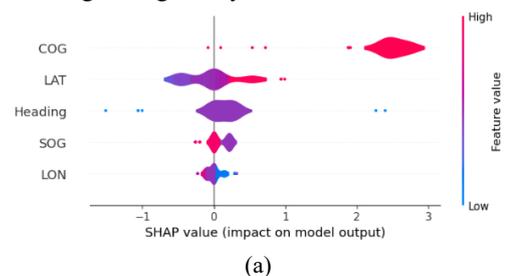


**Fig 7.** Anomaly assessment

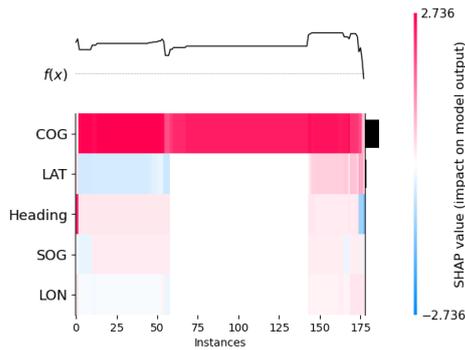
**v. Additional Plots**

There are a number of additional plots that can be generated using SHAP values, providing insights and exploration capabilities, and can be exploited for further analysis of SHAP values and the model depending on the use case. Some of these plots are shown in Fig 8. The Force plot is particularly useful to a maritime traffic analyst for comparing the relative impact of different features on a vessel's cluster assignment and for real-time traffic monitoring, where it is important to understand sudden changes in patterns. Examination of the Heatmap plot can help to identify features that have strong interactions, and enable understanding the underlying relationships in the data. The violin plot enables the determination of consistent trends or variations, e.g. if the speed is showing a narrow impact on clustering. It will help the analyst assess that there are expectable speed patterns among vessels. A dependence plot demonstrates the impact of a feature changes with variation of its value and thus helps reveal the relationship to the analyst.

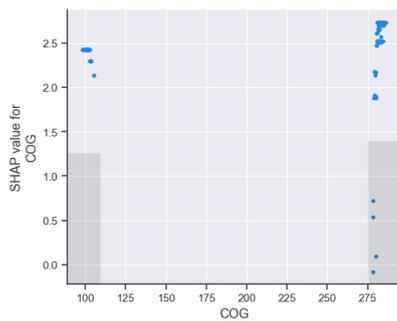
A comparative assessment of global mean SHAP values with region-specific SHAP values can also be undertaken, in which areas of divergence could indicate influential features which may be location-dependent and help in explaining outliers. Thus, targeted SHAP assessments can help discover contextual influences along with globally stable features.



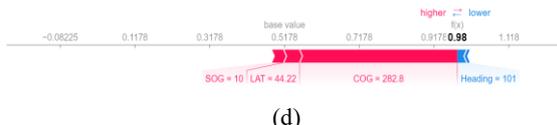
(a)



(b)



(c)

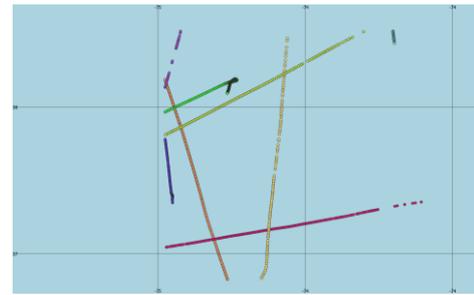


(d)

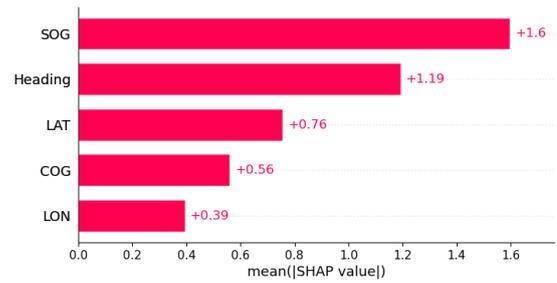
**Fig 8.** Additional SHAP Plots (a) Violin (b) Heatmap (c) Dependence and (d) Force

## 6.2. SHAP based assessment of Hi-DBSCAN Clustering on Dataset 2

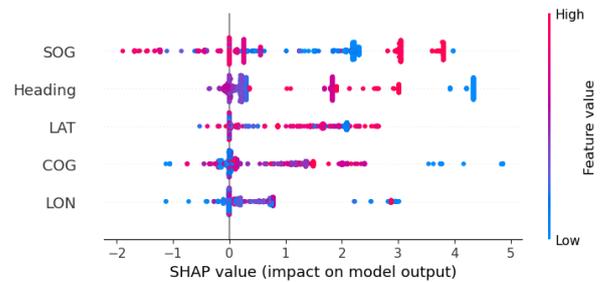
Explanation by SHAP using XGBoost has also been assessed on another data set, in which the selected AIS data has been sourced from the east coast of the USA on 31 December 2020. Hi-DBSCAN clustering result is shown in Fig 9. It comprises of paths of eight vessels. Each of these vessels has a distinct path, with variation in its movement parameters as well as spatial parameters, leading to the formation of eight separate clusters. The Bar plots indicate a higher contribution of SOG and Heading than the others, which also have a certain amount of contribution, unlike Data Set 1. SHAP plots are shown in Fig 10. An analysis of the Beeswarm plot indicates that both high and low values influence the clustering for most variables, with SOG having a higher influence and variation around 0. Two examples of waterfall plots of clusters are shown, which show how SOG and COG can have differing influences in the two clusters.



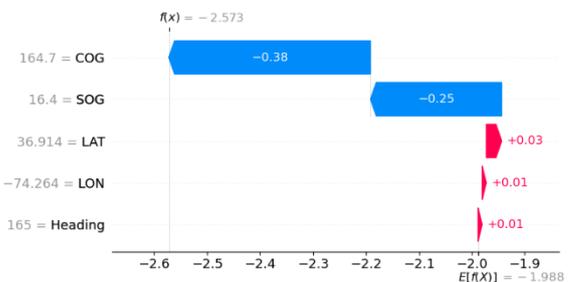
**Fig 9.** Dataset 2 Hi-DBSCAN Clustering.



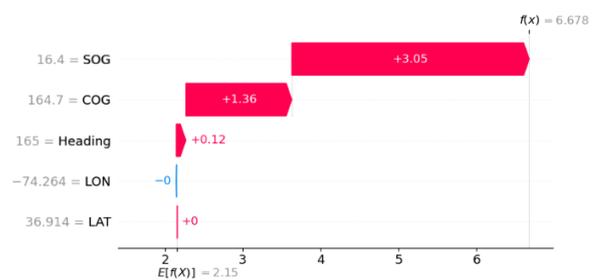
(a)



(b)



(c)



(d)

**Fig 10.** Dataset 2 SHAP plots (a) Bar (b) Beeswarm (c) Waterfall Cluster 0 and (d) Waterfall Cluster 1.

## 7. Discussions

The experiments and results demonstrate that SHAP has significant application in assessing the influence of individual features in the output of a higher dimension DBSCAN clustering of AIS data. Using the global assessments through Beeswarm and Bar plots, the average influence exerted by all the variables on the model can be determined to understand the overall impact of clustering. The local assessments using Waterfall plots, on the other hand, enable determining the contribution each variable has to the final cluster that has been determined. These two assessments enable the explainability of the clustering and provide an insight into the internal functioning of the clustering model. Thus, using SHAP analysis as a second stage after implementing the clustering machine learning model in the first stage enables explanations of the clustering results. This study has explored the clustering results of AIS data and enables an explanation of the factors influencing the overall clustering as well as individual instances. This has wide-ranging applications in the maritime domain while exploring maritime movement information such as AIS and providing valuable insights from explanations using SHAP, including, assessment of anomalies and outliers. The transparency and interpretability obtained through XAI techniques like SHAP enables maritime stakeholders to understand the factors affecting the clustering results and thereafter make more learned decisions. Therefore, there is a significant impact that this framework will have towards enhancing the understanding and management of maritime activities, and facilitate contributing to a safer, more efficient, and sustainable maritime domain, including the following:

- (a) **Enhancing Maritime Situational Awareness:** It has been seen that the factors driving vessel behaviour patterns can be understood through SHAP analyses of Hi-DBSCAN clustering results. The traffic flows can be anticipated and managed better, thereby enhancing safety and reducing congestion.
- (b) **Improved Anomaly Detection:** It emerges that the proposed approach can help identify outliers or unusual vessel behaviours. These may indicate illegal activities, safety violations, or potential hazards, which would assist the maritime analysts in their processes.
- (c) **Formulating Policy Decisions:** Since the framework enables understanding of the influencing factors behind vessel movements, these can enable the

formulation of more effective and targeted policies related to areas like maritime safety, environmental protection, and port management.

- (d) **Optimised Logistics and Efficiency:** The insights that are determined using this approach into vessel traffic patterns can help in optimising shipping routes, improving logistics planning, and reducing fuel consumption and emissions.

While XAI has been used to interpret supervised learning techniques in maritime applications, its application to unsupervised methods like clustering remains under-explored, particularly in the context of AIS data and DBSCAN. The model proposed in this paper addresses a gap in existing research by applying SHAP to explain Hi-DBSCAN clustering outcomes. As a future scope of work, comparative studies with other classification algorithms and the XAI model are recommended, which then facilitate additional assessment studies.

The results of the study directly contribute to enhancing the exploration of maritime movement information by providing a framework for ascertaining not just what clusters are formed, but also why those clusters formed. This in-depth understanding because of the interpretability brought in by SHAP, enables maritime stakeholders to progress ahead from just pattern recognition towards actionable insights and informed decision-making.

## 8. Conclusion and Future Scope

SHAP, an XAI method, has been exploited to bring in explainability to higher dimension DBSCAN clustering of maritime movement information, viz. AIS data. It has been shown how the various features of AIS data influence the formation of clusters, thereby making it more understandable to both model developers and policymakers. For a given scenario, the combination of AIS data, Clustering and SHAP is effectively utilised to glean valuable information for further understanding and assessments.

Significantly, the study moves beyond just unsupervised clustering results of AIS data, diving instead into the rationale for the formulation of the clusters. One of the crucial highlights of the research includes examining the interplay of both global and local SHAP assessments and their ability to uncover insights. This research introduces an innovative framework for maritime movement data combining higher dimension unsupervised clustering with explainability through SHAP, enabling an insightful understanding of the

dynamics at play within the maritime movement information.

Machine Learning and AI have considerable potential in the maritime field, and this study has demonstrated the way in which XAI models such as SHAP can be effectively utilised for the explainability of clustering models.

Future opportunities in the utility of SHAP plots in the analysis and application of clustering through DBSCAN are considerable. It has been shown how it can be used for anomaly/ outlier detections, which can be dwelled into in more detail, including a deeper analysis of regions towards the determination of anomaly detection systems. SHAP assessments can also be utilised to determine the weighting of certain features depending on the outcomes required. It can also be used in studies to determine significant parameters and thereafter optimise further processing. Also, additional features of AIS data can be considered in the higher dimensional DBSCAN, and clustering can be assessed using SHAP. Some other XAI models and classification techniques can also be studied, and a comparative assessment can be undertaken.

In conclusion, this study demonstrates a powerful framework for explaining the clustering results in the complex maritime movement information landscape, enabling more transparency to all concerned towards informed decisions by examination of global, regional and outlier behaviours. This can help in more informed exploration and understanding of maritime movement data and facilitate advances in secure and efficient maritime traffic movements.

## Acknowledgement

AIS data from the Bureau of Ocean Energy Management (BOEM) and the National Oceanic and Atmospheric Administration (NOAA). MarineCadastre.gov. Vessel Traffic Data. Retrieved November 26, 2023, from [marine-cadastre.gov/data](https://marine-cadastre.gov/data).

## References

Abuella, M., Atoui, M. A., Nowaczyk, S., Johansson, S., & Faghani, E. (2023). Data-Driven Explainable Artificial Intelligence for Energy Efficiency in Short-Sea Shipping. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*: Vol. 14175 LNAI. [https://doi.org/10.1007/978-3-031-43430-3\\_14](https://doi.org/10.1007/978-3-031-43430-3_14)

AccessAIS - MarineCadastre.gov. (n.d.). Retrieved November 26, 2023, from <https://marinecadastre.gov/accessais/>

Alvarez-Garcia, M., Ibar-Alonso, R., & Arenas-Parra, M. (2024). A comprehensive framework for explainable cluster analysis. *Information Sciences*, 663, 120282. <https://doi.org/10.1016/J.INS.2024.120282>

Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, 58, 82–115. <https://doi.org/10.1016/J.INFFUS.2019.12.012>

Barredo-Arrieta, A., Lana, I., & Del Ser, J. (2019). What Lies Beneath: A Note on the Explainability of Black-box Machine Learning Models for Road Traffic Forecasting. 2019 IEEE Intelligent Transportation Systems Conference, ITSC 2019, 2232–2237. <https://doi.org/10.1109/ITSC.2019.8916985>

Bobek, S., Kuk, M., Szelazek, M., & Nalepa, G. J. (2022). Enhancing Cluster Analysis With Explainable AI and Multidimensional Cluster Prototypes. *IEEE Access*, 10, 101556–101574. <https://doi.org/10.1109/ACCESS.2022.3208957>

Han, X., Armenakis, C., & Jadidi, M. (2020). DBscan optimization for improving marine trajectory clustering and anomaly detection. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, 43(B4), 455–461. <https://doi.org/10.5194/isprs-archives-XLIII-B4-2020-455-2020>

Han, X., Armenakis, C., & Jadidi, M. (2021). Modeling vessel behaviours by clustering ais data using optimized dbscan. *Sustainability (Switzerland)*, 13(15). <https://doi.org/10.3390/su13158162>

Handayani, M. P., Kim, H., Lee, S., & Lee, J. (2023). Navigating Energy Efficiency: A Multifaceted Interpretability of Fuel Oil Consumption Prediction in Cargo Container Vessel Considering

- the Operational and Environmental Factors. *Journal of Marine Science and Engineering*, 11(11). <https://doi.org/10.3390/jmse11112165>
- He, J., Hao, Y., & Wang, X. (2021). An interpretable aid decision-making model for flag state control ship detention based on SMOTE and XGboost. *Journal of Marine Science and Engineering*, 9(2), 1–19. <https://doi.org/10.3390/jmse9020156>
- Horel, E., Giesecke, K., Storchan, V., & Chittar, N. (2020). Explainable clustering and application to wealth management compliance. *ICAIF 2020 - 1st ACM International Conference on AI in Finance*. <https://doi.org/10.1145/3383455.3422530>
- Huang, C., Qi, X., Zheng, J., Zhu, R., & Shen, J. (2023). A maritime traffic route extraction method based on density-based spatial clustering of applications with noise for multi-dimensional data. *Ocean Engineering*, 268, 113036. <https://doi.org/10.1016/J.OCEANENG.2022.113036>
- Huang, I. L., Lee, M. C., Nieh, C. Y., & Huang, J. C. (2023). Ship Classification Based on AIS Data and Machine Learning Methods. *Electronics* 2024, Vol. 13, Page 98, 13(1), 98. <https://doi.org/10.3390/ELECTRONICS13010098>
- Kim, D., Antariksa, G., Handayani, M. P., Lee, S., & Lee, J. (2021). Explainable anomaly detection framework for maritime main engine sensor data. *Sensors*, 21(15). <https://doi.org/10.3390/s21155200>
- Kim, D., Handayani, M. P., Lee, S., & Lee, J. (2023). Feature Attribution Analysis to Quantify the Impact of Oceanographic and Maneuverability Factors on Vessel Shaft Power Using Explainable Tree-Based Model. *Sensors*, 23(3). <https://doi.org/10.3390/s23031072>
- Lan, H., Wang, S., & Zhang, W. (2024). Predicting types of human-related maritime accidents with explanations using selective ensemble learning and SHAP method. *Heliyon*, 10(9). <https://doi.org/10.1016/j.heliyon.2024.e30046>
- Lee, J., Eom, J., Park, J., Jo, J., & Kim, S. (2024). The Development of a Machine Learning-Based Carbon Emission Prediction Method for a Multi-Fuel-Propelled Smart Ship by Using Onboard Measurement Data. *Sustainability (Switzerland)*, 16(6). <https://doi.org/10.3390/su16062381>
- Li, X., Ha, J., & Lee, S. (2024). Unveiling the roles of public bike systems: From leisure to multimodal transportation. *Travel Behaviour and Society*, 34. <https://doi.org/10.1016/j.tbs.2023.100705>
- Li, Z., Li, R., Yuan, L., Cui, J., & Li, F. (2024). A benchmarking framework for eye-tracking-based vigilance prediction of vessel traffic controllers. *Engineering Applications of Artificial Intelligence*, 129. <https://doi.org/10.1016/j.engappai.2023.107660>
- Lo Duca, A., & Marchetti, A. (2022). Towards the Evaluation of Date Time Features in a Ship Route Prediction Model. *Journal of Marine Science and Engineering*, 10(8). <https://doi.org/10.3390/jmse10081130>
- Lötsch, J., & Malkusch, S. (2021). Interpretation of cluster structures in pain-related phenotype data using explainable artificial intelligence (XAI). *European Journal of Pain (United Kingdom)*, 25(2), 442–465. <https://doi.org/10.1002/ejp.1683>
- Lover, J., Gjaerum, V. B., & Lekkas, A. M. (2021). Explainable AI methods on a deep reinforcement learning agent for automatic docking. *IFAC-PapersOnLine*, 54(16), 146–152. <https://doi.org/10.1016/J.IFACOL.2021.10.086>
- Lundberg, S. M., Allen, P. G., & Lee, S.-I. (2017). A Unified Approach to Interpreting Model Predictions. *Advances in Neural Information Processing Systems*, 30. <https://github.com/slundberg/shap>
- Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., & Lee, S. I. (2020). From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence* 2:1, 2(1), 56–67. <https://doi.org/10.1038/s42256-019-0138-9>
- Lundberg, S. M., Nair, B., Vavilala, M. S., Horibe, M., Eisses, M. J., Adams, T., Liston, D. E., Low, D. K. W., Newman, S. F., Kim, J., & Lee, S. I. (2018).

- Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. *Nature Biomedical Engineering* 2018 2:10, 2(10), 749–760. <https://doi.org/10.1038/s41551-018-0304-0>
- Ma, Y., Zhao, Y., Yu, J., Zhou, J., & Kuang, H. (2023). An Interpretable Gray Box Model for Ship Fuel Consumption Prediction Based on the SHAP Framework. *Journal of Marine Science and Engineering*,11(5). <https://doi.org/10.3390/jmse11051059>
- Morichetta, A., Casas, P., & Mellia, M. (2019). Explain-IT: Towards explainable AI for unsupervised network traffic analysis. *Big-DAMA 2019 - Proceedings of the 3rd ACM CoNEXT Workshop on Big Data, Machine Learning and Artificial Intelligence for Data Communication Networks, Part of CoNEXT 2019*, 22–28. <https://doi.org/10.1145/3359992.3366639>
- Su, M., Lee, H. J., Wang, X., & Bae, S.-H. (2024). Fuel consumption cost prediction model for ro-ro carriers: a machine learning-based application. *Maritime Policy and Management*. <https://doi.org/10.1080/03088839.2024.2303120>
- UNSD\_MM. (2020). Overview of AIS dataset - AIS Handbook - UN Statistics Wiki. <https://unstats.un.org/wiki/display/AIS/Overview+of+AIS+dataset>
- Veerappa, M., Anneken, M., Burkart, N., & Huber, M. F. (2022). Validation of XAI explanations for multivariate time series classification in the maritime domain. *Journal of Computational Science*, 58. <https://doi.org/10.1016/j.jocs.2021.101539>
- Wang, H., Yan, R., Wang, S., & Zhen, L. (2023). Innovative approaches to addressing the tradeoff between interpretability and accuracy in ship fuel consumption prediction. *Transportation Research Part C: Emerging Technologies*, 157. <https://doi.org/10.1016/j.trc.2023.104361>
- Wang, L., Gopal, R., Shankar, R., & Pancras, J. (2022). Forecasting venue popularity on location-based services using interpretable machine learning. <https://doi.org/10.1111/Poms.13727>, 31(7), 2773–2788. <https://doi.org/10.1111/POMS.13727>
- Yan, R., Wu, S., Jin, Y., Cao, J., & Wang, S. (2022). Efficient and explainable ship selection planning in port state control. *Transportation Research Part C: Emerging Technologies*, 145. <https://doi.org/10.1016/j.trc.2022.103924>
- Zhang, C., Zou, X., & Lin, C. (2022). Fusing XGBoost and SHAP Models for Maritime Accident Prediction and Causality Interpretability Analysis. *Journal of Marine Science and Engineering*, 10(8). <https://doi.org/10.3390/jmse10081154>
- Zhang, T., Yin, J., Wang, X., & Min, J. (2023). Prediction of container port congestion status and its impact on ship's time in port based on AIS data. *Maritime Policy and Management*. <https://doi.org/10.1080/03088839.2023.2165185>

## AUTHOR BIOGRAPHIES



**Nitin Newaliya** received his M.Tech. degree in Electronics and Telecommunications Engineering from the University of Pune, India. He is currently pursuing Ph.D. degree at Maharshi Dayanand University, Rohtak, India. His research interest includes data analytics in the maritime domain, machine learning and artificial intelligence. He has worked for more than 20 years in a research role in the data and systems integration and engineering field. He is a Fellow in the IETE. He has published papers on maritime data analytic subjects and presented at various international conferences.



**Dr Vikas Siwach** received his M.S Computer Engineering degree from University of Massachusetts, Lowell, Boston, USA. He completed his PhD in Computer Science and Engineering from MDU, Rohtak, Haryana. He has more than 20 years of experience in teaching in the field of Computer Science and Engineering. He is presently guiding 8 Ph.D students and has guided more than 25 students at the Masters level. He has more than 30 research publications. Presently, he is working with Maharshi Dayanand University, Rohtak, Haryana as an Associate Professor.



**Dr Harkesh Sehrawat** completed his Ph.D. in Computer Science and Engineering from Guru ambheshwar University of Science and Technology, Hisar. His research areas include machine learning, data analytics and software defined networks. He has more than 15 years of experience and has presented papers in veracious national and international conferences and also published in various journals. He is presently in the faculty of University Institute of Engineering and Technology at Maharshi Dayanand University, Rohtak, Haryana as an Associate Professor.



**Professor Yudhvir Singh** received his M.Tech. in Computer Engineering from Thapar Institute of Engineering & Technology, Patiala, Punjab, India in 1997 and his Ph.D. from Guru Jambheshwar University of Science and Technology, Hissar, India, in 2010. He has guided a number of Ph.D. students and also has authored a considerable number of papers in the Computer Engineering domain in different Journals of International/National repute. He has delivered invited talks and chaired sessions in various conferences. He is presently a Professor and Director of the University Institute of Engineering and Technology at Maharshi Dayanand University, Rohtak, India.

# Research on the path of coordinated development of innovation model and supply chain capability building of new energy vehicle enterprises in China

Yang Zhuang<sup>1</sup>, Liu Bin<sup>2</sup>, Kuang Guihua<sup>3</sup>, Zhang Wenxin<sup>4\*</sup>

<sup>1</sup>Guangdong University of Science and Technology

<sup>2</sup>FAW-VOLKSWAGEN(FOSHAN), Parts Procurement Department.

<sup>3</sup>Guangdong University of Finance and Economics

<sup>4\*</sup> Guangdong University of Finance and Economics

\*Corresponding author E-mail: wencyzz99@gmail.com

(Received 17 January 2024; Final version received 24 May 2024; Accepted 31 May 2024)

## Abstract

Facing the increasingly complex competitive environment of China's automobile industry, the innovative development mode and the long-term and stable cooperative relationship between vehicle manufacturers and suppliers are the key factors for its high-quality development. In particular, the new energy automobile enterprises, as the shakers of the traditional automobile industry, the coordinated development of the two has become one of the important topics in the process of their rapid development. Based on the supplier theory and related theoretical literature, the synergistic development model of market, main engine plants and supplier research object is constructed. And with FAW-VOLKSWAGEN (FOSHAN) Branch and its suppliers as the research object to carry out field research activities. We use linear regression method to further clarify the mutual promoting factors of innovation mode and manufacturing organization ability. The results show that enterprises can promote the manufacturing organization ability of the supply chain by predicting the market and innovating the procurement mode; and the improvement of the manufacturing organization ability can promote the innovation of the core technology and promote the development of market diversification, which effectively strengthens the core competitiveness of enterprises. On the basis of the existing research results, this paper makes an in-depth analysis of the development strategy of China's new energy vehicle manufacturers and suppliers and the policy suggestions for the development of the new energy vehicle market, so as to help the coordinated development of the innovation mode and manufacturing organization ability of China's new energy vehicle enterprises from the theoretical level.

*Keywords: coordinated development, innovation mode, new energy vehicles, supply chain capability building*

## References

- AKABANE Jun etc. (2017). From product design to product, process and domain design capabilities of local tier 2 suppliers: lessons form case studies in Japan, Thailand and China. *Int. J. Automotive Technology and Management*, 17(4)385-408.
- Chen, F. & Wei, J.G. (2015). Research on Collaborative Innovation and Evolution of Emerging Industries: New Energy Vehicles as an Example, *Research Management*, 2015, 36(1):26-33. (in Chinese)
- Han, W. etc. (2021). How does business model innovation shape the differences in business ecosystem attributes? A cross-case longitudinal study and theoretical model construction based on two start-ups, *Management World*, 37(1):88-107+7. (in Chinese)
- Hou, G.M. etc. (2021) Research on innovation performance and role path of collaborative innovation model of new energy vehicle enterprises in China, *Technology Economics*, 40(11):13-22. (in Chinese)
- Jiao, W. etc. (2020). Joint procurement strategy of automobile OEMs based on supplier compensation

- incentive, *Journal of Systems Engineering*, 35(6):760-772+805.n (in Chinese)
- Liu, G.W. & Shao, Y.F. (2020). Evolution and Collaborative Measurement of Cooperation Network of Strategic Emerging Industries from the Perspective of Industrial Chain Innovation: A Case Study of New Energy Vehicle Industry, *Science of Science and Management of Science and Technology*, 41(8):43-62. (in Chinese)
- Ma, M.M. etc. (2023). Impact of dual credit policy on new energy vehicles technology innovation with information asymmetry, *Applied Energy*, 332.
- Sun, C. (2006). Analysis of the theoretical model of the outsourcing system of Japanese enterprises: A case study of Sato model and Asanuma model, *Modern Japanese Economy*, (1):5-10. (in Chinese)
- Sun, W. & Ye C.S. (2023). Why does government procurement affect enterprise innovation: On the synergy between the "pull" of demand-side policies and the "push" of supply-side policies, *China Industrial Economics*, (1):1-19. (in Chinese)
- Suo, Q. & Li, C.S. (2023). Research on the dynamic mechanism of the evolution of China's new energy vehicle industry from the perspective of technology-cooperation network, *Chinese Journal of Management*, 20(3):329-338. (in Chinese)
- Wang, S.N. & Zhao, S. (2020) Social Trust, Supplier Relationship and Enterprise Innovation, *Journal of Zhongnan University of Economics and Law*, (6):25-34. (in Chinese)
- Xiong, Y.Q. & Qin, S.F. (2023). What kind of innovation has the new energy vehicle industry policy promoted? *Scientific Research Management*, 44(3):102-111. (in Chinese)
- Xiong, Y.Q. & Wang, X. (2021). Innovation incentive effect and mechanism of heterogeneous demand for new energy vehicles: A perspective on the comparison of "government procurement", "commercial operation" and "private passenger" demand, *Journal of Finance and Economics*, 47(7):48-62. (in Chinese)
- Zhang, C.L. etc. (2016). Market opening, capital subsidy and the evolution of emerging industry market: A case study of new energy vehicle industry, *Shanghai Economic Research Journal*, (5):47-57. (in Chinese)
- Zhang, J. (2020). Research on development mechanism of new energy vehicle enterprises based on open source innovation model, *IOP Conference Series: Earth and Environmental Science*, 619(1):012028.
- Zhang, J.J. etc. (2020). Research on business model innovation of new energy vehicles from the perspective of ecological consumption, *Ecological Economy*, 36(3):72-77. (in Chinese)
- Zhang, X.Y. & Li, W.Y. (2021). Research on the incentive of government subsidy in the innovation ecosystem of Guangzhou new energy automobile industry, *E3S Web of Conferences*, 235:01001
- Zhao, Z.Y. etc. (2021). Reconstruction of spatial organization of China's automobile industry cluster under modular production: A case study of FAW-Volkswagen, *Acta Geographica Sinica*, 76(8):1848-1864. (in Chinese)
- Zhou, Y. & Pan, Y. (2019). Financial subsidy and tax reduction: policy analysis of new energy vehicle industry from the perspective of transaction cost, *Management World*, 35(10):133-149. (in Chinese)

# 中国新能源汽车企业创新模式与制造组织能力协同发展的因素分析

杨壮<sup>1</sup>, 刘彬<sup>2</sup>, 匡桂华<sup>3</sup>, 张文昕<sup>4\*</sup>

<sup>1</sup>广东科技学院

<sup>2</sup>一汽-大众(佛山)零部件采购部

<sup>3</sup>广东财经大学

<sup>4\*</sup>广东财经大学

\*通讯作者 E-mail: wencyzz99@gmail.com

## 摘要

中国汽车产业面对日趋复杂的竞争环境,创新发展模式及整车制造商与供应商长期稳定的合作关系是其高质量发展关键因素。这是新能源汽车企业建立早期优势的关键。本论依据浅沼万里供应商理论及相关理论文献构建了以市场、主机厂和供应商研究对象的协同发展的作用模型。并以一汽大众佛山分公司及其供应商为研究对象展开了实地调研活动。本论在验证阶段运用线性回归法进行分析,进一步明确了创新模式和制造组织能力的相互促进因素。结果表明,企业可以通过适应市场及创新采购模式来促进提升供应链的制造组织能力;且制造组织能力的提升可推动企业核心技术的创新及促进市场多元化的发展,有利于企业核心竞争力加强。本论在现有研究成果的基础上,针对中国新能源整车制造商及供应商的发展策略及新能源汽车市场发展的政策建议进行了深入探析,从理论层面上助力中国新能源汽车企业的创新模式与制造组织能力的协同发展。

关键词:协同发展,创新模式,新能源汽车,供应链管理

## 1. 引言

为了促进汽车产业绿色发展,推动企业转型升级,中国采取了一系列相关政策支持新能源汽车产业发展。其能源清洁和智能高效的特点也逐渐成为推动经济可持续发展的重要引擎。近年来,伴随着经济全球化与日趋激烈的市场竞争,新能源汽车企业出现了过于依赖政策补贴,核心技术难以突破,创新动力下降等一系列问题(赵梓渝等,2021)。传统封闭的发展模式固然对传统燃油车的发展提供动力,但是对于新能源汽车企业来说,容易造成组织架构固化等负面影响,一定程度上阻碍企业的长期发展。

致力于解决以上问题,各企业正在不断地寻求新思路、新方法(熊勇清、秦书锋,2023)。广东省汽车产业作为中国重要的汽车生产基地之一,其创新模式和组织架构可为中国新能源汽车产业的供应链管理提供借鉴和参考经验。2020年国务院办公厅印发的《新能源汽车产业发展规划(2021-2035年)》(以下简称《规划》)明确了新能源汽车产业的发展目标。《规划》指出要加强新能

源汽车核心技术攻关工程,提高技术创新能力,要深化开放合作,调动相关企业的积极性,推动产业融合发展。

中国作为新能源汽车消费大国,新能源汽车产业发展前景良好,这有利于改善中国能源消费结构,推动企业转型升级(熊勇清、王溪,2021)。然而,较多企业目前难以从短暂的合作关系中积累到长期发展的经验,无法形成较强的核心竞争力。如何在资源有限的情况下,借助于主机厂与供应商之间的默契配合、供应商之间的竞争与合作,针对销售市场变化对企业现有(生产、采购、核心技术等)能力进行创新,对企业的长期发展具有重要作用。浅沼供应商理论将“关系的技能”定义为企业为了满足核心企业的需求或要求而有效率地作出反应。“制造组织能力”即为从顾客企业长期稳定的合作关系中积累到的“关系的技能”。本研究将从“产品设计能力”“工程工序设计能力”和“事业领域设计能力”三个方面分析新能源汽车企业的制造组织能力,探索如何通过企业创新构建适合的产业发展模式,从而实现协同发展。

进一步迎合市场，提升品牌竞争力，对推动新能源汽车产业的高质量发展具有重要意义。

## 2. 理论和假设

作为中国重点发展的新兴产业之一，新能源汽车产业的发展已驶入“快车道”。2015年以来，中国乘用车总销量处于高位，市场份额较高。随着政策优惠和配套环境日益优化，新能源汽车销量也稳步提升。2020年以来，新冠疫情带来的冲击逐渐缓和，汽车市场开始回暖。2022年中国新能源汽车销量达到688.7万辆，同比增长93.4%，连续8年保持全球第一。



图 1. 2015-2022 年中国汽车销售市场变化组合图 (单位: 万辆)

注: 数据来源于中国汽车工业协会官网  
[http://www.caam.org.cn/chn/4/cate\\_31/list\\_1.html](http://www.caam.org.cn/chn/4/cate_31/list_1.html) (2023 年 10 月 14 日)。

随着新一轮产业发展变革的深入，新能源汽车企业也需要针对自身发展模式探索新的发展路径，提升其品牌竞争力(Zhang and Li, 2021)。其中，一汽大众自 2020 年高尔夫纯电投入市场以来，凭借其良好的技术实力、品牌信誉及稳定的售后服务等优势，现稳居年中国新能源销量榜前十。为了符合新能源的设计生产，一汽 VW 采用全新的纯电动汽车平台——MEB (Make Everything Better) 平台，主要用于工艺变革及设备改造，促进传统能源汽车逐渐向新能源汽车转型的目标的实现。由此，本论选取一汽大众主机厂和零部件供

应商为调研对象并收集相关数据，为后续的研究做充分的准备。

不同于传统汽车企业，新能源汽车具有智能化、便利化的特点，为了适应市场发展的新模式，企业需要逐步在生产端、产品端和消费端进行创新变革(孙薇、叶初升, 2023)。随着供应链管理理念的提出，采购模式从库存驱动逐步向订单驱动转变，新的市场需求拉动企业进行产品创新。例如，一汽大众的生产管理模式从预测订单生产模式逐步转变为客户订单与预测订单相结合的模式，增强了其市场竞争力。有学者(焦薇等, 2020)认为把握市场变化能够降低风险，提升制造能力，驱动新能源汽车企业的创新发展。深入思考创新模式(张静静等, 2020)对新能源汽车企业的作用关系，市场变化对新能源汽车企业产品设计能力具有正向作用，对主机厂的协调能力具有重要影响(索琪、李长升, 2023)。基于以上，本论提出以下两个假设：

H1: 市场变化对新能源汽车企业产品设计能力具有正向作用。

H2: 新能源汽车主机厂零部件采购模式的创新对产品设计能力具有提升作用。

日本学者浅沼万里(1997)从生产技术的角度，阐述了日本汽车零部件供应商制造组织能力的演进路线。制造组织能力是企业在与供应商长期合作关系中不断积累形成的默契配合的能力，并在市场竞争力，供应链绩效等多个方面体现出优势(孙川, 2006)。有学者(刘国巍、邵云飞, 2020)认为上下游企业的合作有利于资源整合和价值创造。新能源汽车主机厂和供应商之间通过长期稳定的合作逐步积累经验(王生年、赵爽, 2020)，在生产过程中不断自主优化，自主创新，最终在供应链整体上带来更大的效益(AKABANE, etc. 2017)。基于以上，提出假设：

H3: 新能源汽车供应商的工程设计能力对核心技术的创新具有正向作用。

对于新能源汽车企业来说，其目标客户的需求变化快，产品更新迭代速度快，协同发展的创新模式能够帮助企业快速适应环境(Ma, etc. 2023)(陈芳、睦纪刚, 2015)实现价值创造。例如，一汽大众对于供应商的开放管理，为供应商提供生产指导帮助其高效率生产，致力于在开放供应链体系下与民族供应商一起成长。有学者(侯光明等, 2021)基于协同创新和产业价值链展开研

究，提出良好的供应商关系能够提升企业价值；认为协同创新对新能源汽车企业长期发展有积极作用（韩炜等，2021）。因此，研究零部件供应商创新模式对于企业协同发展的影响因素，对增强产业综合竞争力具有重要意义。由此提出以下假设：

H4：新能源汽车企业制造组织能力的提升对增强市场竞争力具有正向作用。

协同发展作为企业创新模式的新方向，学者们(Zhang, 2020)从组织架构和作用路径展开研究。新能源汽车主机厂和供应商的协同创新有利于获取外界信息，占据早期市场优势（张长令等，2016）。从供应链管理的角度出发，如一汽大众致力于打造全过程数字化的高质量新能源样板基地，建立智慧工厂（周燕、潘遥，2019）。在保证企业在现有产品的基础上响应顾客创新需求，实现新产品、新业务的多元化经营，从而增强综合竞争力，为市场多元化提供动力，即企业事业领域能力的体现。因此，提出以下假设：

H5：新能源汽车主机厂和供应商的协同发展对市场多元化具有正向作用。

本研究分为市场，新能源汽车主机厂和供应商三个主体；首先、在市场方面分为销售市场变化、市场竞争力、市场多元化；其次、在新能源汽车主机厂和供应商之间探究其制造组织能力（即产品设计能力、工程工序设计能力和事业领域设计能力），进而在企业间合作关系的基础上，分析企业间协同发展对市场方面的综合作用。综上所述，本研究构建的概念模型如图 2 所示。其中，市场变化对新能源汽车主机厂的产品设计能力有影响，同时主机厂的制造组织能力也会影响企业核心竞争力。主机厂和供应商之间通过采购或技术创新等因素相互影响，主机厂和供应商构成的协同发展模式也作用于市场多元化。

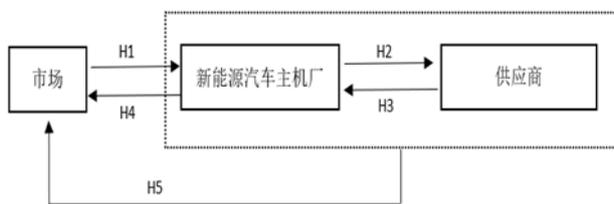


图 2 理论模型框架

### 3. 研究设计

#### 3.1 问卷设计

为了保证变量的测量效度，使其与中国新能源汽车主机厂和供应商之间的关系情况更加契合。研究者通过预调研，采用走访和座谈会的形式，收集并整理当前市场情况，对量表进行完善。问卷设计共分为四个部分：第一部分是人口统计学变量调查，包括性别、年龄、职位等，第二部分是企业在市场和创新方面的调查，第三部分是关于制造组织能力的调查，包括创新情况和协同发展情况等，第四部分则调查企业未来的发展方向。为了避免选项集中趋势，本论采用 likert5 级量表对变量进行衡量，其中 1 代表非常不同意，5 代表非常同意。

#### 3.2 描述性统计分析

本论研究对象为中国汽车主机厂及零部件供应商，共计向相关汽车企业发放问卷 255 份。剔除具有明显规律性的问卷及有悖于反转问题的问卷，总计回收有效问卷 220 份，有效回收率为 86%。

表 1 问卷调查描述性统计分析表

基本属性	分类	频次	百分比
年龄	18-25 岁	15	6.82%
	26-40 岁	164	74.55%
	41-50 岁	37	16.82%
	51 岁以上	4	1.82%
职业类型	企业所有者	6	2.73%
	供应链关系主管	43	19.55%
	客户关系主管	14	6.36%
	生产管理主管	60	27.27%
	其他	97	44.09%

如表 1 所示，受访者年龄分布呈正态分布，主要集中于 26—50 岁的中层管理者。其次、在受访者职业类型中，与本研究直接相关的管理层人员如企业所有者、供应链关系、客户关系及生产管理主管人数占比 55.81%、其他为企业基层管理者如班组长以上岗位。图 3 为本研究调查问卷对象企业的企业属性及事业领域相关信息。本次问卷调查合资零部件供应商占比 51%、民族资本零部件供应商占比 24%、外资零部件供应商占比 23%、整车制造商为 2%。其中，企业主营产品为

内饰相关零部件的生产销售占比 49%、从事发动机、变速箱、底盘或电池、电机、电控等零部件生产销售的企业占比 40%，从事钣金件等相关零部件生产销售的企业占比 8%，整车生产与销售企业占比 3%。

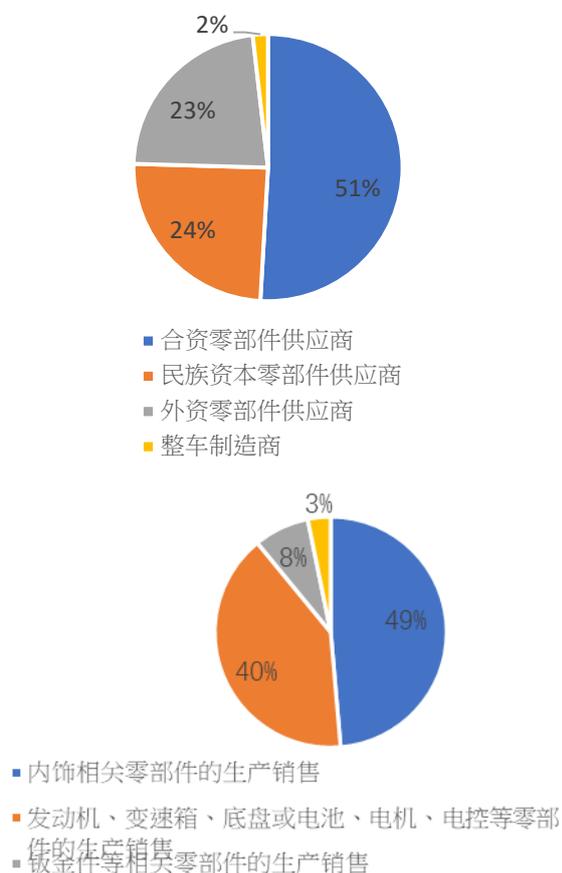


图 3. 问卷调查描述性统计分析图

### 3.3 变量选择

当前激烈变化的市场情况下，新能源汽车主机厂会更加注重迎合市场需求，探索新的商业模式，降低牛鞭效应的影响。企业制造组织能力更是高效决策和长期健康发展的重要指标，进一步推动主机厂和供应商之间的协同发展。综上，以图 2 概念模型为基础，本论选择创新模式、制造组织能力和协同发展作为研究变量。

#### (1) 创新模式

创新模式可以使企业在采购、生产、销售等环节不断变革和更新，采取更合适自身和目标市场的发展模式，从而提升核心竞争力。已有文献通过研发投入和专利数量来衡量企业创新，但是由于企业创新并不是全部体现在实质性的产出，

零部件供应商的创新也很难从专利数量上体现。因此，基于以往学者研究，本论将创新模式的衡量分为市场预测 (X1)、研发投入 (X2) 以及采购创新 (X3) 三个方面。将产品设计能力 (y1) 和工程工序设计能力 (y2) 分别作为因变量得到回归模型。

$$y_1 = 0.341 + 0.128X_1 + 0.524X_2 + 0.242X_3$$

$$y_2 = 2.601 + 0.091X_1 + 0.444X_2$$

(1)

#### (2) 制造组织能力

基于长期稳定的合作关系，企业在追求产品品质、成本、准时交货期的同时，在设计、开发、生产、交货的各个阶段不断积累定向的关系技能，从而形成制造组织能力的进阶。本论将自变量制造组织能力分为产品设计能力 (X4)，工程工序设计能力 (X5) 以及事业领域设计能力 (X6) 三个方面来衡量。产品设计能力即企业对于产品图纸设计方面的参与程度，例如从借出图成长为承认图。工程工序设计能力即企业对于生产线及设备的设计优化能力。事业领域能力即企业针对市场多元化采取的系统化、规模化、专一化的生产模式。以市场竞争力 (y3) 为因变量构建回归模型如下。

$$y_3 = 0.445 + 0.083X_4 + 0.022X_5 + 0.113X_6$$

(2)

#### (3) 协同发展

制造组织能力的提升需要双方或多方共同推动，能够在企业发展的过程中资源互补，促进创新。已有学者通过企业间分工和协调机制等因素表现协同发展。本论基于创新模式和制造组织能力之间的联系，从企业角度出发，采用协同合作的意愿 (X7) 和优势互补的程度 (X8) 为自变量，市场多元化 (y4) 作为因变量，该回归模型如下。

$$y_4 = 1.06 + 0.325X_7 + 0.38X_8$$

(3)

### 3.4 信度和效度检验

信息指量表的一致性和稳定性，主要通过 Cronbach's  $\alpha$  系数和组成信度 CR(composite reliability) 来检测。本论的数据处理软件为 Excel 和 SPSS Statistics 26.0。由表 1 所示，Cronbach's  $\alpha$  系

数和组成信度 CR 均大于 0.6，平均方差提取量 AVE 大于 0.5，且在 0.001 水平上显著，说明量表具有较好的内部一致性和收敛效度。

表 2. 信效度分析结果

潜变量	观察变量	符号	标准负荷	Cronbach's $\alpha$	组成信度 CR	AVE
创新模式	市场预测	X <sub>1</sub>	0.800	0.714	0.8268	0.6166
	研发投入	X <sub>2</sub>	0.673			
	采购创新	X <sub>3</sub>	0.870			
制造组织能力	产品设计能力	X <sub>4</sub>	0.776	0.727	0.7512	0.6016
	工程工序设计能力	X <sub>5</sub>	0.739	0.695	0.7094	0.5525
	事业领域设计能力	X <sub>6</sub>	0.821	0.781	0.8047	0.6733
协同发展	协同合作的意愿	X <sub>7</sub>	0.850	0.871	0.8229	0.6991
	优势互补的程度	X <sub>8</sub>	0.822			

## 4. 回归分析

### 4.1 回归结果分析

为了防止变量之间的多重共线性的问题，本论选取线性回归分析法。依据理论假设和变量定

义构建回归模型，结果如下表所示，VIF 值均小于 5，说明模型不存在多重共线性的问题，模型构建良好。由 F 检验的结构分析可知，显著性 P 值均为 0.000\*\*\*，在水平上呈现显著性。

表 3. 回归分析结果

		B	标准误	Beta	t	P	VIF	R <sup>2</sup>	调整 R <sup>2</sup>	F
H1-H2	常数	0.341	0.311	-	1.098	0.273	-	0.407	0.399	F=49.456 P=0.000***
	X <sub>1</sub>	0.128	0.069	0.105	1.864	0.064*	1.162			
	X <sub>2</sub>	0.524	0.064	0.481	8.123	0.000***	1.279			
	X <sub>3</sub>	0.242	0.069	0.202	3.505	0.001***	1.215			
因变量：产品设计能力										
H3	常数	2.601	0.254	-	10.26	0.000***	-	0.295	0.288	F=45.345 P=0.000***
	X <sub>1</sub>	0.091	0.06	0.093	1.517	0.131	1.157			
	X <sub>2</sub>	0.444	0.054	0.502	8.183	0.000***	1.157			
因变量：工程设计能力										
H4	常数	0.445	0.276	-	1.616	0.108	-	0.522	0.511	F=46.81 P=0.000***
	X <sub>4</sub>	0.083	0.059	0.085	1.358	0.178	1.747			
	X <sub>5</sub>	0.022	0.064	0.021	0.35	0.727	1.612			
	X <sub>6</sub>	0.113	0.069	0.102	1.57	0.37	1.774			
因变量：市场竞争力										
H5	常数	1.06	0.26	-	4.082	0.000***	-	0.374	0.368	F=64.832 P=0.000***
	X <sub>7</sub>	0.325	0.089	0.307	3.636	0.000***	2.474			
	X <sub>8</sub>	0.38	0.094	0.342	4.054	0.000***	2.474			
因变量：市场多元化										
注：***、**、*分别代表 1%、5%、10%的显著性水平										

增长点,从而扩大市场占有率。因此,假设 5 得到验证。

## 4.2 假设检验结果讨论

基于表 2 及表 3 可知,本论提出的假设全部成立,接下来将逐一进行深入分析。市场变化对新能源汽车企业产品设计能力具有正向作用。回归分析结果显示,市场因素与产品设计能力之间的回归系数值(B 值)大于 0,表明预测市场的能力与产品设计能力呈正相关关系。具体而言,企业对于市场消费需求和消费趋势的把握,将有利于制定下一步的生产计划。这种预测市场变化的能力越强,对于产品更新改造的能力便越清晰。并且拟合系数 R 方为 0.407,说明变量之间具有 40% 的解释能力。因此,验证假设 1 成立。

零部件采购模式的创新对产品设计能力具有提升作用。随着市场环境的变化,供应链的优势逐步在企业发展中体现出来。在按照物料清单采购零部件的传统模式上,逐步增加对于小规模零部件的采购,以供企业满足个性化订单和产品创新的需求。由模型结果可知,采购模式的创新与产品设计能力之间具有显著正相关关系,假设 2 得到验证。

新能源汽车供应商的工程工序设计能力对核心技术的创新具有正向作用。企业在生产的过程中逐步生产线规模化,生产设备智能化。不仅能够提高生产效率,而且在优化的过程中,进一步改造生产技术,积累经验,有利于技术的创新。由回归结果可知,回归系数值(B 值)为正,表明企业的工程工序设计能力与企业核心技术的创新呈正相关关系,假设 3 得到验证。

制造组织能力的提升对新能源汽车企业增强市场竞争力具有正向作用。在上下游成员企业的长期合作中,新能源汽车企业的制造组织能力逐步提升。在产品的图纸设计,工程工序的优化以及事业领域探索三个方面的表现将直接作用于市场竞争力。从回归结果来看,回归系数(B 值)为正,具有显著地正相关关系,并且解释程度达到 52%,模型基本满足要求,假设 4 得到验证。

主机厂和供应商的协同发展对市场多元化具有正向作用。根据回归结果可知,标准化系数(B 值)达到 0.325 和 0.38,说明协同发展对于市场多元化具有正向影响关系。具体而言,新能源汽车主机厂和供应商在建立长期的合作关系之后,不断扩展经营空间,通过产品创新建立新的消费

## 5. 研究结论和启示

为了更好地把握企业转型升级的正向推动条件,本论采用中国汽车企业的调查数据,构建回归模型,验证了创新模式和制造组织能力协同发展的正向作用。从市场、供应链管理、制造组织能力和协同发展四个维度得出以下研究结论和启示。

据一汽大众官方数据显示,2023 年,一汽大众 ID.系列车型累计销量达到 4.87 万辆,获得中国市场纯电 SUV 合资品牌销量冠军。企业在适应销售市场的变化过程中不断调整运营策略,不断创新优化产品图纸,采用先进的生产技术和高效的管理方法,以在激烈的市场竞争中脱颖而出,实现长远的发展目标。新能源汽车转型升级以来,市场竞争愈加激烈,一汽大众在保证质量的前提下推出了包括 ID.3、ID.4 CROZZ、ID.6 CROZZ 等至少 20 款新能源车型。在与顾客企业的长期合作中,逐步积累经验,提升产品设计能力,从产品多样性的角度收获消费者的青睐,逐步向市场多元化发展。因此,企业要立足于新的市场需求,提高对销售市场信息的迅速感知能力和及时处理能力。企业管理者应充分认识到与顾客企业长期合作关系的重要性,即重视制造组织能力的进阶(周燕、潘遥,2019)。同时,政府要继续加强保障市场信息透明化,完善市场监管制度,充分发挥市场在资源配置中的作用,逐步推进企业转型升级。

零部件采购方式的变革帮助企业完善产品设计图纸,推动产品设计能力的提升。对于新能源汽车来说,采购模式的创新不仅能够提高生产过程的稳定性和效率,而且在完成个性化订单的同时,积累到的生产经验则会转变为对产品图纸的改进能力,体现在企业后续的产品优化设计中。在供应链管理方面,零部件的采购是供应链中重要的环节之一。企业在计划、采购、生产、销售等核心环节中突破,实现精细化、规模化管理,最大化实现降本增效,有利于建立高质量发展的长效机制(韩炜等,2021)。因此,在政策上建议立足于创新的发展模式,将重点逐步转向运营环节,在供应链环节上提高效率,为企业转型升级的路径探索作铺垫。

从制造组织能力上, 工程工序的系统化设计可以帮助企业发现生产流程中产品设计的不足, 增强对核心技术的创新能力。新能源汽车产业是技术密集型产业, 优良的生产线运作能够优化产品设计, 提高劳动生产率, 降低生产成本, 长期稳定的合作关系也会加持提升企业制造组织能力。随着新能源汽车产业链在各个环节逐步成熟, 建议政府加大创新支持力度, 为产业发展提供技术和人才保障。企业也要建立协调发展机制, 在现有的资源和优势的条件下, 适时引入智能化设备和高素质人才。通过创新模式的增持与上下游供应商协同发展, 向产品多元化和生产规模化发展, 才能促进新能源汽车产业发展再上新台阶。

面对当前市场的新环境, 多方合作共赢是企业长期稳健发展的必由之路。在协同发展方面, 主机厂和供应商协同发展的配合程度越高, 企业对于潜在市场的拓展能力越强, 市场竞争力越强。相较于其他以短期发展为主的企业来说, 具备良好制造组织能力的企业面对市场环境的变化和企业竞争的加大时, 能够领先一步展开行动, 找到合适的解决方法, 更加迅速地在供应链上建立早期竞争优势, 实现多元化、高效率的发展模式。随着智能化技术的推广加速, 在政策上应当引导多领域推进新能源汽车的使用, 在建立数字化平台上给予支持。例如, 一汽大众在保障正常生产的基础上为线上消费者定制化购车提供上万种搭配选择方案, 致力于打造整合的线上数字化营销平台。既满足了新生代消费者用户的个性化需求, 又提高产品及品牌的竞争力。最后, 新能源汽车企业在塑造制造组织能力和创新能力的过程中, 也要充分评估现有的发展阶段。分类有序地推进产业政策, 合理运用资源优势, 强化落实措施, 形成良性市场竞争机制。

## 6. 结语

本论在新能源汽车企业制造组织能力的基础上, 验证了创新模式与其协同发展的正向影响因素, 从市场、主机厂和供应商方面进行了彼其间协同发展的理论探索, 并提出了实践建议, 一定程度上弥补了理论与实践之间的空隙, 为企业在激烈的市场竞争中提供了更准确具体的战略支持。但考虑到新能源汽车品种和供应商企业性质的多样性, 本论存在如下研究限制: 企业调研的数据均来自于一汽大众佛山工厂及其所属零部件

供应商, 在数据采集的广度上有待在其后的研究中进一步加强。其次, 新能源产业作为新兴产业, 产业链尚未完全成熟, 而一汽大众作为传统燃油车转型新能源的企业, 与中国造车新势力企业存在产品设计、供应链管理及销售模式上存在差异。因此, 将本论的研究方法付诸于中国造车新势力的研究亦是接下来的重点研究方向。

## 基金项目:

广东省哲学社会科学规划项目“粤港澳大湾区生产性服务业与制造业高质量发展的协同演化机制研究”(编号:GD20CYJ15)、教育部人文社科青年项目“中国生产性服务业与制造业高质量发展的协同演化机制与路径研究”(编号:21YJC790058)、2023 广东省社科共建项目“广东省汽车产业供应链与创新链协同发展的路径研究”(编号:GD23XGL013)。

## 参考文献

- 赵梓渝、王士君、陈肖飞, 2021, 模块化生产下中国汽车产业集群空间组织重构——以一汽-大众为例, 地理学报, 76(8):1848-1864。(Zhao and Wang and Chen, 2021)
- 熊勇清、秦书锋, 2023, 新能源汽车产业政策促进了何种创新?, 科研管理, 44(3):102-111。(Xiong and Qin, 2023)
- 熊勇清、王溪, 2021, 新能源汽车异质性需求的创新激励效应及作用机制——“政府采购”“商业运营”与“私人乘用”需求比较的视角, 财经研究, 47(7):48-62。(Xiong and Yu, 2021)
- 孙薇、叶初升, 政府采购何以牵动企业创新——兼论需求侧政策“拉力”与供给侧政策“推力”的协同, 中国工业经济, (1):1-19。(Sun and Ye, 2023)
- 焦薇、倪明、吴知轩, 基于供应商补偿激励的汽车主机厂联合采购策略, 系统工程学报, 35(6):760-772+805.n。(Jiao and Ni and Wu, 2020)

张静静、刘璐、李剑玲，生态消费视角下的新能源汽车商业模式创新研究，生态经济，36(3):72-77。(Zhang and Liu and Li, 2020)

索琪、李长升，技术-合作网络视角下中国新能源汽车产业演化动力机制研究，管理学报，20(3):329-338。(Suo and Li, 2023)

孙川，日本企业下包制度理论模型分析—以佐藤模型和浅沼模型为对象，现代日本经济，(1):5-10。(Sun, 2006)

刘国巍、邵云飞，产业链创新视角下战略性新兴产业合作网络演化及协同测度—以新能源汽车产业为例，科学学与科学技术管理，41(8):43-62。(Liu and Shao, 2020)

王生年、赵爽，社会信任、供应商关系与企业创新，中南财经政法大学学报，(6):25-34。(Wang and Zhao, 2020)

陈芳、睦纪刚，新兴产业协同创新与演化研究:新能源汽车为例，科研管理，36(1):26-33。(Chen and Sui, 2015)

侯光明、景睿、石秀，中国新能源汽车企业协同创新模式的创新绩效及作用路径研究，技术经济，40(11):13-22。(Hou and Jing and Shi, 2021)

张长令、马焱、杜玖玉，市场开放、资金补贴与新兴产业市场演化—以新能源汽车产业为例，上海经济研究，(5):47-57。(Zhang and Ma and Du, 2016)

韩炜、杨俊、胡新华等，商业模式创新如何塑造商业生态系统属性差异?—基于两家新创企业的跨案例纵向研究与理论模型构建，管理世界，37(1):88-107+7。(Han and Yang and Hu, 2021)

周燕、潘遥，财政补贴与税收减免—交易费用视角下的新能源汽车产业政策分析，管理世界，35(10):133-149。(Zhou and Pan, 2019)

(2017). From product design to product ,process and domain design capabilities of local tier 2 suppliers: lessons form case studies in Japan, Thailand and China. *Int. J. Automotive Technology and Management*, 17(4)385-408. MA Miaomiao & MENG Weidong & LI Yuyu, et al. (2023). Impact of dual credit policy on new energy vehicles technology innovation with information asymmetry, *Applied Energy*, 332.

CHEN Fang & WEI Jigang. (2015). *Research Management*,2015,36(1):26-33. (in Chinese)

HAN Wei & YANG Jun & HU Xinhua, etc. (2021). How does business model innovation shape the differences in business ecosystem attributes? A cross-case longitudinal study and theoretical model construction based on two start-ups, *Management World*, 37(1):88-107+7. (in Chinese)

HOU Guangming & JING Rui & SHI Xiu. (2021) Research on innovation performance and role path of collaborative innovation model of new energy vehicle enterprises in China, *Technology Economics*, 40(11):13-22. (in Chinese)

Jiao Wei & Ni Ming & Wu Zhixuan. (2020). Joint procurement strategy of automobile OEMs based on supplier compensation incentive, *Journal of Systems Engineering*, 35(6):760-772+805.n (in Chinese)

jie Zhang.(2020). Research on development mechanism of new energy vehicle enterprises based on open source innovation model, *IOP Conference Series: Earth and Environmental Science*, 619(1):012028.

LIU Guowei & SHAO Yunfei. (2020). Evolution and Collaborative Measurement of Cooperation Network of Strategic Emerging Industries from the Perspective of Industrial Chain Innovation: A Case Study of New Energy Vehicle Industry, *Science of Science and Management of Science and Technology*, 41(8):43-62. (in Chinese)

SUN Chuan. (2006). Analysis of the theoretical model of the outsourcing system of Japanese enterprises: A case study of Sato model and Asanuma model, *Modern Japanese Economy*, (1):5-10. (in Chinese)

## References

AKABANE Jun & Tsuchiya Yasuo & Inoue Ryuichirou & Yamamoto Hajime & Yang Zhuang.

SUN Wei & YE Chusheng. (2023). Why does government procurement affect enterprise innovation: On the synergy between the "pull" of demand-side policies and the "push" of supply-side policies, *China Industrial Economics*, (1):1-19. (in Chinese)

Suo Qi & Li Changsheng. (2023). Research on the dynamic mechanism of the evolution of China's new energy vehicle industry from the perspective of technology-cooperation network, *Chinese Journal of Management*, 20(3):329-338. (in Chinese)

WANG Shengnian & ZHAO Shuang. (2020) *Journal of Zhongnan University of Economics and Law*, (6):25-34. (in Chinese)

Xiong Yongqing & Qin Shufeng. (2023). What kind of innovation has the new energy vehicle industry policy promoted?, *Scientific Research Management*, 44(3):102-111. (in Chinese)

Xiong Yongqing & Wang Xi. (2021). Innovation incentive effect and mechanism of heterogeneous demand for new energy vehicles: A perspective on the comparison of "government procurement", "commercial operation" and "private passenger" demand, *Journal of Finance and Economics*, 47(7):48-62. (in Chinese)

Xueyu Zhang & Wenyong Li. (2021). Research on the incentive of government subsidy in the innovation ecosystem of guangzhou new energy automobile industry, *E3S Web of Conferences*, 235:01001

ZHANG Changling & MA Ben & DU Jiuyu. (2016). Market opening, capital subsidy and the evolution of emerging industry market: A case study of new energy vehicle industry, *Shanghai Economic Research Journal*, (5):47-57. (in Chinese)

ZHANG Jingjing & LIU Lu & LI Jianling. (2020). Research on business model innovation of new energy vehicles from the perspective of ecological consumption, *Ecological Economy*, 36(3):72-77. (in Chinese)

ZHAO Ziyu. & WANG Shijun & CHEN Xiaofei. (2021). Reconstruction of spatial organization of

China's automobile industry cluster under modular production: A case study of FAW-Volkswagen, *Acta Geographica Sinica*, 76(8):1848-1864. (in Chinese)

ZHOU Yan & PAN Yao. (2019). Financial subsidy and tax reduction: policy analysis of new energy vehicle industry from the perspective of transaction cost, *Management World*, 35(10):133-149. (in Chinese)

## 作者簡介



楊壯自 2018 年進入高校任職，現為廣東科技學院管理學院專任講師。他曾掛職唯品會資訊科技有限公司任自動化管理總部研發與創新專員。楊老師從日本櫻美林大學（J.F.Oberlin）獲得經濟學博士學位、工商管理碩士學位、並從長春光華學院獲得文學學士學位。研究領域包括中日汽車產業、製造組織能力、中小企業革新能力、危機管理。

## INSTRUCTIONS TO AUTHORS

### *Submission of papers*

The International Journal of Systematic Innovation is a refereed journal publishing original papers four times a year in all areas of SI. Papers for publication should be submitted online to the IJoSI website (<http://www.ijosi.org>) In order to preserve the anonymity of authorship, authors shall prepare two files (in MS Word format or PDF) for each submission. The first file is the electronic copy of the paper without author's (authors') name(s) and affiliation(s). The second file contains the author's (authors') name(s), affiliation(s), and email address(es) on a single page. Since the Journal is blind refereed, authors should not include any reference to themselves, their affiliations or their sponsorships in the body of the paper or on Figs and computer outputs. Credits and acknowledgement can be given in the final accepted version of the paper.

### *Editorial policy*

Submission of a paper implies that it has neither been published previously nor submitted for publication elsewhere. After the paper has been accepted, the corresponding author will be responsible for page formatting, page proof and signing off for printing on behalf of other co-authors. The corresponding author will receive one hardcopy issue in which the paper is published free of charge.

### *Manuscript preparation*

The following points should be observed when preparing a manuscript besides being consistent in style, spelling, and the use of abbreviations. Authors are encouraged to download manuscript template from the IJoSI website, <http://www.ijosi.org>.

1. *Language.* Paper should be written in English except in some special issues where Chinese may be acceptable. Each paper should contain an abstract not exceeding 200 words. In addition, three to five keywords should be provided.
2. *Manuscripts.* Paper should be typed, single-column, double-spaced, on standard white paper margins: top = 25mm, bottom = 30mm, side = 20mm. (The format of the final paper prints will have the similar format except that double-column and single space will be used.)
3. *Title and Author.* The title should be concise, informative, and it should appear on top of the first page of the paper in capital letters. Author information should not appear on the title page; it should be provided on a separate information sheet that contains the title, the author's (authors') name(s), affiliation(s), e-mail address(es).
4. *Headings.* Section headings as well as headings for subsections should start from the left-hand margin.
5. *Mathematical Expressions.* All mathematical expressions should be typed using Equation Editor of MS Word. Numbers in parenthesis shall be provided for equations or other mathematical expressions that are referred to in the paper and be aligned to the right margin of the page.
6. *Tables and Figs.* Once a paper is accepted, the corresponding author should promptly supply original copies of all drawings and/or tables. They must be clear for printing. All should come with proper numbering, titles, and descriptive captions. Fig (or table) numbering and its subsequent caption must be below the Fig (or table) itself and as typed as the text.
7. *References.* Display only those references cited in the text. References should be listed and sequenced alphabetically by the surname of the first author at the end of the paper. For example:

Altshuller, G. (1998). *40 Principles: TRIZ Keys to Technical Innovation*, Technical Innovation Center.  
Sheu, D. & Lee, H. (2011). A Proposed Process for Systematic Innovation, International Journal of Production Research, Vol. 49, No. 3, 2011, 847-868.

## The International Journal of Systematic Innovation Journal Order Form

<b>Organization Or Individual Name</b>	
<b>Postal address for delivery</b>	
<b>Person to contact</b>	Name: _____ e-mail: _____ Position: _____ School/Company: _____
<b>Order Information</b>	<p><b>I would like to order</b> ___ copy(ies) of the <i>International Journal of Systematic Innovation</i>:</p> <p><b>Period Start: 1<sup>st</sup>/ 2<sup>nd</sup> half</b> ____, <b>Year:</b> ____ (Starting 2010)  <b>Period End : 1<sup>st</sup>/ 2<sup>nd</sup> half</b> ____, <b>Year:</b> ____</p> <p><b>Price:</b>  <b>Institutions: US \$150 (yearly) / NT 4,500 (In Taiwan only)</b>  <b>Individuals: US \$50 (yearly) / NT 1500 (In Taiwan only)</b>  (Local postage included. International postage extra)  <b>E-mail to:</b> <a href="mailto:IJoSI@systematic-innovation.org">IJoSI@systematic-innovation.org</a> or fax: +886-3-572-3210</p> <p>Air mail desired <input type="checkbox"/> (If checked, we will quote the additional cost for your consent)</p>
<b>Total amount due</b>	US\$
<b>Payment Methods:</b> 1. <b>Credit Card (Fill up the following information and e-mail/ facsimile this form to The Journal office indicated below)</b> 2. <b>Bank transfer</b> 3. <b>Account:</b> The Society of Systematic Innovation 4. <b>Bank Name:</b> Mega International Commercial BANK 5. <b>Account No:</b> 020-53-144-930 6. <b>SWIFT Code:</b> ICBCTWTP020 7. <b>Bank code :</b> 017-0206 8. <b>Bank Address:</b> No. 1, Xin'an Rd., East Dist., Hsinchu City 300, Taiwan (R.O.C.)	

### VISA / Master/ JCB/ AMERICAN Cardholder Authorization for Journal Order

#### Card Holder Information

Card Holder Name	(as it appears on card)		
Full Name (Last, First Middle)			
Expiration Date	/ (month / year)	Card Type	<input type="checkbox"/> VISA <input type="checkbox"/> MASTER <input type="checkbox"/> JCB
Card Number	□□□□-□□□□-□□□□-□□□□	Security Code	□□□ 
Amount Authorized		Special Messages	
Full Address (Incl. Street, City, State, Country and Postal code)			

Please Sign your name here \_\_\_\_\_ (same as the signature on your card)

### The Society of Systematic Innovation

6 F, #352, Sec. 2, Guanfu Rd,  
Hsinchu, Taiwan, 30071, R.O.C.