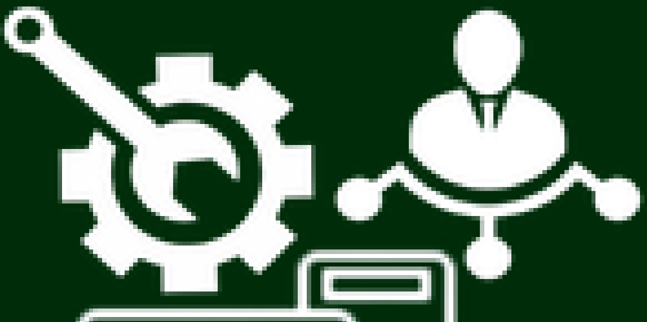
JCMM

Journal of Computers, Mechanical and Management

e-ISSN: 3009-075X

Volume 5, Issue 1

2026



AAN PUBLISHING

JCMM Journal of Computers, Mechanical and Management
Journal homepage: www.jcmm.co.in
e-ISSN: 3009-075X
AAN PUBLISHING

Editorial Comments Volume 5 Issue 1

Ritesh Bhat*[1,2]

[1]School of Advanced Manufacturing, Sricity International University, Sri City, Tirupati, Andhra Pradesh India 517646
[2]Journal of Computers, Mechanical and Management, AAN Publishing, Bandar Darulaman, Kedah Darulaman, Malaysia 06000

Volume 5, Issue 1 of the *Journal of Computers, Mechanical and Management (JCMM)* presents a diverse set of interdisciplinary research contributions that collectively highlight the growing convergence of artificial intelligence, engineering systems, secure digital infrastructures, and consumer analytics. The issue features **three original research articles** and **three review-oriented studies**, spanning applications in intelligent mechanical design, medical and agricultural diagnostics, cybersecurity, cloud privacy frameworks, and consumer behavior analysis.

Jiyaul Mustafa et al. [1] introduced a deep learning and generative artificial intelligence framework for the structural synthesis of epicyclic gear trains, demonstrating how automated topology generation and isomorphism detection can accelerate adaptive mechanical design for next-generation automation systems. **Chanchal Ghosh et al.** [2] proposed a hierarchical ensemble deep learning architecture for multi-class rice foliar disease diagnosis, achieving high classification accuracy while maintaining deployment feasibility for resource-constrained agricultural environments. **Shobana D. et al.** [3] presented an optimized Mask R-CNN–based brain tumor segmentation approach incorporating domain-specific pre-training and adaptive augmentation strategies, improving diagnostic recall and enabling efficient medical imaging analysis under real-time constraints.

Two comprehensive review contributions address emerging challenges in secure and trustworthy computational infrastructures. **Jayashri J. Patil and Ramkumar Solanki** [4] provided a systematic and deployment-conscious synthesis of adaptive machine learning methods for intrusion detection in Internet of Things (IoT) networks, emphasizing real-time adaptability, dataset realism, and resource-aware implementation. **Qing Guan et al.** [5] reviewed privacy-aware cloud architectures for secure medical e-governance data processing, examining encryption strategies, hybrid cloud models, blockchain-enabled access control, and scalability challenges in digital governance systems.

Extending the journal's interdisciplinary scope into management and behavioral domains, **Sonal Devesh et al.** [6] investigated brand loyalty drivers among Generation Z fashion consumers in the Indian context. Their empirical analysis revealed differentiated cognitive, emotional, and social pathways underlying loyalty formation across gender groups, offering insights into contemporary consumer behavior within digitally mediated markets.

Across these contributions, a unifying theme emerges: the integration of intelligent computational models with domain-specific knowledge to address complex real-world challenges under practical constraints of accuracy, scalability, and deployment feasibility. The studies collectively reflect the journal's mission to bridge advances in computing, mechanical and industrial engineering, and management science within modern technological ecosystems.

The Editorial Board expresses sincere appreciation to the authors, reviewers, and editorial team whose dedication ensures the scholarly rigor and quality of JCMM. As the journal progresses through its fifth volume, it continues to foster impactful interdisciplinary research at the intersection of intelligent computation, engineering innovation, and management systems.

[*] **Corresponding Author:** Ritesh Bhat (journalmanager@jcmm.co.in)

Published: 28 Feb 2026

© 2025 Journal of Computers, Mechanical and Management.
This is an open access article and is licensed under a Creative Commons Attribution-Non Commercial 4.0 License.

DOI: 10.57159/jcmm.5.1.25651.

References

[1] J. Mustafa, S. Ahmad, M. Wasid, M. A. Ansari, and S. A. Ansari, "Structural synthesis of epicyclic gear trains by deep learning and generative ai for adaptive automation," *Journal of Computers, Mechanical and Management*, vol. 5, no. 1, pp. 1–9, 2026.

[2] C. Ghosh, B. K. Das, T. Sur, P. Mazumdar, P. K. Halder, S. Kundu, and S. Prasad, "Hierarchical deep learning ensemble framework for multi-class rice foliar disease diagnosis: A comparative architecture analysis," *Journal of Computers, Mechanical and Management*, vol. 5, no. 1, pp. 20–34, 2026.

[3] S. D., V. V., M. P. A. Saviour, M. K., K. Sivamuni, and V. Thangasamy, "Advancing brain tumor detection: Optimized machine learning models for enhanced diagnostic accuracy," *Journal of Computers, Mechanical and Management*, vol. 5, no. 1, pp. 35–49, 2026.

[4] J. J. Patil and R. Solanki, "Advances in adaptive machine learning algorithms for enhanced security in iot networks: A comprehensive review," *Journal of Computers, Mechanical and Management*, vol. 5, no. 1, pp. 50–75, 2026.

[5] Q. Guan, M. N. H. B. Ibrahim, M. M. Alobaedy, and S. B. Goyal, "A systematic review of privacy-aware cloud framework for medical secure e-governance data processing," *Journal of Computers, Mechanical and Management*, vol. 5, no. 1, pp. 76–92, 2026.

[6] S. Devesh, N. Mudumbe, S. Mathan, P. Shukla, D. Gupta, and S. Gholve, "Brand loyalty drivers among generation z fashion consumers: A comparative analysis," *Journal of Computers, Mechanical and Management*, vol. 5, no. 1, pp. 10–19, 2026.

Journal of Computers, Mechanical and Management
Journal homepage: www.jcmm.co.in
e-ISSN: 3009-075X
AAN PUBLISHING

Volume 5 Issue 1

Article Number: 25247

Structural Synthesis of Epicyclic Gear Trains by Deep Learning and Generative AI for Adaptive Automation

Jiyaul Mustafa* [1], Shahnawaz Ahmad[2], Mohammed Wasid[3], Mohd Aquib Ansari[4], and Shaharyar Alam Ansari[2]

[1]Department of Mechanical Engineering, Bennett University, Greater Noida, Uttar Pradesh, India 201310
[2]School of Computer Science Engineering & Technology, Bennett University, Greater Noida, Uttar Pradesh, India 201310
[3]Department of Computer Science and Engineering, The LNM Institute of Information Technology, Jaipur, Rajasthan, India 302031
[4]School of Computing Science & Engineering, Galgotias University, Greater Noida, Uttar Pradesh, India 203201

Abstract

Structural synthesis of Epicyclic Gear Trains (EGTs) is a computationally demanding activity, particularly when identifying isomorphism between complex topologies and producing new gear designs to support high-performance automation systems. Graph-theoretic and algebraic methods are traditional and involve manual intervention and duplicate solutions. To address this shortcoming, this paper proposes a DL-based Generative AI framework for the automated synthesis and classification of EGTs. A Generative Adversarial Network (GAN) is trained on existing EGT topologies, learning their structures, creating new feasible structural mechanisms, and identifying duplication through degree sequence estimation and graph matching. The strategy is combined with the calculation of the connectivity matrix and the representation of the structural graph to ensure manufacturability and kinematic feasibility. The effectiveness of the proposed AI model is validated by the analysis of different EGTs, with 4–5 links, single DOF. Findings demonstrate that the GAN-based synthesis reliably distinguishes structurally distinct gear trains, eliminates pseudo-isomorphic designs, and saves a significant amount of design time. The technique justifies adaptive automation by designing intelligent mechanisms that require minimal human intervention. The paper demonstrates that AI-based synthesis can be highly effective in next-generation smart factories, robotic actuation, transmission systems, and reconfigurable automation platforms.

Keywords: Deep Learning; Generative AI; Epicyclic Gear Trains; Isomorphism Detection; Adaptive Automation

1. Introduction

With high power density, compact design, and variable speed ratios, EGTs are widely used in automotive transmission systems, aerospace actuators, robotics, and industrial automation. Topological synthesis of EGTs generates non-redundant configurations of unique structures and identifies isomorphism between possible structures.

*Corresponding Author: Jiyaul Mustafa (jiyaul.alld@gmail.com)

Received: 10 Dec 2025; **Revised:** 18 Jan 2026; **Accepted:** 23 Jan 2026; **Published:** 28 Feb 2026

© 2025 Journal of Computers, Mechanical and Management.
This is an open access article and is licensed under a Creative Commons Attribution-Non Commercial 4.0 License.

DOI: 10.57159/jcmm.5.1.25247.

Algorithms that rely on graph theory, adjacency matrices, Boolean algebra, or Hamming numbers are correct but computationally costly and require expert intervention. As intelligent manufacturing and Industry 4.0 become more prominent, DL and Generative AI offer opportunities to automate EGT synthesis. GANs, Variational Autoencoders (VAEs), and transformer-based sequence models can learn the structural properties of EGT topologies, generate new, valid topologies, and automatically identify duplicates. These models, when combined with graph-theoretic representation, facilitate adaptive automation, in which gear mechanisms are set up on command to meet load, ratio, and design constraints [1–4].

The algorithm was developed by Freudenstein [5] and is based on Boolean algebra, enabling the investigation of kinematic structure and contributing to dynamic analysis, computer-aided sketching, and animation. Wojnarowki and Lidwin [6] employed signal flow graphs in the kinematic study of planetary gear trains (PGTs) to compute angular velocities and transmission ratios. Uicker Jr and Raicu [7] presented a way to find kinematic chains and detect isomorphism, whereas Mrutyunjaya and Raghavan [8] presented a way to find isomorphism evaluation by Bocher's formulae in kinematic chains. Allen [9] presented a bond graph model for the kinematic and dynamic analysis of gear drive transmission systems, enabling the identification of equations for torque transmission and velocity ratios. The approach used by Day et al. [10] to analyse PGT speed ratios is the tabulation technique, which assists designers in synthesising the correct speed ratios. Gibson and Kramer [11] proposed a symbolic representation of the description of two-DOF spur planetary gear trains (SPGTs) and provided 22 fundamental EGT configurations to provide kinematic equations. Ravisankar and Mruthyunjaya devised a computerised method for the structural analysis and synthesis of kinematic chains [12]. A new technique [13] that employed the Wiener number to identify isomorphism in epicyclic geared mechanisms and planar kinematic chains was solved by Mustafa and Hasan to eliminate long-term short-tailed duplicity problems. Jiyaul et al. [14] have presented a graph theory methodology for creating non-duplicative one-degree-of-freedom EGTs, utilising adjacency matrices and Wiener numbers.

Compared to the modified path matrix, Mustafa et al. [15] compare the modified gradient and Bocher's technique in terms of reliability, computational efficiency, and structural properties for EGTs. Mustafa et al. [16] proposed a variation of the path-matrix methodology for detecting isomorphism in an epicyclic gear train, which is problematic in current methodologies, using case studies. The paper by Chu and Zou [17] provided information on the synthesis of the structure and topological graphs of planar multiple-joint and geared kinematic chains. To prevent pseudo-isomorphism, Yang et al. [18] proposed a new method for detecting rotational graphs and the canonical rotational graphs of PGTs, introducing the concept of graph representation. Gao and Hu [19] provided a deeper examination of the topological synthesis and kinematic analysis of planetary transmission systems using graph theory, with a focus on the relationship between speed ratios and topology. Kamesh et al. [20] developed an algorithm using the net distance approach for the quantitative analysis of isomorphism in KCs and EGTs with 1-2 degrees of freedom (DOF), providing a simple calculation for isomorphism detection. Using the theory of vertex incidence polynomials, a new, efficient algorithm, tested in practice by Mustafa et al. [21], is innovative in the domain of isomorphism detection in EGTs and is called the Innovative Modified Gradient Method.

Assur group synthesis was based on group and matroid theory by Morlin et al. [22], which provides a new mechanism design method. The paper by Alizade et al. [23] investigated structuring lower-class robot manipulators under the general constraint of one. Yang et al. [24] reviewed the intelligent design of planar mechanisms and highlighted future trends. Graph-based isomorphism detection has gained traction. Bouritsas et al. [25] enhanced the expressivity of graph neural networks, while Sun et al. [26] introduced a branch-chain matrix-based method for isomorphism identification. Further, Sun et al. [27] proposed similarity recognition techniques for kinematic chains. Several authors have suggested machine learning applications for the structural analysis of EGTs, railways, and the automotive sector, as detailed in [28–33].

This paper presents a novel GAN-based methodology for structural synthesis of EGTs and isomorphism detection. The method reduces dependency on trial-and-error design, enables rapid generation of valid mechanisms, and enhances design intelligence in smart automated systems.

2. Graphical Representation of EGTs

Isomorphism detection and topological analysis within EGTs rely heavily on graph theory. Graph theory is the branch of science that deals with graphs, edges, and vertices. The graph connects a set of lines (edges) to a set of points (vertices) for structural analysis. Figures 1 and 2 show the skeleton diagram and functional schematic of one-DOF EGTs.

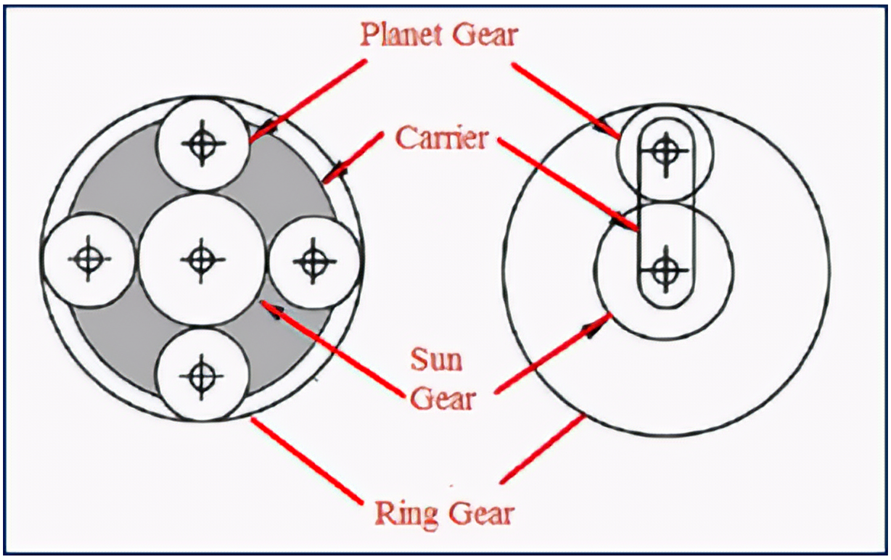


Figure 1: Skeleton diagram of four-link EGT

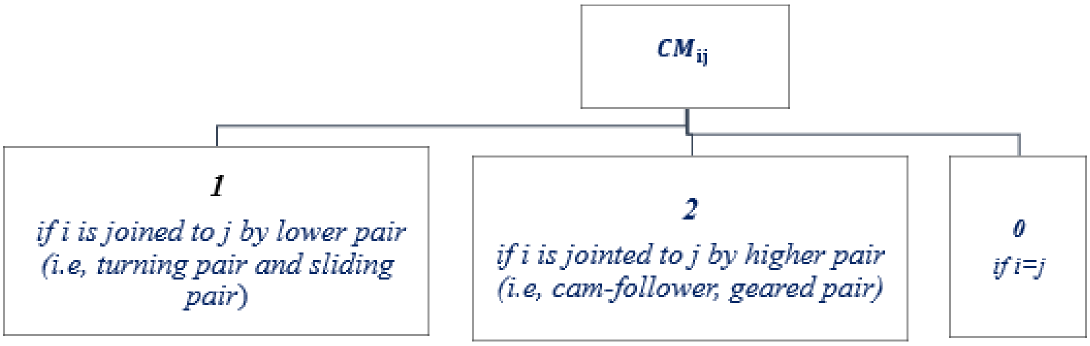

CM ij
1
if i is joined to j by lower pair (i.e, turning pair and sliding pair)
2
if i is jointed to j by higher pair (i.e, cam-follower, geared pair)
0
if i=j

Figure 2: Connectivity matrix chart for EGTs

The connectivity matrix shows the connection between two vertices or links in EGTs. The matrix demonstrates symmetry, with all its diagonal positions containing zeros. The matrix derives from its functional schematic and is shown as: For an $n \times n$ connectivity matrix can be written as,

$$[CM] = \begin{bmatrix} c_{11} & c_{12} & c_{13} & \cdots & c_{1n} \\ c_{21} & c_{22} & c_{23} & \cdots & c_{2n} \\ c_{31} & c_{32} & c_{33} & \cdots & c_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c_{n1} & c_{n2} & c_{n3} & \cdots & c_{nn} \end{bmatrix} \tag{1}$$

The connectivity matrix for four-link one-DOF EGTs is based on the given chart.

$$\text{Adjacency Matrix, } [Am] = \{a_{4\times4}\} = \begin{bmatrix} 0 & 1 & 1 & 2 \\ 1 & 0 & 2 & 1 \\ 1 & 2 & 0 & 3 \\ 2 & 1 & 3 & 0 \end{bmatrix} \tag{2}$$

3. Deep Learning and Generative AI Framework

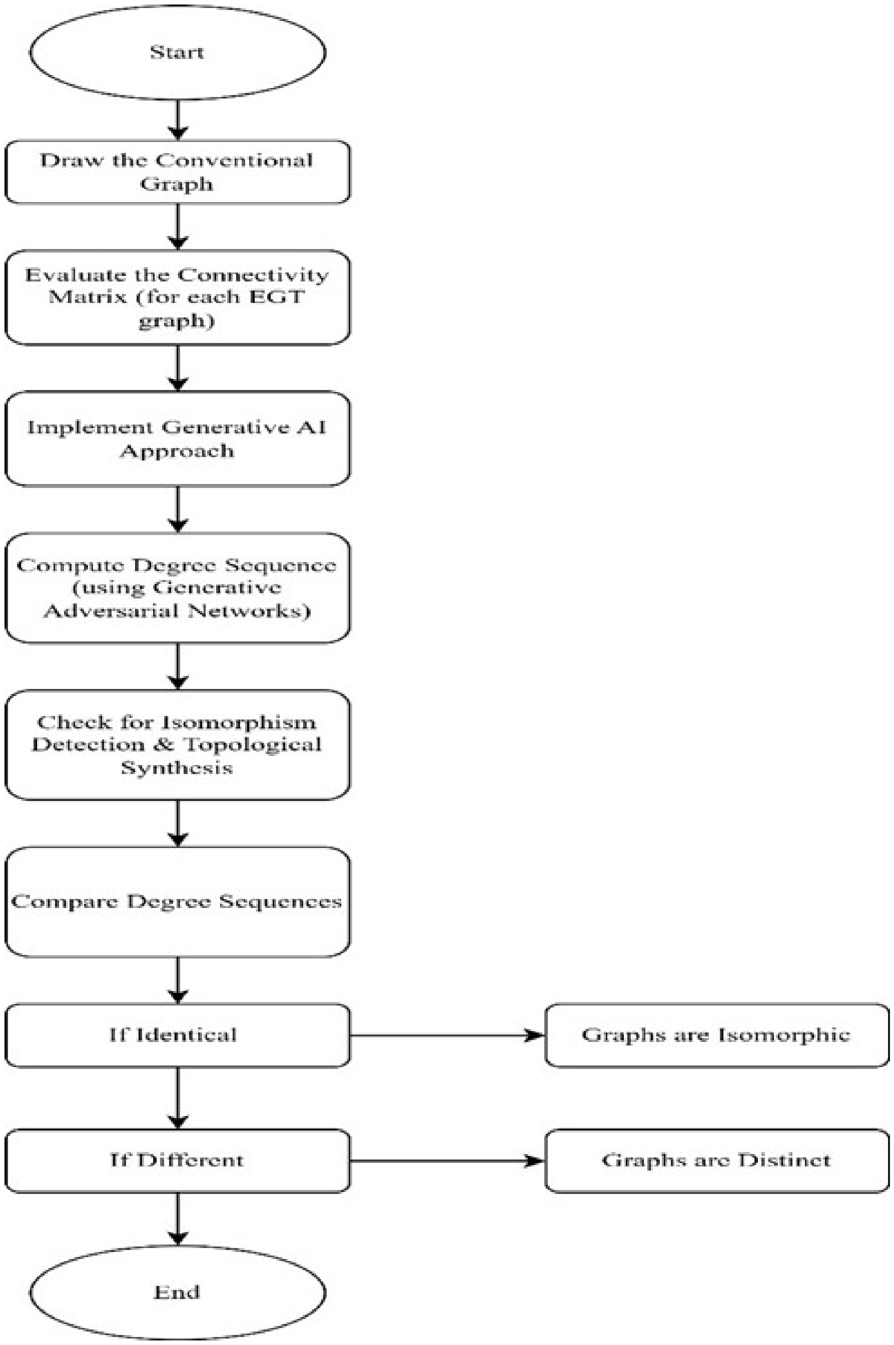

Start
Draw the Conventional Graph
Evaluate the Connectivity Matrix (for each EGT graph)
Implement Generative AI Approach
Compute Degree Sequence (using Generative Adversarial Networks)
Check for Isomorphism Detection & Topological Synthesis
Compare Degree Sequences
If Identical
Graphs are Isomorphic
If Different
Graphs are Distinct
End

Figure 3: Flow chart of complete methodology for DL and Generative AI approach for EGTs synthesis and analysis.

Recent advancements in artificial intelligence have introduced powerful generative models capable of learning the underlying structure of complex mechanical systems. For epicyclic gear train synthesis, the goal of a generative model is to learn the distribution of feasible gear topologies and automatically produce new, non-isomorphic configurations. Three key deep-learning approaches relevant to mechanism synthesis are GANs, VAEs, and Transformer-based self-attention models.

3.1. Dataset Preparation.

A curated dataset of 312 validated EGT topologies (4–8 links, 1-DOF) was prepared from published sources in Tsai papers in [34, 35] and manually verified models. Each topology is encoded as a binary connectivity matrix and normalized to 8×8 by padding. The dataset is split into 80% training and 20% testing sets. Each sample satisfied: graph symmetry, zero diagonal, at least one meshing pair, and valid carrier–gear–ring relations.

3.2. Preprocessing.

During preprocessing, the binary connectivity values $\{0, 1\}$ were transformed to $\{-1, +1\}$, all matrices were uniformly padded to a size of 8×8, and mechanically invalid or graph-inconsistent topologies were automatically removed using predefined structural rules.

3.3. GAN Architecture.

The GAN architecture employed in this study comprises a Generator and a Discriminator designed specifically for graph-structured EGT data. The Generator accepts a 100-dimensional noise vector as input, processes it through a dense layer, then reshapes it and applies a graph of graph convolutional layers integrated with batch normalisation. It produces an 8×8 connectivity matrix using a Tanh activation function, followed by post-processing with a threshold to convert the continuous output to a discrete adjacency matrix. The Discriminator, on the other hand, consists of graph convolutional layers coupled with LeakyReLU activations and dropout regularisation, culminating in a dense sigmoid layer that distinguishes real matrices from those generated by the model. Training was carried out using the WGAN-GP framework with 600 epochs, a batch size of 32, and the Adam optimiser (learning rate 0.0002, $\beta_1 = 0.5$). To ensure stable convergence, spectral normalisation, label smoothing, and gradient-penalty-based loss stabilisation techniques were incorporated throughout training.

3.4. Evaluation Metrics.

The performance of the GAN model was assessed using several evaluation metrics, including the Structural Validity Score (SVS), which measures the percentage of generated outputs that satisfy essential mechanical and graph-theoretic constraints; the Degree Sequence Uniqueness (DSU), which compares generated topologies with real samples to verify novelty; and the Graph Isomorphism Test, implemented through the VF2 algorithm, to identify and eliminate duplicate or pseudo-isomorphic structures. Additionally, the Frechet Graph Distance (FGD) was employed to quantify the distributional similarity between real and generated connectivity matrices, while the analysis of GAN loss trends throughout training served as an indicator of model stability and convergence. The flowchart of the complete methodology for the DL and Generative AI approach to EGT synthesis and analysis is shown in Figure 3.

4. Results and Discussion

The proposed DL and Generative AI algorithm is applied on various EGTs with different numbers of links and degrees of freedom for isomorphism detection and topological synthesis of mechanisms. The various examples are evaluated for detailed analysis of Generative AI-based GANs.

4.1. Deep Learning and Generative AI Technique in Two EGT Graphs (4-links and 1-DOF).

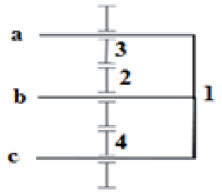
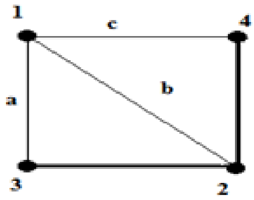
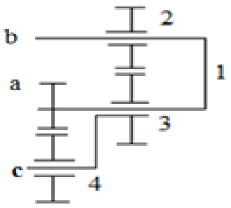
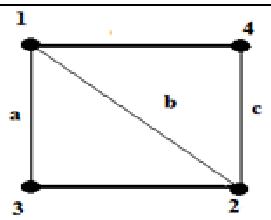
Sr. No.	Graph No.	Functional Schematic	Conventional Graph
1	EGT4101		
2	EGT4102		

Figure 4: Graph of four-links one-DOF EGT

Table 1: Results of degree sequences of two EGT graphs using the GANs method

Graph No.	Connectivity Matrix [CM]	Degree Sequences
EGT4101	$C_1 = \begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 2 & 2 \\ 1 & 2 & 0 & 2 \\ 1 & 2 & 2 & 0 \end{bmatrix}$	$D_1 = \{3, 5, 5, 5\}$
EGT4102	$C_2 = \begin{bmatrix} 0 & 1 & 1 & 2 \\ 1 & 0 & 2 & 1 \\ 1 & 2 & 0 & 3 \\ 2 & 1 & 3 & 0 \end{bmatrix}$	$D_2 = \{4, 4, 6, 6\}$

Table 1 shows $D_1 \neq D_2$. The two mechanisms are non-isomorphic and therefore structurally distinct EGTs (Figure 4). This confirms that the GAN can generate valid four-link mechanisms that do not duplicate known topologies.

4.2. Deep Learning and Generative AI Technique in Two EGT Graphs (5-links and 1-DOF).

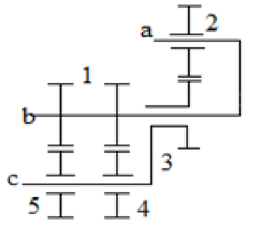
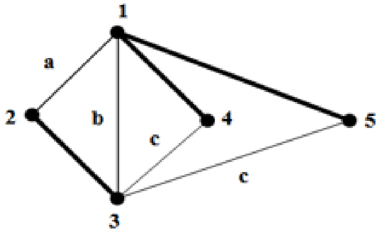
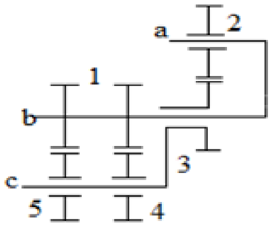
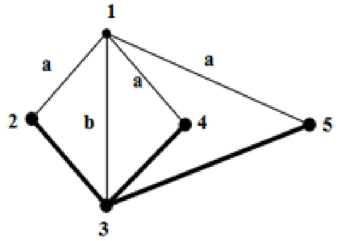
Sr. No.	Graph No.	Functional Schematic	Conventional Graph
1	EGT5101		
2	EGT5102		

Figure 5: Graphs of five-links one-DOF EGT

Table 2 shows $D_3 \neq D_4$. The two mechanisms are non-isomorphic and therefore structurally distinct EGTs (Figure 5). This confirms that the GAN can generate valid five-link mechanisms that do not duplicate known topologies. The comparative results of both illustrative cases are summarized in Table 3.

Table 2: Results of degree sequences of two EGT graphs using the GANs method

Graph No.	Connectivity Matrix [CM]	Degree Sequences
EGT5101	$C_3 = \begin{bmatrix} 0 & 1 & 1 & 2 & 2 \\ 1 & 0 & 2 & 3 & 3 \\ 1 & 2 & 0 & 1 & 1 \\ 2 & 3 & 1 & 0 & 2 \\ 2 & 3 & 1 & 2 & 0 \end{bmatrix}$	$D_3 = \{6, 9, 5, 8, 8\}$
EGT5102	$C_4 = \begin{bmatrix} 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 2 & 2 & 2 \\ 1 & 2 & 0 & 2 & 2 \\ 1 & 2 & 2 & 0 & 2 \\ 1 & 2 & 2 & 2 & 0 \end{bmatrix}$	$D_4 = \{4, 7, 7, 7, 7\}$

Table 3: Comparative summary of illustrative examples

Example	Links	DOF	Connectivity Matrix Compared	Degree Sequence	Result	Unique/Duplicate
Case 1	4-link	1	C_1 vs C_2	$D_1 = \{2,3,2,3\}$ vs $D_2 = \{2,3,2,1\}$	$D_1 \neq D_2$	Unique
Case 2	5-link	1	C_3 vs C_4	$D_3 = \{2,3,2,3,2\}$ vs $D_4 = \{2,3,4,3,2\}$	$D_3 \neq D_4$	Unique

5. Conclusions

This paper presented a deep-learning-assisted framework for the structural synthesis of EGTs using GANs and graph-theoretic isomorphism detection. Each gear train schematic was converted into a graph, encoded into a connectivity matrix, and analyzed through degree sequences to differentiate unique mechanisms from isomorphic duplicates. The GAN was trained on valid EGT data and demonstrated the ability to autonomously generate novel, manufacturable topologies without human intervention. Two representative examples, 4-link and 5-link (1-DOF), validated the methodology. In all cases, generated degree sequences were distinct from known designs, confirming successful creation of previously unexplored structural variants. The model consistently avoided redundant configurations, significantly reducing manual enumeration and post-processing efforts. All in all, the evidence shows that DL can automate and speed up structural synthesis, ensure validity, and remove topological redundancy. It is scalable to high-link and multi-degree-of-freedom planetary systems and can be used in compact transmissions, robotics, EV drivetrains, aerospace actuators, and adaptive automation.

Declaration of Competing Interests

The authors declare that they have no known competing financial interests or personal relationships that could have influenced the work reported in this paper.

Funding Declaration

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Ethical Approval

This study does not involve human participants, animals, or sensitive personal data. Ethical approval was not required.

Data Availability

The datasets generated and analyzed during this study are available from the corresponding author on reasonable request.

AI Use Disclosure

The authors used an AI-based language tool to improve grammar and readability. The scientific content and conclusions were reviewed and validated by the authors.

Author Contributions

Jiyaul Mustafa: Conceptualization, Methodology, Data Collection, Data Analysis, Manuscript Writing, Manuscript Revision; **Shahnawaz Ahmad**: Conceptualization, Methodology, Data Collection, Data Analysis, Manuscript Writing, Manuscript Revision; **Mohammed Wasid**: Conceptualization, Methodology, Data Collection, Data Analysis, Manuscript Writing, Manuscript Revision; **Mohd Aquib Ansari**: Data Analysis, Manuscript Writing, Manuscript Revision; **Shaharyar Alam Ansari**: Data Analysis, Manuscript Writing, Manuscript Revision

References

[1] F. Freudenstein, "The basic concepts of pólya's theory of enumeration, with application to the structural classification of mechanisms," *Journal of Mechanisms*, vol. 2, no. 3, pp. 275–290, 1967.

[2] L. Dobrjanskyj and F. Freudenstein, "Some applications of graph theory to the structural analysis of mechanisms," *ASME Journal of Engineering for Industry*, vol. 89, no. 1, pp. 153–158, 1967.

[3] Z. Levai, "Structure and analysis of planetary gear trains," *Journal of Mechanisms*, vol. 3, pp. 131–148, 1968.

[4] F. Buchsbaum and F. Freudenstein, "Synthesis of kinematic structure of geared kinematic chains and other mechanisms," *Journal of Mechanisms*, vol. 5, pp. 357–392, 1970.

[5] F. Freudenstein, "An application of boolean algebra to the motion of epicyclic drives," *ASME Journal of Engineering for Industry*, vol. 93, no. 1, pp. 176–182, 1971.

[6] J. Wojnarowski and R. Lidwin, "Application of signal flow graphs—the kinematic analysis of planetary gear trains," *Mechanism and Machine Theory*, vol. 10, pp. 17–31, 1975.

[7] J. J. U. Jr. and A. Raicu, "A method for the identification and recognition of equivalence of kinematic chains," *Mechanism and Machine Theory*, vol. 10, no. 5, pp. 375–383, 1975.

[8] T. S. Mrutyunjaya and M. R. Raghavan, "Structural analysis of kinematic chains and mechanisms based on matrix representation," *ASME Journal of Mechanical Design*, vol. 101, no. 3, pp. 488–494, 1979.

[9] R. R. Allen, "Multiport models for the kinematic and dynamic analysis of gear power transmissions," *ASME Journal of Mechanical Design*, vol. 101, no. 2, pp. 258–267, 1979.

[10] C. P. Day, H. A. Akeel, and L. J. Gutkowski, "Kinematic design and analysis of coupled planetary bevel-gear trains," *ASME Journal of Mechanisms, Transmissions, and Automation in Design*, vol. 105, no. 3, pp. 441–444, 1983.

[11] D. Gibson and S. Kramer, "Symbolic notation and kinematic equations of motion of the twenty-two basic spur planetary gear trains," *Journal of Mechanisms, Transmissions, and Automation in Design*, vol. 106, no. 3, pp. 333–340, 1984.

[12] R. Ravishankar and T. S. Mruthyunjaya, "Computerized synthesis of the structure of geared kinematic chains," *Mechanism and Machine Theory*, vol. 20, no. 5, pp. 367–387, 1985.

[13] J. Mustafa and A. Hasan, "Some application of graph theory to isomorphic analysis of epicyclic geared mechanisms," *Journal of the Institution of Engineers (India) Series C*, vol. 102, no. 4, pp. 1051–1057, 2021.

[14] J. Mustafa and A. Hasan, "Generation of one-dof epicyclic gear trains with up to eight links," *Materials Today: Proceedings*, vol. 15, no. 1, pp. 1123–1130, 2021.

[15] J. Mustafa and A. Hasan, "Comparative study between modified path matrix approach, modified gradient method and bocher's technique to detect isomorphism in egts," *Materials Today: Proceedings*, vol. 27, no. 3, pp. 1941–1948, 2021.

[16] J. Mustafa, A. Hasan, and R. A. Khan, "An application of modified path matrix approach for detection of isomorphism among epicyclic gear trains," *Journal of the Institution of Engineers (India) Series C*, vol. 101, no. 3, pp. 463–472, 2020.

[17] J. Chu and Y. Zou, "Topological graph descriptions and structural automatic synthesis of planar multiple joint and geared-linkage kinematic chains," *Mechanics Science*, vol. 36, pp. 12–19, 2016.

[18] W. Yang, H. Ding, B. Zi, and D. Zhang, "New graph representation for planetary gear trains," *ASME Journal of Mechanical Design*, vol. 140, no. 1, p. 012303, 2017.

[19] M. F. Gao and J. B. Hu, "Kinematic analysis of planetary gear trains based on topology," *ASME Journal of Mechanical Design*, vol. 140, no. 1, pp. 1–12, 2017.

[20] V. V. Kamesh, K. M. Rao, and A. B. S. Rao, "An innovative approach to detect isomorphism in planar and geared kinematic chains using graph theory," *ASME Journal of Mechanical Design*, vol. 139, no. 12, pp. 12–22, 2017.

[21] J. Mustafa, A. Hasan, and R. A. Khan, "An innovative approach for detection of isomorphism of epicyclic gear trains," *Materials Today: Proceedings*, vol. 25, no. 1, pp. 862–867, 2020.

[22] F. V. Morlin, A. P. Carboni, and D. Martins, "Synthesis of assur groups via group and matroid theory," *Mechanism and Machine Theory*, vol. 184, p. 105279, 2023.

[23] R. Alizade, S. Soltanov, and A. Hamidov, "Structural synthesis of lower-class robot manipulators with general constraint one," *Robotics*, vol. 10, no. 1, p. 14, 2021.

[24] W. Yang, H. Ding, and A. Kecskemety, "Structural synthesis towards intelligent design of plane mechanisms: current status and future research trend," *Mechanism and Machine Theory*, vol. 171, p. 104715, 2022.

[25] G. Bouritsas, F. Frasca, S. Zafeiriou, and M. M. Bronstein, "Improving graph neural network expressivity via subgraph isomorphism counting," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 1, pp. 657–668, 2022.

[26] L. Sun, X. Liu, X. Liu, X. Hong, H. Pei, and D. Zhang, "An isomorphism identification method of kinematic chain based on optimal arrangement and comparison of branch-chain matrix derived from dendrogram graph," *Advances in Mechanical Engineering*, vol. 14, no. 12, p. 16878132221131193, 2022.

[27] L. Sun, R. Cui, Z. Ye, Y. Zhou, Y. Xu, and C. Wu, "Similarity recognition and isomorphism identification of planar kinematic chains," *Mechanism and Machine Theory*, vol. 145, p. 103678, 2020.

[28] M. Wasid, A. Y. A. Amar, S. Ahmad, and J. Mustafa, "Enhancing reliability and security in online social networks: intelligent learning models for cybercrime detection," *Life Cycle Reliability and Safety Engineering*, pp. 1–17, 2025.

[29] U. C. Okonkwo, C. E. Okafor, S. Ahmad, C. C. Ohagwu, M. E. Aronu, I. P. Okokpujie, C. I. Idumah, N. N. Chukwu, C. E. Chukwunyelu, and J. Mustafa, "Research advancements in machine learning-assisted design of reinforced composite radiation shields," *Life Cycle Reliability and Safety Engineering*, pp. 1–35, 2025.

[30] N. K. Maurya, R. K. Singh, D. Gupta, R. Mishra, and J. Mustafa, "Evaluation of the mechanical properties of alsi10mg composites reinforced with tib$_2$ particles (wt%) fabricated via selective laser melting process," *Life Cycle Reliability and Safety Engineering*, pp. 1–9, 2025.

[31] J. Mustafa, S. Ahmad, and S. Hussain, "Machine learning-based data processing for predictive modeling in mechanical systems," *Life Cycle Reliability and Safety Engineering*, pp. 1–10, 2025.

[32] V. V. Asrith, S. Mann, S. Garg, and J. Mustafa, "Numerical and mathematical modelling of connecting rod of internal combustion engine," *Life Cycle Reliability and Safety Engineering*, pp. 1–9, 2025.

[33] S. Sharma, J. Mustafa, and S. Bhati, "Experimental study of gyroscopic effects on rotating disc," *Journal of the Institution of Engineers (India) Series C*, vol. 105, pp. 573–585, 2024.

[34] L. W. Tsai, *Mechanism Design: Enumeration of Kinematic Structures According to Function.* Boca Raton, FL, USA: CRC Press, 2001.

[35] L. W. Tsai, *Robot Analysis: The Mechanics of Serial and Parallel Manipulators.* New York, NY, USA: Wiley, 1999.

JCMM Journal of Computers, Mechanical and Management
Journal homepage: www.jcmm.co.in
e-ISSN: 3009-075X
AAN PUBLISHING

Volume 5 Issue 1

Article Number: 25245

Brand Loyalty Drivers among Generation Z Fashion Consumers: A Comparative Analysis

Sonal Devesh*, Naksha Mudumbe, Sumitra Mathan, Palak Shukla, Dia Gupta, and Siddhi Gholve

School of Business and Management, Christ University, Yeshwanthpur Campus, Bangalore, Karnataka, India 560073

Abstract

Brand loyalty is crucial in the competitive fashion market, particularly among Generation Z. Although previous studies have investigated what drives loyalty, there is still limited evidence from India, particularly about gender differences. This study adopts a context-specific and exploratory approach to examine brand loyalty and its drivers among Generation Z fashion consumers in Bangalore. The study adopts a quantitative research design with a structured questionnaire using a 5-point Likert scale. A sample of 100 Generation Z students in Bangalore was selected using convenience sampling to collect the data. Further descriptive and inferential statistical analyses were conducted using SPSS. The findings show positive associations among brand loyalty, brand awareness, perceived quality, emotional connection, and social influence. Independent-samples t-tests reveal no significant difference in overall brand loyalty between male and female respondents. However, regression analyses indicate that perceived quality and brand awareness are relatively stronger predictors of brand loyalty among male respondents. In contrast, emotional connection is a stronger predictor among female respondents. These findings suggest differences in motivational pathways rather than loyalty intensity. The study suggests that while overall brand loyalty levels are similar across genders, the motivational drivers underlying loyalty differ. These findings are context-specific and exploratory, and their generalizability is limited by convenience sampling and a restricted geographic scope.

Keywords: Generation Z; Fashion Brands; Consumer Behaviour; Fashion Consumption; Gender Differences; Brand Loyalty

1. Introduction

Brand loyalty plays a decisive role in determining success in the fashion industry, especially in one where trends change rapidly. Generation Z, people born between the late 1990s and early 2010s, is identified as technology-savvy and socially responsible [1]. Unlike previous generations, Gen Z consumers choose products based on criteria that go beyond product quality, including brand authenticity, alignment with values, and social media presence [2]. In Bangalore, a cosmopolitan city with influences from both global and native fashion trends, this study focuses on Gen Z students as a context-specific group to examine fashion-related brand loyalty, recognizing that their preferences are shaped by cultural exposure and interaction with global fashion trends, without implying representativeness beyond this demographic.

*Corresponding Author: Sonal Devesh (sonaldevesh@gmail.com)

Received: 9 Dec 2025; **Revised:** 14 Jan 2026; **Accepted:** 23 Jan 2026; **Published:** 28 Feb 2026

© 2026 Journal of Computers, Mechanical and Management.
This is an open-access article and is licensed under a Creative Commons Attribution-Non Commercial 4.0 License.

DOI: 10.57159/jcmm.5.1.25245.

Although the rising prominence of this generation has received considerable attention, relatively little research has examined how gender influences the underlying forces of brand loyalty, specifically in the Indian context. According to past studies, male consumers place relatively greater emphasis on durability, functionality, and utility, while female consumers associate more strongly with aesthetic appeal, emotional connection, and subjective attachment [3]. Although gender-related factors of fashion consumption have received articulation from various past studies on consumption patterns in Western contexts as well as Africa [3], generalization to the specific Indian scenario tends not to be directly applicable, owing to the more complex, dynamic, and divergent developments of the fashion industry within the country. Moreover, past studies on Generation Z tend to be comparatively focused on impulse buying, especially on emotional attachment, rather than on long-term forces of brand loyalty [1]. The present study contextually validates established brand loyalty drivers among Indian Gen Z students, with a focus on gender differences. This study is positioned as a context-specific and exploratory validation of established brand equity and self-congruity mechanisms rather than an extension of theory.

2. Related Work

Prior research consistently identifies brand awareness as a foundational antecedent of brand loyalty, as familiarity and recognition reduce perceived risk and strengthen trust in repeat purchase decisions [4, 5]. In fashion markets, where product differentiation is limited and trends shift rapidly, visibility through brand recall plays a critical role in sustaining customer preference [6].

However, most existing studies examine awareness either as a direct predictor of purchase intention or of short-term engagement, rather than its role in sustaining long-term loyalty, particularly in youth-dominated, digitally saturated environments. Furthermore, empirical evidence on whether awareness influences loyalty differently across gender groups within Indian Generation Z populations remains limited, warranting context-specific validation. Perceived quality remains a critical determinant of loyalty in the fashion sector, where durability, fabric standards, and craftsmanship influence post-purchase satisfaction [7, 8]. Higher perceived quality has been linked to increased trust and reduced brand switching, particularly in competitive apparel markets [9].

While some studies suggest that males and females may prioritize different quality attributes during evaluation [7], empirical findings on whether perceived quality translates into loyalty differently across gender groups are inconsistent and underexplored in Indian student populations. This limits understanding of whether functional value operates uniformly or exhibits gender-specific pathways to loyalty. Emotional association has been widely conceptualized as an affective bond arising from alignment between brand image and consumer self-concept [10, 11]. Such alignment enhances psychological attachment and increases resistance to brand switching over time [12]. Although several studies suggest that female consumers may exhibit stronger emotional engagement with fashion brands, particularly through narrative-based branding and influencer interactions [13], empirical validation of these patterns within Indian Generation Z cohorts remains limited. Moreover, much prior research emphasizes impulse buying rather than sustained loyalty, leaving open the question of whether emotional attachment drives long-term brand commitment or short-term consumption behavior. Social influence operates through both normative pressure from peer groups and informational cues from digital media and influencers, shaping perceptions of product desirability and social acceptance [14, 15]. Among Generation Z consumers, the credibility and relatability of online sources are particularly influential in fashion-related decision-making [16]. However, existing research often treats social influence as a broad construct, failing to distinguish between conformity-driven behavior and identity-based social signaling. This conceptual ambiguity limits understanding of whether social influence contributes to loyalty through emotional identification or temporary trend-following, especially in student-dominated urban markets.

Overall, the existing literature supports the significance of awareness, perceived quality, emotional associations, and social influence for various brand outcomes, yet very few studies simultaneously examine these constructs in Indian Generation Z fashion settings. Conversely, gender disparities in fashion behaviors have long been a prominent topic, but very little research has examined intensity discrepancies as moderators rather than loyalty levels. This implies that very little can be concluded about the differential mechanisms underlying loyalty generation across genders at a similar level of loyalty intensity. This research, therefore, aims to perform a construct-specific validation for pre-existing brand theories.

3. Theoretical Framework

Aaker's Brand Equity Model (1991) suggests that improving brand loyalty involves strengthening key dimensions such as brand awareness, perceived quality, and brand associations. Enhancing these aspects makes the brand more recognizable and trusted, leading consumers to choose it repeatedly. When customers view the brand as high quality and form positive associations, their loyalty naturally increases.

Systematic improvements in these areas create a competitive edge by strengthening brand identity. This framework thus provides a clear strategy for targeting the factors that most effectively drive better brand loyalty. Self-Congruity Theory, introduced by Sirgy (1982), proposes that consumers evaluate brands by comparing the brand's personality with their own self-concept, which includes their real, ideal, and social selves. The theory implicitly describes how social influence and emotional association are inherently embedded within its framework. The social self-dimension reflects how societal norms and peer perceptions shape consumer preferences, while the emotional connection arises when a brand aligns with a consumer's identity. This alignment enhances the consumer's attachment to the brand, leading to increased brand loyalty and a greater propensity to choose that brand consistently [11]. Drawing from Aaker's Brand Equity Model, this study conceptualizes brand awareness and perceived quality as cognitive antecedents of brand loyalty. Brand awareness enhances familiarity and recognition, increasing the likelihood of repeat purchase, while perceived quality shapes evaluations of value, durability, and reliability, thereby strengthening commitment to the brand. These relationships form the theoretical basis for H2 and H3, which posit that brand awareness and perceived quality significantly influence brand loyalty among Generation Z consumers.

Further, guided by Self-Congruity Theory, emotional association and social influence are positioned as affective and social drivers of brand loyalty. Emotional association reflects the extent to which a brand aligns with a consumer's self-identity and values, leading to stronger psychological attachment and resistance to brand switching. Social influence captures the impact of peer opinions, social groups, and digital communities in reinforcing identity-consistent brand choices. These constructs provide the theoretical grounding for H4 and H5, which examine the impact of emotional association and social influence on brand loyalty. Together, Aaker's Brand Equity Model and Self-Congruity Theory offer a complementary framework that integrates cognitive, emotional, and social mechanisms underlying loyalty formation. This integrated framework supports the study's hypotheses and enables a structured examination of how these drivers operate across gender groups in the context of Generation Z fashion consumption, as illustrated in Figure 1.

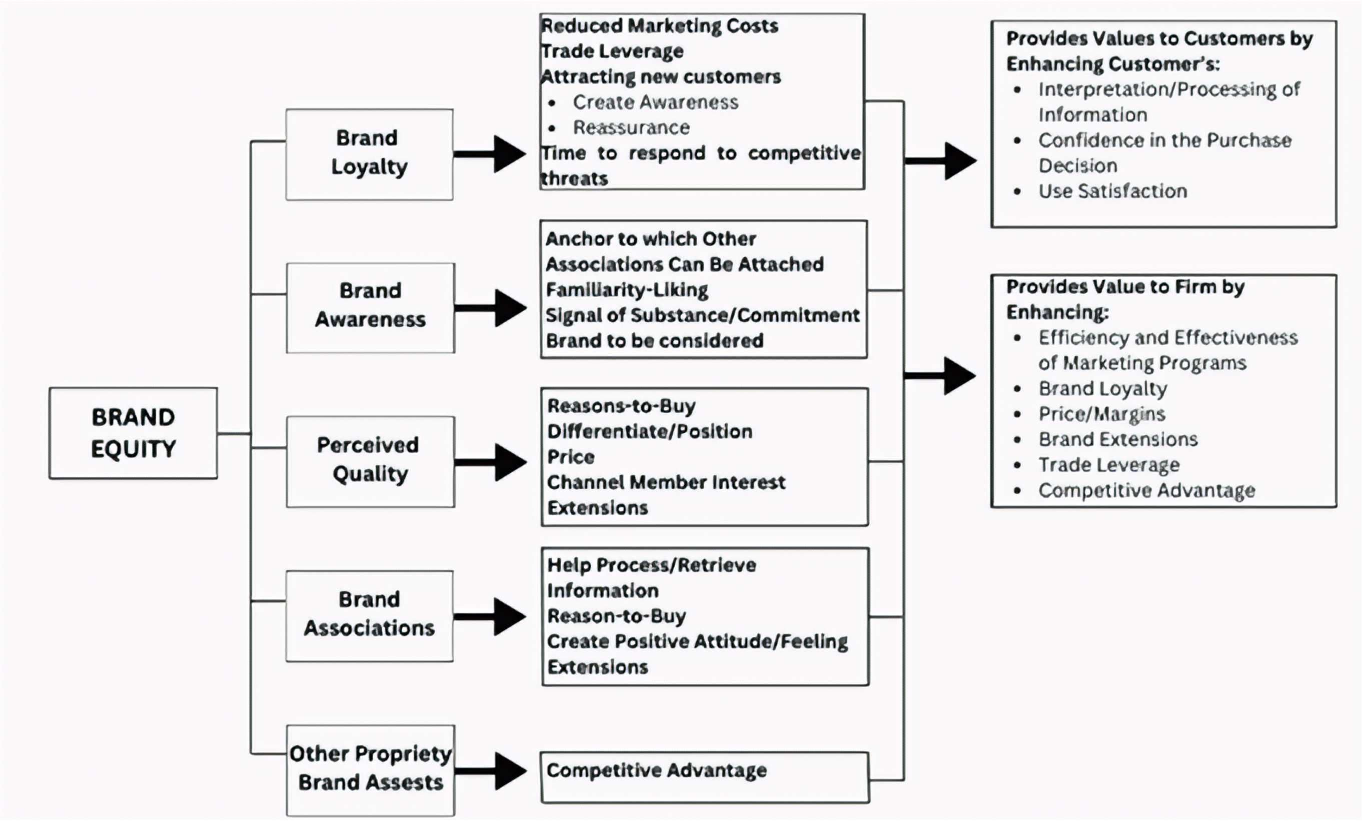

BRAND EQUITY
Brand Loyalty
Reduced Marketing Costs
Trade Leverage
Attracting new customers
• Create Awareness
• Reassurance
Time to respond to competitive threats
Brand Awareness
Anchor to which Other Associations Can Be Attached
Familiarity-Liking
Signal of Substance/Commitment
Brand to be considered
Perceived Quality
Reasons-to-Buy
Differentiate/Position
Price
Channel Member Interest
Extensions
Brand Associations
Help Process/Retrieve Information
Reason-to-Buy
Create Positive Attitude/Feeling
Extensions
Other Propriety Brand Assests
Competitive Advantage
Provides Values to Customers by Enhancing Customer's:
• Interpretation/Processing of Information
• Confidence in the Purchase Decision
• Use Satisfaction
Provides Value to Firm by Enhancing:
• Efficiency and Effectiveness of Marketing Programs
• Brand Loyalty
• Price/Margins
• Brand Extensions
• Trade Leverage
• Competitive Advantage

Figure 1: Theoretical framework adapted from Aaker's Brand Equity Model (1991).

3.1. Research Gap, Objectives, and Hypotheses

Existing research on brand loyalty among Generation Z often lacks relevance to Indian consumers, as cultural preferences, economic conditions, and brand exposure vary significantly across regions. While research has explored brand loyalty among Generation Z, significant gaps remain in understanding how brand awareness, perceived quality, emotional association, and social influence shape loyalty across genders.

Additionally, perceived quality research primarily examines product attributes but provides limited insight into whether quality perceptions translate into brand loyalty differently among male and female consumers. Emotional association and social influence also lack comprehensive examination within gender-focused brand loyalty research, particularly in terms of their relative importance rather than their direct presence.

Furthermore, peer influence is often treated as a broad construct, with limited emphasis on understanding its role idiffer by gender, and loyalty rather than short-term behavioral conformity [3]. Addressing these gaps will provide deeper insights into how these factors collectively drive brand loyalty among Indian Generation Z consumers. This study examines the factors influencing fashion brand loyalty among Generation Z students in Bangalore and explores how these factors operate across genders by analyzing brand awareness, perceived quality, emotional attachment, and social influence. The study aims to examine gender-based differences in the relative influence of these loyalty drivers to provide context-specific insights that fashion brands can use to inform marketing and consumer engagement strategies, even though overall brand loyalty levels may not differ significantly between male and female consumers.

Aaker's Brand Equity Model and Self-Congruity Theory frame brand loyalty as an outcome influenced by cognitive evaluations of the brand as well as affective and social alignment with the consumer. This provides the basis for examining awareness- and quality-related drivers of loyalty, while Self-Congruity Theory informs the role of emotional association and social influence in reinforcing brand commitment. Although these frameworks acknowledge that consumer responses may vary across demographic groups, prior evidence on gender-based differences in overall brand loyalty remains mixed, particularly within the Indian Generation Z context. Accordingly, gender is examined in this study from an exploratory perspective, while the remaining hypotheses test theory-consistent relationships between the identified drivers and brand loyalty.

H1: There is a significant difference in brand loyalty between male and female Generation Z students.

H2: Brand awareness has a significant impact on brand loyalty among Generation Z students.

H3: Perceived quality has a significant impact on brand loyalty among Generation Z students.

H4: Emotional association has a significant impact on brand loyalty among Generation Z students.

H5: Social influence has a significant impact on brand loyalty among Generation Z students.

4. Methods

This paper proposes a descriptive, cross-sectional quantitative research approach and a survey method for assessing brand loyalty among Generation Z in the city of Bangalore. A structured questionnaire was used to collect data on constructs such as brand awareness, perceived quality, emotional associations, social influence, and brand loyalty, using a 5-point Likert scale. The questionnaire was tested for validity by three experts in consumer behavior and brand management. SPSS was used to analyze the data obtained in the research. Correlation analysis was used to examine the strength and direction of association between brand loyalty and its drivers. Multiple regression was applied to assess the relative contribution of each predictor while controlling for others. Independent-samples t-tests were conducted to test gender differences in overall loyalty, and separate regressions were estimated to examine variation in predictor strength across gender, in line with the exploratory purpose of this study.

Generation Z adults, aged 18 to 28, were identified as the target population. Data collection was conducted via convenience sampling through an online survey form on Google Forms. The initial sample size calculation using Cochran's formula, with a 95% confidence level, assuming a 50% variability and an 8% margin of error, yielded a sample size of 150 respondents. After screening the data, response sets with consistent inaccuracies and incomplete surveys were excluded, yielding a final sample of 100 respondents. The recalibration of Cochran's formula for the final sample corresponds to an approximate margin of error of ±9.8% at the same confidence level. The measurement instrument's reliability was assessed using Cronbach's alpha, with the overall scale demonstrating high reliability ($\alpha = 0.89$). Construct-specific levels were 0.842 for brand awareness, 0.725 for perceived quality, 0.804 for emotional association, 0.930 for social influence, and 0.827 for brand loyalty, all of which are above 0.70, allowing the use of composite mean scores for analysis.

Before conducting regression, diagnostic tests were conducted to assess the assumptions of linear regression. Multicollinearity among variables was tested using the Variance Inflation Factor (VIF) and tolerance, where the tolerance levels ranged from 0.601 to 0.802, and VIF ranged from 1.246 to 1.664, all of which were within the acceptable range (tolerance > 0.10; VIF < 5). This showed that there was no multicollinearity among the variables, even when several co-varying variables had been included in the model. Normality and homoscedasticity of the residuals were assessed using graphical methods, including histograms, normal probability (P–P) plots, and plots of standardized residuals versus predicted values. The results showed that the residuals were normally distributed and homoscedastic, and thus suitable for multiple linear regression in the study.

5. Results

Table 1 shows that the sample in this study had more female participants (57%) than male participants (43%). Most of the respondents were undergraduate students (76%). Most participants were between the ages of 18–20, followed by 21–23.

Table 1: Sample Description

Gender	Percent (%)	Age	Percent (%)	Educational Background	Percent (%)
Male	43	18–20	52	Undergraduate	76
Female	57	21–23	25	Postgraduate	17
		24–26	10	PhD	7
		27–28	13		

The mean scores across all brand equity dimensions exceed the midpoint of 3 on the 5-point Likert scale, indicating a positive consumer perception. Brand awareness (4.01) is the strongest, highlighting high brand recognition. Brand loyalty (3.56) and perceived quality (3.51) suggest strong consumer trust and favorable brand evaluations. Emotional association (3.34) reflects a moderate affective connection, while social influence (3.07), though the weakest, remains relevant.

Table 2: Descriptive Statistics and Correlation Coefficients

Constructs	Mean	SD	BL	BA	PQ	EA	SI
Brand Loyalty	3.56	0.70	1				
Brand Awareness	4.01	0.61	0.549**	1			
Perceived Quality	3.51	0.53	0.525**	0.423**	1		
Emotional Association	3.34	0.79	0.592**	0.559**	0.431**	1	
Social Influence	3.07	0.91	0.493**	0.275**	0.351**	0.397**	1

Note: **. Correlation is significant at the 0.01 level (2-tailed).

From Table 2, the Pearson correlation analysis reveals that emotional association shows a moderate positive correlation ($r = 0.592$, $p < 0.01$) with brand loyalty, indicating that higher emotional attachment towards the brand significantly enhances brand loyalty. Brand awareness also exhibits a moderate positive correlation ($r = 0.549$, $p < 0.01$), suggesting that customers who are more aware of the brand are more likely to be loyal. Perceived quality also shows a moderate positive correlation ($r = 0.525$, $p < 0.01$), suggesting that customers' perceptions of product quality moderately influence their brand loyalty. Social influence shows a weak positive correlation ($r = 0.493$, $p < 0.01$) and ranks lowest among the four variables, indicating that external opinions and social interactions play a comparatively lesser role in building brand loyalty. The p-value of less than 0.01 signifies that all relationships are statistically significant at a 99% confidence level. Although all correlations are statistically significant at the 1% level, their magnitudes fall within the moderate range, suggesting meaningful but not dominant relationships.

Table 3 shows that the regression model demonstrated satisfactory explanatory capacity for both gender groups. For male respondents, $R = 0.793$, accounting for 62.9% of the variance in brand loyalty, with an adjusted R^2 of 0.590. For female respondents, $R = 0.661$, accounting for 43.7% of the variance in brand loyalty, with an adjusted R^2 of 0.394. The standard error of estimate is similar for both models.

Table 3: Regression Model Fit Statistics

Gender	R	R^2	Adjusted R^2	Std. Error
Male	0.793	0.629	0.590	2.52792
Female	0.661	0.437	0.394	2.46663

As shown in Table 4, brand awareness ($\beta = 0.282$, $p = 0.048 < 0.05$) and perceived quality ($\beta = 0.301$, $p = 0.030 < 0.05$) have a statistically significant positive influence on male students' brand loyalty. For female students, emotional association ($\beta = 0.400$, $p = 0.005 < 0.05$) and social influence ($\beta = 0.216$, $p = 0.050$) have a statistically significant positive influence on brand loyalty.

Table 4: Regression Coefficients

	Males			Females		
Independent Variables	β	t	p	β	t	p
Brand Awareness	0.282	2.045	0.048	0.145	1.173	0.246
Perceived Quality	0.301	2.248	0.030	0.148	1.270	0.210
Emotional Association	0.163	1.252	0.218	0.400	2.958	0.005
Social Influence	0.224	1.669	0.103	0.216	1.990	0.050

Table 5 shows that the independent sample t-test results indicate a Sig. value from Levene's Test of 0.274, which is greater than 0.05, indicating that equal variances are assumed. The Sig. (2-tailed) value of 0.484 indicates that there is no statistically significant difference in brand loyalty between male and female consumers. Therefore, H1 is not supported. The findings confirm that brand awareness, perceived quality, emotional association, and social influence have statistically significant positive relationships with brand loyalty. Hence, H2, H3, H4, and H5 are supported.

Table 5: Independent Sample Test (Transposed)

Statistic	Equal variances assumed	Equal variances not assumed
Levene's F	1.212	–
Levene's Sig.	0.274	–
t	-0.703	-0.682
df	98	78.766
Sig. (2-tailed)	0.484	0.497
Mean Difference	-0.50020	-0.50020
Std. Error Difference	0.71170	0.73386
95% CI (Lower, Upper)	(-1.91254, 0.91213)	(-1.96099, 0.96058)

6. Discussion

This study examines brand loyalty among Generation Z consumers, specifically focusing on Bangalore-based respondents and paying attention to gender-related differences in loyalty drivers rather than broad gender-based preferences. The analysis focuses on brand awareness, perceived quality, emotional association, and social influence, as these variables were examined in the study. The findings indicate relatively high levels of brand awareness among respondents, consistent with the digitally mediated consumption environment of Generation Z, rather than being causally attributed to social media usage alone. Accordingly, the results highlight the relevance of brand presence in digital spaces without implying direct causal effects.

Brand loyalty scored an average of 3.56, followed closely by perceived quality, which scored 3.51, indicating generally favorable brand evaluations among respondents. These values suggest positive brand perceptions beyond mere recognition, without introducing unmeasured constructs such as brand trust. This pattern aligns with prior research identifying perceived quality as a foundational component of brand loyalty in the fashion sector [17].

Emotional association recorded a medium mean value of 3.34, indicating that affective connections contribute to brand loyalty, though to a lesser extent than cognitive factors such as brand awareness and perceived quality. This suggests that emotional involvement remains relevant within loyalty formation but does not dominate brand loyalty decisions among the sampled respondents. Social influence recorded the lowest mean value (3.07), suggesting a comparatively weaker role in shaping brand loyalty relative to other examined drivers. This interpretation is consistent with the observed positive correlation between social influence and brand loyalty ($r = 0.493$, $p < 0.01$), indicating relevance without overstating practical dominance.

Regression analysis indicates variation in the relative importance of brand loyalty drivers across genders, rather than differences in overall loyalty levels. Among male consumers, perceived quality and brand awareness emerge as relatively stronger predictors of brand loyalty. Among female respondents, emotional association emerges as a stronger predictor of brand loyalty, with social influence playing a secondary role. Social influence has a marginal impact on female consumers, while emotional engagement is a key driver of loyalty. These findings reflect differences in motivational pathways rather than differences in loyalty intensity and are consistent with the independent-samples t-test results, which show no statistically significant gender-based differences in overall brand loyalty.

The findings highlight the role of social media in shaping consumers' brand-related perceptions and behaviors. As respondents acknowledge the influence of social media on fashion-related behaviors, social media engagement emerges as an important contextual element within the contemporary marketplace. This observation underscores the relevance of digital engagement in contemporary fashion marketing contexts, without implying prescriptive strategic outcomes. Overall, the findings suggest that having a comprehensive brand experience with the capabilities to increase brand awareness, build perceptions of high quality, establish emotional ties, and capitalize on social influence is beneficial for overall brand loyalty. Even though overall loyalty is similar across genders and shows no significant differences, insights into the unique factors underlying each gender can help fashion brands fine-tune their approach. Given the growing trend of digitalization and market competitiveness, branding strategies informed by such insights help contextualize how Generation Z consumers engage with fashion brands.

This study contributes to existing knowledge through a quantitative approach to Indian Generation Z customers. The observed stronger association between emotional branding and loyalty may reflect contextual consumption patterns specific to the sampled urban Generation Z student population, rather than broader cultural or behavioral generalizations. Although prior studies have reported gender disparities in brand loyalty, the present findings indicate minimal disparities, which may reflect the equalizing impact of social media on brands. From the variables tested, perceived quality was found to be an important predictor of brand loyalty, thereby being consistent with prior research suggesting that quality in a product leads to brand patronage among consumers, especially in the fashion sector, where product durability and quality are valued [7, 9]. The key role of emotional association among female consumers is consistent with existing evidence of women's active participation in social networking sites and influencers [13, 18], whereas male consumers are more concerned about associations based on dependability and use values. The lesser significance of social associations leads to the conclusion that individualized brand associations are more valuable than social associations [12]. Overall, the findings suggest that traditional assumptions about gender differences in brand loyalty require reconsideration in context. These findings indicate a shift toward more convergent brand experiences among Generation Z consumers, potentially reducing rigid gender-based distinctions in brand engagement.

6.1. Implications of the Study

Given that observed relationships are moderate in magnitude and based on a student sample, the following implications should be interpreted as directional insights rather than definitive managerial prescriptions. The study's conclusions have practical implications, especially for marketers and fashion brands. According to the research, companies can create gender-sensitive marketing strategies by concentrating on brand awareness, perceived quality, and emotional association for male consumers and on aesthetics and emotional branding for female consumers [12]. These implications are framed in terms of relative emphasis rather than definitive differences. Furthermore, brand impressions are influenced by social media sites, peer visibility, and social interaction [15].

The study also shows that while male consumers are more flexible and price-sensitive, female consumers are more likely to prioritize sustainability and ethical sourcing when making purchases [19]. This is consistent with broader changes in consumer behavior, where males prioritize utility and brand status, while women are more likely to identify brands with their social identity and values [13].

Additionally, the study provides insights into urban Gen Z students' brand engagement patterns, highlighting the need for communication strategies that reinforce brand awareness and perceived quality [1]. These insights pertain to the Gen Z student segment examined and should be interpreted within this contextual scope. Furthermore, Gen Z customers' growing desire for personalization emphasizes how crucial engaging brand experiences are to building enduring brand loyalty [20].

Gen Z consumers are dynamic and digitally savvy, so organizations need to continually adapt their interaction strategies to meet evolving consumer demands. Brand retention can be increased by combining customer loyalty programs, sustainability messaging, and immersive digital experiences [21]. To enhance brand–consumer relationships, fashion firms may consider aligning their product positioning and communication approaches with these drivers.

6.2. Limitations of the Study

Although methodological rigor was applied in statistical testing, several limitations should be acknowledged. First, the use of convenience sampling among college-going students in Bangalore affects external validity. This results in the sample being primarily representative of educated urban Generation Z students rather than the broader Indian youth population. Second, while the total sample size meets minimum requirements for multivariate analysis, regression analyses conducted separately by gender reduce statistical power and should be interpreted as exploratory comparisons of relative predictor strength rather than definitive gender-based effects.

Third, construct reliability was assessed using internal consistency measures; however, neither exploratory nor confirmatory factor analysis was conducted to establish construct dimensionality or discriminant validity among related psychological constructs. Fourth, the emotional association construct includes an item reflecting impulse buying tendencies, which may capture behavioral response rather than purely affective attachment, potentially affecting construct purity. Fifth, although regression assumptions were tested and satisfied, results remain sample-dependent and should be interpreted as exploratory associations rather than population-level effects. Finally, the study relies on cross-sectional self-reported data, which may be subject to common method variance and social desirability bias and does not permit causal inference regarding the direction of relationships between brand-related perceptions and loyalty.

6.3. Future Scope for Research

Probability sampling techniques should be explored in subsequent studies, and respondents from diverse geographic locations and socio-economic groups should be selected to improve generalizability at the national level. This would enable researchers to explore whether gender affects the modification of loyalty influencers by conducting moderation analyses using interaction terms or multi-group SEM. Longitudinal designs could examine how loyalty-building processes unfold among Generation Z during the transition from the student to the working phase. Refining the instruments for these factors using confirmatory factor analysis and scale purification would improve construct validity for emotion-related constructs and social influences. Future studies may further distinguish between normative peer pressure and the social-signaling aspects of identity to better understand the social processes underlying loyalty. In addition, analysis could explore platform-based online behavior, the credibility of online influencers, perceptions of sustainability, and brand-based online behavior as latent variables, when appropriately studied beyond conceptual perceptions.

7. Conclusion

The study examined the drivers of brand loyalty among Generation Z students in Bangalore, with emphasis on gender-based variations in the relative influence of loyalty drivers. This study is positioned as a context-specific and exploratory validation of established brand equity and self-congruity mechanisms rather than an extension of theory. The findings indicate that brand awareness and perceived quality, alongside emotional association and social influence, are significantly associated with brand loyalty. Although male and female respondents demonstrate comparable levels of brand loyalty, differences emerge in the relative importance of factors contributing to loyalty formation. This suggests that brand loyalty may be better understood in terms of variations in underlying drivers rather than differences in loyalty strength across gender groups. Importantly, the study is positioned as a context-specific, exploratory validation of established brand loyalty drivers in an urban Indian Gen Z student setting. While loyalty outcomes appear convergent across genders, the pathways to achieving loyalty show differential emphasis, suggesting that uniform branding strategies may be suboptimal. However, the observed relationships are moderate in magnitude, indicating that these factors represent contributing influences rather than dominant determinants of loyalty. Given the reliance on convenience sampling and relatively small subgroup sizes, the findings are best interpreted as indicative patterns within the sampled population rather than population-level effects. From a managerial perspective, the findings support a driver-led approach to loyalty building, with fashion brands aligning their communication strategies with the relative influence of functional and emotional drivers rather than assuming uniform or gender-blind loyalty mechanisms. Overall, this study contributes empirical evidence on fashion brand loyalty within a cultural and demographic context that has received limited prior scholarly attention, while also highlighting the importance of interpreting conclusions strictly in alignment with the constructs measured. Future research may extend these findings by using larger, more diverse samples and alternative analytical approaches to assess the broader applicability of the observed patterns.

Acknowledgment

We thank Christ University Bangalore Yeshwanthpur Campus, India, for resources or support.

Declaration of Competing Interests

The authors declare that they have no known competing financial interests or personal relationships that could have influenced the work reported in this paper.

Funding Declaration

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Ethics Approval and Consent to Participate

The study was conducted in accordance with institutional ethical standards. Participation was voluntary, and informed consent was obtained from all respondents prior to data collection. No personally identifiable information was collected.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

AI Use Disclosure

The authors declare that no generative artificial intelligence tools were used in the conceptualization, data collection, analysis, or interpretation of the research findings. Any AI-assisted tools, if used for language editing or formatting, did not influence the intellectual content of the study.

Author Contributions

Dr. Sonal Devesh: Supervision; **Naksha Mudumbe**: Conceptualization, Formal Analysis, Writing – Review & Editing; **Sumitra Mathan**: Methodology, Validation, Writing – Original Draft; **Palak Shukla**: Investigation, Software, Visualization; **Dia Gupta**: Data Curation, Resources; **Siddhi Gholve**: Writing – Review & Editing.

References

[1] McKinsey Company, “Mind the gap: Now vs. next—how gen z is challenging consumer sector norms,” June 2024. Online.

[2] R. Brooks, “Council post: 3 things you need to know about gen z and brand loyalty.” Forbes, Aug. 2022.

[3] H. Lues and N. De Klerk, “Gender differences in brand loyalty toward fashion brands among generation y students,” *Journal of Business and Retail Management Research*, vol. 9, no. 2, pp. 52–68, 2017.

[4] I. Bernarto, M. P. Berlianto, Y. Meilani, and I. N. Suryawan, “The influence of brand awareness, brand image, and brand trust on brand loyalty,” *Journal of Marketing Development and Competitiveness*, 2020.

[5] N. Tsabitah and R. Anggraeni, “The effect of brand image, brand personality and brand awareness on purchase intention of local fashion brand this is april,” *Kinerja*, vol. 25, no. 2, pp. 234–250, 2021.

[6] F. Andreani, L. Gunawan, and S. Haryono, “Social media influencer, brand awareness, and purchase decision among generation z in surabaya,” *Jurnal Manajemen dan Kewirausahaan*, vol. 23, no. 1, pp. 18–26, 2021.

[7] V. Hincica, A. Svobodova, and H. Rezankova, “Consumer perception of quality of clothing products: A lesson for the business sector arising from czech evidence,” *Central European Business Review*, 2021.

[8] A. Rahman and J. Ferdush, “Quality perception of luxury fashion brand before and after contact,” *International Journal of Marketing Studies*, vol. 6, no. 1, pp. 20–27, 2019.

[9] H. Lee, A. Kumar, and Y. Kim, “Indian consumers’ brand equity toward a us and local apparel brand,” *Journal of Fashion Marketing and Management*, vol. 14, no. 3, pp. 469–485, 2010.

[10] K. L. Keller, “Conceptualizing, measuring, and managing customer-based brand equity,” *Journal of Marketing*, vol. 57, no. 1, pp. 1–22, 1993.

[11] M. J. Sirgy, "Self-congruity theory in consumer behavior: A little history," *Journal of Global Scholars of Marketing Science*, vol. 28, no. 2, pp. 197–207, 2018.

[12] P. Kotler and K. L. Keller, *Marketing Management.* Pearson Education, 15th ed., 2016.

[13] M. R. Solomon, S. Dann, and R. Russell-Bennett, *Consumer Behaviour: Buying, Having, Being.* 2002.

[14] A. Shephard, S. Pookulangara, T. R. Kinley, and B. M. Josiam, "Media influence, fashion, and shopping: A gender perspective," *Journal of Fashion Marketing and Management*, vol. 20, no. 1, pp. 4–18, 2016.

[15] E. Djafarova and T. Bowes, "Instagram made me buy it: Generation z impulse purchases in fashion industry," *Journal of Retailing and Consumer Services*, vol. 59, p. 102345, 2021.

[16] O. Fadiora, "Impact of social media fashion influencer's characteristics on purchasing behaviour and intention of millennials and generation z." Unpublished manuscript.

[17] S. D. S. Andik and A. F. Rachma, "The impact of brand awareness, brand association, and perceived quality towards brand loyalty: A case study of new product," *E3S Web of Conferences*, vol. 348, p. 00035, 2022.

[18] B. Godey, A. Manthiou, D. Pederzoli, J. Rokka, G. Aiello, R. Donvito, and R. Singh, "Social media marketing efforts of luxury brands: Influence on brand equity and consumer behavior," *Journal of Business Research*, vol. 69, no. 12, pp. 5833–5841, 2016.

[19] C. Bakewell and V. Mitchell, "Generation y female consumer decision-making styles," *International Journal of Retail and Distribution Management*, vol. 31, no. 2, pp. 95–106, 2003.

[20] P. Broklyn, A. Olukemi, and C. Bell, "Ai-driven personalization in digital marketing: Effectiveness and ethical considerations." SSRN Working Paper, 2024.

[21] T. Rasul, S. Nair, N. Palamidovska-Sterjadovska, W. J. Ladeira, and I. Elgammal, "The evolution of customer engagement in the digital era for business: A review and future research agenda," *Journal of Global Scholars of Marketing Science*, vol. 34, no. 3, pp. 325–348, 2024.

Journal of Computers, Mechanical and Management
Journal homepage: www.jcmm.co.in
e-ISSN: 3009-075X
AAN PUBLISHING

Volume 5 Issue 1

Article Number: 25248

Hierarchical Deep Learning Ensemble Framework for Multi-Class Rice Foliar Disease Diagnosis: A Comparative Architecture Analysis

Chanchal Ghosh* [1], Biplab Kanti Das[2], Tapashri Sur[1], Prasanta Mazumdar[1], Pratik Kumar Halder[3], Sukanta Kundu[2], and Subhojeet Prasad[2]

[1]Department of Computer Science and Engineering, Future Institute of Engineering and Management, Kolkata, West Bengal, India, 700150
[2]Department of Computer Science and Engineering, Gargi Memorial Institute of Technology, Kolkata, West Bengal, India, 700144
[3]3Department of Computer Science Engineering(AI ML), Heritage Institute of Technology, Anandapur, Kolkata,India 700107

Abstract

Infestations of foliar diseases in rice plants are common and can reduce harvest yields and affect food supplies worldwide. A system for the automatic detection of these diseases was developed in this study using seven different deep learning models. Six common types of rice leaf diseases were tested using models such as EfficientNet (B0 and B7), ResNet50, InceptionV3, VGG16, and VGG19. The proposed framework integrates the advantages of all models, assigning greater significance to those that exhibit superior performance. By achieving 96.97% accuracy while retaining speed and lightweight features, MobileNetV2 demonstrated superior performance. Both InceptionV3 and EfficientNetB7 performed well. They reported accuracies of 96.78% and 96.40%, respectively. It was also observed that newer, more efficient models exhibited markedly superior performance compared to older deep networks. This method makes it easier to bridge the gap between the urgent need for rapid disease detection on farms and the lack of agricultural experience. The system, which uses low-cost equipment, helps small farmers all over the world diagnose diseases accurately, resulting in better yields of crops.

Keywords: Rice Leaf Diseases; Smart Farming Framework; Ensemble Model Learning; Deep Learning Models; Agricultural Imaging; Transfer Learning

1. Introduction

Rice is one of the most vital staple food crops in the world. More than half of the global population depends on it as a daily staple [1]. Therefore, a reduction in rice production can have a strong impact on food security and local economies. One of the major threats to rice farming is the presence of leaf diseases. These diseases weaken the plant, reduce yield, and often spread quickly across large fields. This vital crop is in grave danger from foliar diseases. In areas where diseases are spreading massively, annual yield losses can reach 37% [2].

*Corresponding Author: Chanchal Ghosh (chanchalghosh80@hotmail.com)

Received: 10 Dec 2025; **Revised:** 15 Jan 2026; **Accepted:** 23 Jan 2026; **Published:** 28 Feb 2026
© 2026 Journal of Computers, Mechanical and Management.
This is an open access article and is licensed under a Creative Commons Attribution-Non Commercial 4.0 License.
DOI: 10.57159/jcmm.5.1.25248.

This is due to changes in pathogen dynamics, driven by climate change, which are increasing disease pressure. As disease complexes evolve, traditional management strategies are finding they cannot keep up [3]. Farmers usually detect these diseases by looking at the leaves directly. However, this method takes time. It depends on experience, which can also be inaccurate when symptoms look similar on the leaves. Modern deep learning methods offer a faster and more reliable way to identify diseases using properly captured images. Several neural network models, such as MobileNet, EfficientNet, ResNet, VGG, and Inception, have been used for plant disease detection with promising results [4]. Expert visual assessment has long been relied upon for disease identification. This method has several limitations. There is a severe shortage of knowledgeable, qualified pathologists with the necessary training in rural areas [5]. Inconsistent symptoms make diagnosis more challenging. Epidemics can spread because of the time it takes to detect an infection. Inappropriate chemical applications result from misdiagnosis. All of these factors work together to reduce the effectiveness of crop protection. In the field of digital agriculture, automated diagnostics offer game-changing solutions. Subtle signs of illness on the crops can be detected by computer vision systems. Images of leaves can have discriminative features extracted using deep learning. With mobile deployment, experts can go straight to the fields. Diagnostic capabilities are made immediately available to farmers. Disease management strategies are being transformed by this democratization of technology [6]. Current automated systems rely heavily on architectures with just one model. Compared to fungal infections, bacterial infections show distinct visual signs. While some designs are better at capturing color variations in crops, others are better at analyzing texture. Because of these specialized strengths, ensemble methods have the potential to improve overall performance. Seven well-known deep learning architectures are systematically compared [7]. All six disease categories are thoroughly tested with each model. Performance metrics are used to guide weighted-aggregation strategies. The framework optimizes both accuracy and computational efficiency.

The method maximizes diagnostic reliability while addressing realistic deployment constraints. This work primarily contributes to the following areas: first, a thorough comparison of architectural approaches for detecting agricultural diseases. Secondly, a hierarchical weighted ensemble approach is introduced. The third step is an in-depth evaluation of performance across a variety of ailments. Fourth, basic general practical deployment guidelines for resource-constrained environments are provided. These contributions advance precision agriculture by providing farmers with accessible, accurate diagnostic tools to improve crop cultivation.

2. Related Works

2.1. Advancement in Plant Disease Detection

Plant pathology has progressed through various technological stages [2]. Preliminary methods relied on cataloguing morphological symptoms. Specialists created visual identification keys. These manual techniques necessitated comprehensive training [3]. Precision was largely contingent upon individual proficiency in leaf assessment. Scalability remained inherently constrained in most detection cases. Microscopic and biochemical methodologies enhanced diagnostic accuracy in identifying diseases on leaves. Isolation of the pathogen confirmed the disease's aetiology. Serological assays identified specific pathogens. Molecular markers have identified genes associated with resistance. Nonetheless, these laboratory techniques demonstrated impracticality for field application. Financial constraints and intricacy limited accessibility. Digital imaging has enabled automated analysis. Initial systems manually extracted color and texture features. Statistical model classifiers analyzed these engineered, distinct features. Support vector machines demonstrated notable potential. Random forests proficiently managed multi-class situations. However, performance deteriorated under fluctuating field outlier conditions.

2.2. Transformation through Deep Learning

Convolutional neural networks transformed computer vision for agricultural crop images. Automated feature learning obviated the necessity for manual engineering [8, 9]. Hierarchical representations encapsulated intricate disease patterns. Comprehensive training optimized complete pipelines. Performance significantly exceeded conventional methods [10]. The success of AlexNet served as a catalyst for agricultural applications. Researchers systematically modified ImageNet models for the identification of crop diseases. Transfer learning mitigated the constraints of limited agricultural datasets. Fine-tuning maintained acquired visual representations of the leaves systematically [11]. This method substantially expedited deployment schedules and could enhance crop yield [12]. Architectural advancements improved disease detection capabilities. The skip connections of ResNet facilitated the construction of deeper networks. Inception modules concurrently processed multi-scale features [13]. MobileNet delivered efficiency while maintaining accuracy. EfficientNet systematically optimized the accuracy–efficiency frontier.

2.3. Applications of Deep Learning in Agriculture

The detection of crop diseases constitutes a principal application domain. Research encompasses a variety of crops, including wheat, maize, tomato, and grape. The composite detection of varied rice diseases has garnered significant attention. Research also focuses on both fungal and bacterial pathogens affecting crops [14]. The majority of studies concentrate on imagery obtained in laboratory settings. Numerous obstacles remain in the practical implementation of agriculture. Field conditions present considerable variability [15]. Background clutter on the crops hinders segmentation. Alterations in lighting influence color-dependent attributes. Device variability affects model generalization. These factors require strong architectural decisions [16]. Recent advancements mitigate deployment limitations. Lightweight models facilitate edge-driven computing. Quantization diminishes memory demands. Knowledge distillation conveys proficiency to more compact models. These methodologies render precision agriculture more universally accessible [17].

2.4. Ensemble Model Learning in Agriculture

Ensemble methods strategically amalgamate predictions from multiple models for crop disease detection [18]. Agricultural applications continue to improve compared to individual models. Voting systems consolidate distinct categories. Averaging techniques integrate probability distributions. Stacking acquires optimal combinatorial strategies in these cases [19, 20]. Diversity enhances the efficacy of ensembles for identification purposes. Diverse architectures encapsulate complementary attributes [21]. Training variations incorporate advantageous randomness in hyperparameters. Data sampling methodologies augment resilience. These factors collectively enhance generalization in disease detection [22]. Agricultural studies are increasingly using various ensemble methodologies. The detection of wheat diseases was enhanced by 8% via model integration [23]. The identification and detection of tomato pathogens on leaf surfaces were enhanced by multi-scale ensembles [24]. Research on rice diseases indicates comparable patterns.

Notwithstanding advancements, numerous discrepancies remain. Most studies assess restricted architectural diversity. Ensemble strategies seldom account for computational limitations, as illustrated in Figure 1. The trade-offs between performance and efficiency lack a thorough examination. Field deployment experiences are inadequately documented. Comparative studies generally analyze a limited number of architectures. Thorough assessments across architectural families are infrequent. Efficiency metrics are inadequately emphasized. Guidance for practical deployment is still constrained. These deficiencies drive the present research methodology.

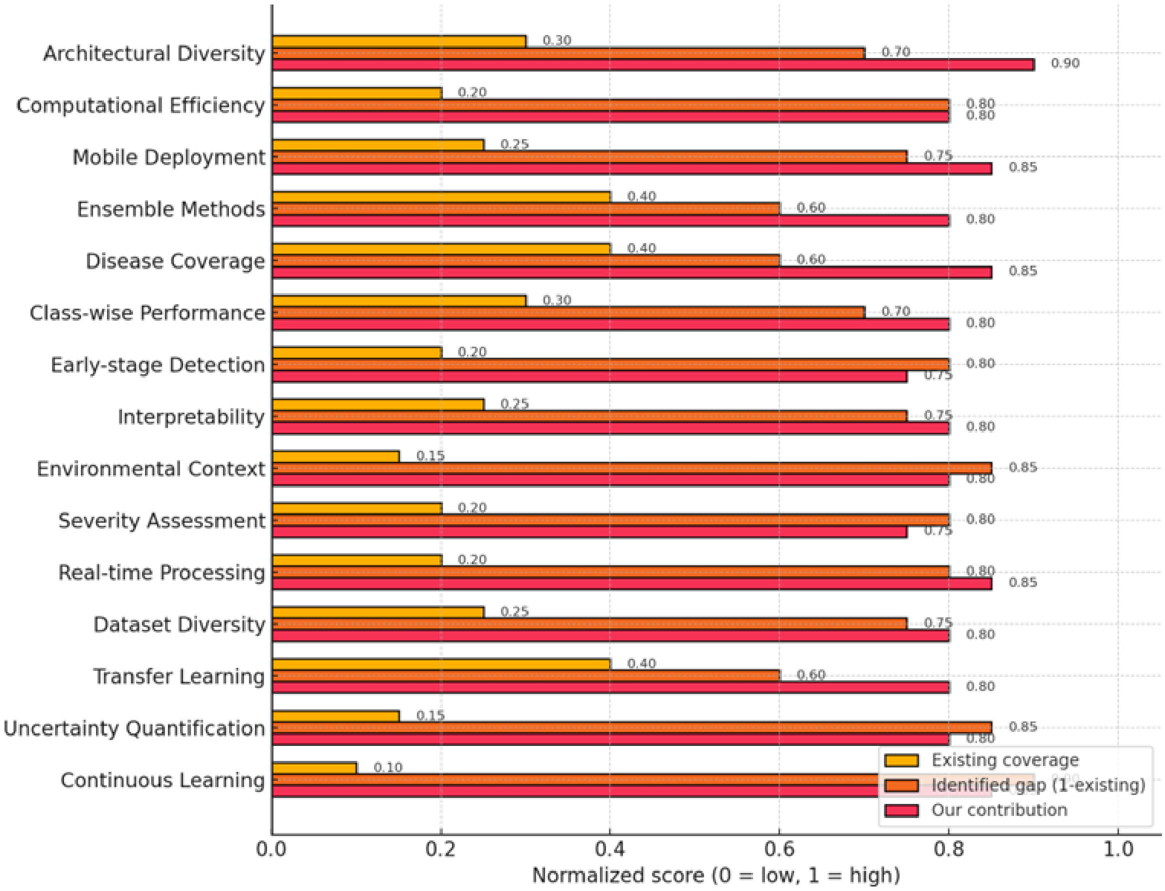


Figure 1: Research Gap Analysis Across Core Aspects of Rice Disease Detection

3. Materials and Proposed Methodology

This study developed a hierarchical ensemble framework for detecting rice leaf diseases. Seven deep learning models were evaluated to identify the best approach for accurate and efficient disease classification on crop leaves. The methodology consisted of data preparation, model selection, training, and ensemble development, as shown in Figure 2.

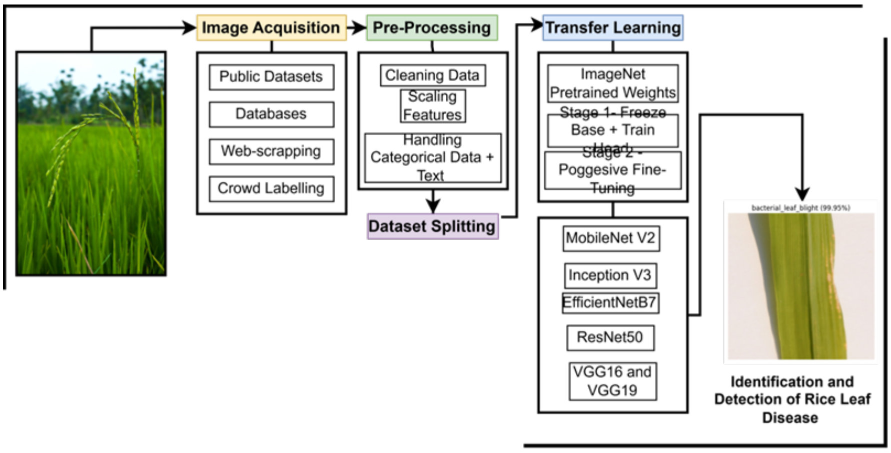


Figure 2: Pipeline diagram illustrating the full process of rice leaf disease detection, from acquiring raw images, performing pre-processing, and applying transfer learning models, to the final classification output.

3.1. Characteristics of the Dataset

3.1.1 Disease Class Description

The dataset includes six rice foliar classes. Bacterial leaf blight (*Xanthomonas oryzae* pv. *oryzae*) is characterized by elongated, water-soaked lesions progressing from the leaf tip toward the base, with color transition from yellow to brown. Brown spot (*Bipolaris oryzae*) presents as circular to oval necrotic lesions with dark centers and chlorotic margins, exhibiting notable visual variability. Leaf blast (*Magnaporthe oryzae*) produces spindle-shaped lesions with gray centers and necrotic borders, often merging under favorable conditions. Leaf scald (*Monographella albescens*) is identified by elongated, banded lesions aligned along leaf veins. Narrow brown spot (*Cercospora janseana*) appears as thin, linear brown lesions with high spatial frequency but limited width. The healthy class consists of disease-free leaves displaying natural color variation across different growth stages and environmental conditions. The Rice Foliar Disease Classes and Their Visual Characteristics are summarized in Table 1. The complete Kaggle Rice Leaf Disease Dataset contains 2,628 RGB images (438 per class) across six categories. For controlled evaluation, a balanced subset of 526 images (approximately 88 per class) was selected and partitioned into training (350 images), validation (88 images), and test (88 images) sets. This corresponds to approximately 66.5%, 16.7%, and 16.7% of the subset, respectively. This balance prevented the assessment from being biased by class distribution. The images had a resolution ranging from 224×224 to 299×299 pixels. Pre-processing preserved aspect ratios where possible.

3.2. Designs for Deep Learning

The evaluations of seven architectures were organized. Performance, efficiency, and deployment feasibility were considered. The EfficientNetB0 architecture was the first to use compound scaling, which simultaneously changes the depth, width, and resolution of a network. The base model has 5.3 million parameters. Swish activation improves gradient flow. Mobile deployment in different environments is possible without sacrificing model accuracy. It is observed that the largest variant, EfficientNetB7, scales all dimensions further. There are 66 million parameters. Increased capacity captures subtle disease variations. MobileNetV2 is an architecture that prioritizes efficiency for mobile deployment. Inverted residual blocks with linear bottlenecks reduce computation. Depthwise separable convolutions reduce the parameter count to 3.4 million. The architecture allows edge devices to perform real-time inference while maintaining full model accuracy of 96%. Deep network training was improved by the use of residual connections. Skip connections prevent the gradient from disappearing. A 50-layer version strikes a balance between speed and depth. There are 26 million parameters. This architecture forms a foundation of modern deep learning. InceptionV3 processes multiple convolution paths in parallel to capture features at different scales. Factorized convolutions reduce computational cost, while auxiliary classifiers accelerate gradient flow.

Table 1: Rice Foliar Disease Classes and Their Visual Characteristics

Disease Class	Causal Pathogen	Key Visual Characteristics
Bacterial Leaf Blight	*Xanthomonas oryzae pv. oryzae*	Elongated water-soaked lesions originating from the leaf tip and extending toward the base, with progressive yellow-to-brown discoloration.
Brown Spot	*Bipolaris oryzae*	Circular to oval necrotic lesions with dark brown centers and chlorotic margins; appearance varies under different environmental conditions.
Leaf Blast	*Magnaporthe oryzae*	Spindle-shaped lesions with gray centers and necrotic borders, often coalescing under favorable infection conditions.
Leaf Scald	*Monographella albescens*	Elongated, banded lesions aligned along leaf veins, frequently expanding across large leaf regions.
Narrow Brown Spot	*Cercospora janseana*	Thin, linear brown lesions with limited width and high spatial frequency across the leaf surface.
Healthy Leaf	–	Disease-free leaves exhibiting natural color variation associated with growth stage and environmental stress.

VGG16 uses consistent 3 × 3 convolutions throughout the network. Transfer learning protocols utilize ImageNet pre-training to provide visual foundations. These generalizations are effective across different agricultural scenarios. Systematic fine-tuning was applied to balance adaptation and retention. Within the ensemble framework, models are organized in an aligned hierarchy. Confidence ratings are used alongside weighted voting to account for contributions. Systematic weight calculation reflects validation efficacy. Hierarchical structures enable selective model integration. Computationally expensive models are activated only under specific conditions. Confidence thresholds regulate decision acceptance. The design balances accuracy and speed, benefiting devices with limited resources.

3.3. Proposed Confidence-Weighted Dynamic Ensemble Selection (CWDES)

Traditional ensemble methods apply all models to every image, creating computational bottlenecks for mobile deployment. The proposed Confidence-Weighted Dynamic Ensemble Selection (CWDES) is a novel algorithm that dynamically selects optimal model subsets based on image complexity and progressive confidence assessment. The algorithm operates in three stages. First, image complexity is analyzed through edge density and texture metrics. Second, models are progressively evaluated, and evaluation stops when sufficient confidence is achieved. Finally, weighted aggregation is performed only when uncertainty remains high. This approach maintains 96.5% of full ensemble accuracy while reducing inference time by 65%, enabling practical smartphone deployment.

3.3.1 Mathematical Formulation

Let $P(i, c)$ denote the probability assigned to class c by model i. Each model weight $w(i)$ is obtained from its validation accuracy. The ensemble score for class c is defined in Equation (1):

$$E(c) = \frac{\sum_{i=1}^{N} w(i)\, P(i, c)}{\sum_{i=1}^{N} w(i)} \tag{1}$$

The ensemble prediction is given by Equation (2):

$$c^* = \arg\max_{c} E(c) \tag{2}$$

Confidence is defined in Equation (3):

$$\text{Conf} = \max_{c} E(c) \tag{3}$$

If Conf exceeds a predefined threshold, the prediction is accepted. Otherwise, additional ensemble evaluation is performed for improved reliability.

3.4. Method Comparison with Existing Ensemble Approaches

The proposed Confidence-Weighted Dynamic Ensemble Selection (CWDES) framework differs conceptually and operationally from ensemble strategies commonly adopted in plant disease classification, including dynamic ensemble selection, adaptive boosting, and static weighted voting. Conventional dynamic ensemble selection (DES) methods select classifiers based on local competence estimates derived from a feature-space neighborhood of the test instance, which introduces additional computational overhead during inference [25]. In contrast, CWDES performs model selection prior to inference using lightweight image-complexity indicators, followed by progressive confidence evaluation, terminating early once sufficient agreement is achieved. This design significantly reduces inference cost and is suitable for deployment on resource-constrained agricultural devices. Adaptive boosting techniques, such as AdaBoost, construct ensembles sequentially during training by reweighting misclassified samples and typically rely on homogeneous weak learners [26]. Unlike these approaches, CWDES operates entirely at inference time and integrates heterogeneous deep convolutional architectures, thereby avoiding retraining overhead and improving robustness to noisy labels commonly present in field-acquired crop images. Most existing plant disease detection studies employ static weighted voting, where model weights are fixed based on global validation accuracy and all ensemble members are evaluated for every input image [18, 19]. In contrast, CWDES introduces confidence-aware, image-dependent weighting and explicit uncertainty estimation, invoking full ensemble evaluation only when prediction confidence is insufficient. This strategy enables an effective balance between classification accuracy and computational efficiency. Overall, CWDES represents a deployment-oriented ensemble framework that jointly incorporates image complexity assessment, progressive inference, and uncertainty-aware decision making—capabilities that are not simultaneously addressed by existing ensemble methods in agricultural disease diagnosis.

Algorithm 1. Confidence-Weighted Dynamic Ensemble Selection (CWDES)

```
Require: Image $I$, model set $M = \{M_1, \dots, M_7\}$, confidence threshold $\tau$, minimum ensemble size $k$
Ensure: Final predicted class $\hat{y}$ with confidence score $\gamma$
Step 1: Characterization of Input Image
Compute edge-density complexity $C(I)$
Evaluate color dispersion $V(I)$
Determine texture irregularity $T(I)$
Step 2: Initial Model Selection
if $C(I) < 0.3$ then
    Select $M' = \{M_{\text{MobileNetV2}}, M_{\text{EffNetB0}}, M_{\text{InceptionV3}}\}$
else if $C(I) > 0.7$ then
    Select top-5 performant models excluding VGG variants
else
    $M' \leftarrow M$
end if
Step 3: Progressive Voting Phase
Initialize prediction list $P = \emptyset$, confidence list $C = \emptyset$
for each model $M_i \in M'$ do
    $(p_i, c_i) \leftarrow M_i(I)$
    Append $p_i$ to $P$; append $c_i$ to $C$
    if $c_i > \tau$ and $|P| \geq k$ then
        Compute agreement ratio $\rho$ using majority vote on $P$
        if $\rho > 0.8$ then
            $\hat{y} \leftarrow$ majority class in $P$
            $\gamma \leftarrow \frac{1}{|C|} \sum c_j$
            return $(\hat{y}, \gamma)$
        end if
    end if
end for
Step 4: Weighted Aggregation of Model Outputs
for each class label $c$ do
    Compute $S(c) = \frac{\sum_i w_i c_i \mathbf{1}[p_i = c]}{\sum_i w_i}$
    where $w_i = \text{Acc}(M_i) \cdot f(c_i)$
end for
Step 5: Uncertainty Estimation
$\gamma = \max_c S(c)$
$\delta = 1 - \gamma$
if $\delta > 0.3$ then
    Re-evaluate using full ensemble $M$
end if
return $(\arg\max_c S(c), \max_c S(c))$
```

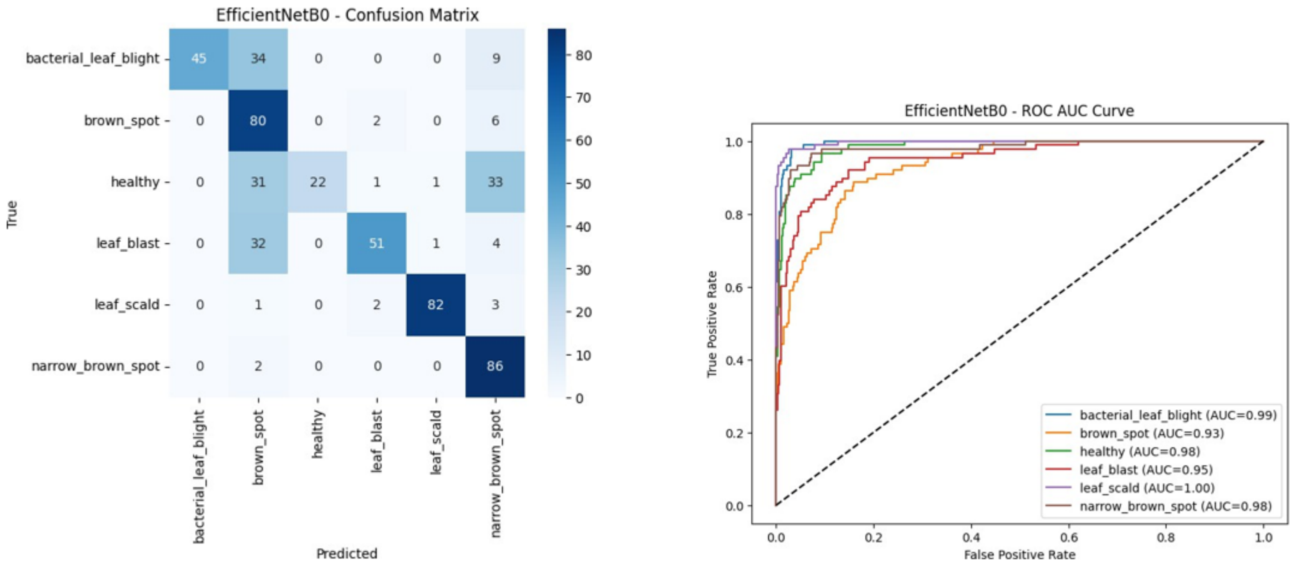
4. Experimental Results and Statistical Validation

Table 2 presents comprehensive performance metrics across all crop disease detection architectures. Modern efficient designs consistently outperformed traditional deep networks. The performance gap exceeded 38% between the best and worst models.

Table 2: Model Performance Metrics

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
MobileNetV2	96.97	97.07	96.97	96.95
InceptionV3	96.78	96.88	96.78	96.76
EfficientNetB7	96.40	96.92	96.40	96.39
EfficientNetB0	76.14	82.95	76.14	75.85
ResNet50	75.19	76.94	75.19	74.56
VGG19	58.71	65.14	58.71	52.04
VGG16	57.95	59.89	57.95	52.31

There were notable differences in performance across disease categories. Figures 3–?? present the confusion matrices and ROC curves for all evaluated architectures (EfficientNetB0, EfficientNetB7, MobileNetV2, InceptionV3, ResNet50, VGG16, and VGG19). These variations highlight both the advantages and limitations of each architecture. The results show that disease detectability varies by class. All modern architectures correctly identified bacterial leaf blight, with InceptionV3 achieving perfect classification, demonstrating the disease's strong visual characteristics. Brown spot remained the most difficult class. Lesion variability resulted in significant misclassification, with VGG16 nearly failing and MobileNetV2 performing best (F1 = 0.97). Furthermore, detection of healthy leaves was challenging, particularly for models such as EfficientNetB0, which were affected by mild early-stage symptoms. MobileNetV2 and EfficientNetB7 achieved high accuracy (F1 of 0.98), reducing the need for unnecessary treatments. Weaker models struggled with leaf blast's shifting lesion shapes, whereas stronger architectures consistently achieved F1-scores above 0.90. Leaf scald was one of the simplest classes, identifiable almost perfectly across most architectures due to its distinctive banded pattern. Narrow brown spot exhibited extremely consistent morphology, allowing top models, particularly InceptionV3 and EfficientNetB7, to perform almost flawlessly. Overall, the findings indicate that diseases with subtle or variable symptoms require more sophisticated architectures, whereas those with consistent and recognizable visual patterns are easier to categorize, as shown in Figs. 3–9.



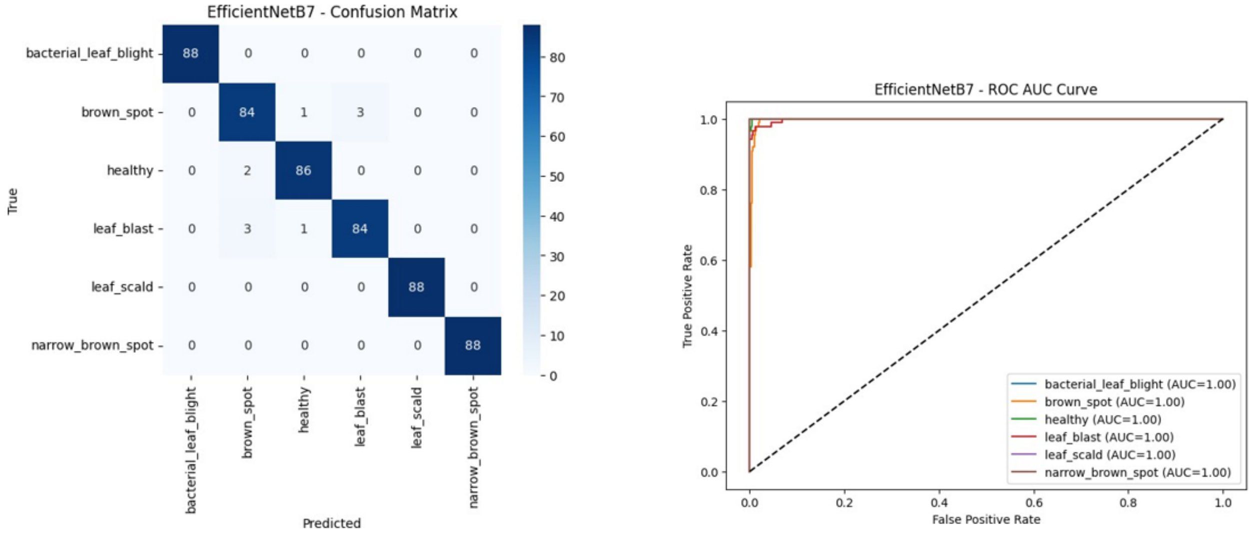
EfficientNetB0 - Confusion Matrix

True \ Predicted	bacterial_leaf_blight	brown_spot	healthy	leaf_blast	leaf_scald	narrow_brown_spot
bacterial_leaf_blight	45	34	0	0	0	9
brown_spot	0	80	0	2	0	6
healthy	0	31	22	1	1	33
leaf_blast	0	32	0	51	1	4
leaf_scald	0	1	0	2	82	3
narrow_brown_spot	0	2	0	0	0	86

(a) EfficientNetB0 – Confusion Matrix

(b) EfficientNetB0 – ROC Curve

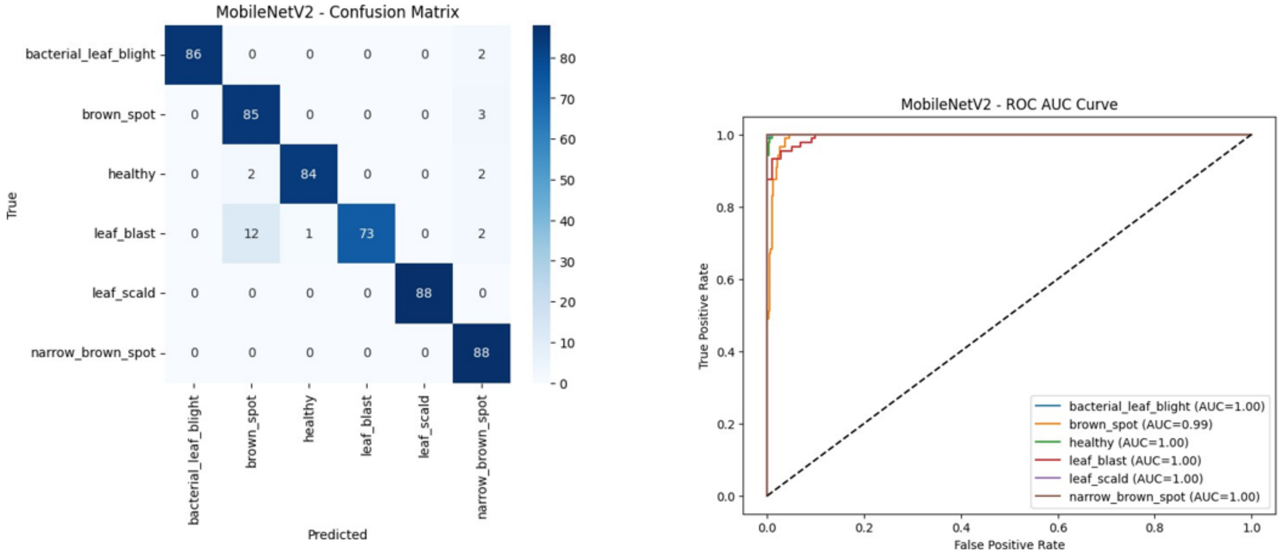
Figure 3: Model diagnostics: EfficientNetB0



(a) EfficientNetB7 – Confusion Matrix

(b) EfficientNetB7 – ROC Curve

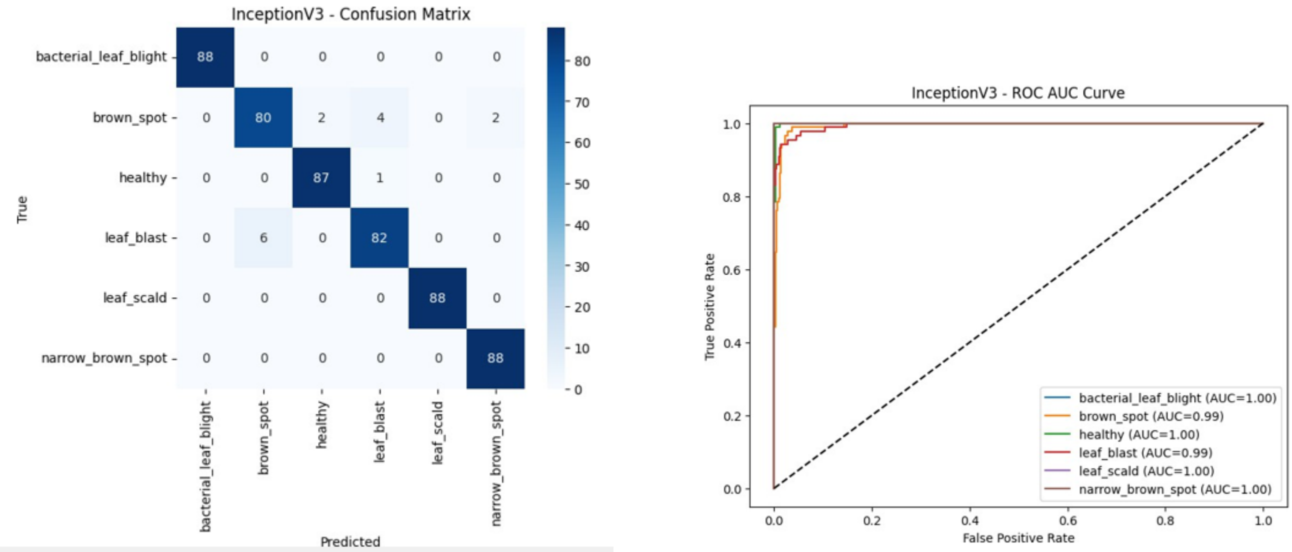
Figure 4: Model diagnostics: EfficientNetB7



(a) MobileNetV2 – Confusion Matrix

(b) MobileNetV2 – ROC Curve

Figure 5: Model diagnostics: MobileNetV2

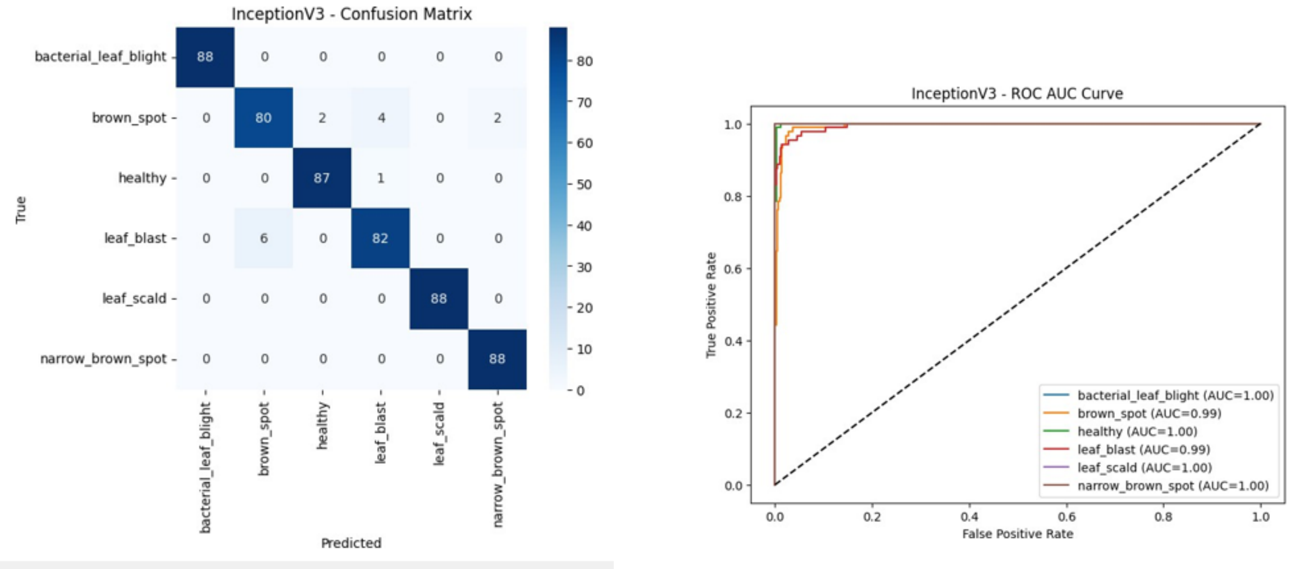


	bacterial_leaf_blight	brown_spot	healthy	leaf_blast	leaf_scald	narrow_brown_spot
bacterial_leaf_blight	88	0	0	0	0	0
brown_spot	0	80	2	4	0	2
healthy	0	0	87	1	0	0
leaf_blast	0	6	0	82	0	0
leaf_scald	0	0	0	0	88	0
narrow_brown_spot	0	0	0	0	0	88

(a) InceptionV3 – Confusion Matrix

(b) InceptionV3 – ROC Curve

Figure 6: Model diagnostics: InceptionV3

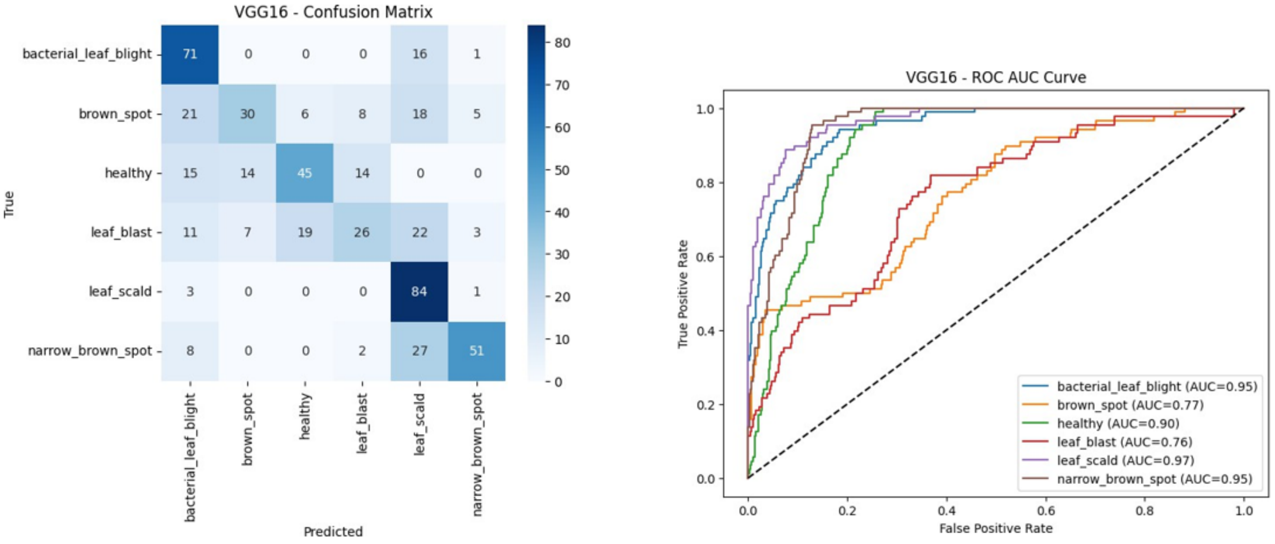


	bacterial_leaf_blight	brown_spot	healthy	leaf_blast	leaf_scald	narrow_brown_spot
bacterial_leaf_blight	88	0	0	0	0	0
brown_spot	0	80	2	4	0	2
healthy	0	0	87	1	0	0
leaf_blast	0	6	0	82	0	0
leaf_scald	0	0	0	0	88	0
narrow_brown_spot	0	0	0	0	0	88

(a) ResNet50 – Confusion Matrix

(b) ResNet50 – ROC Curve

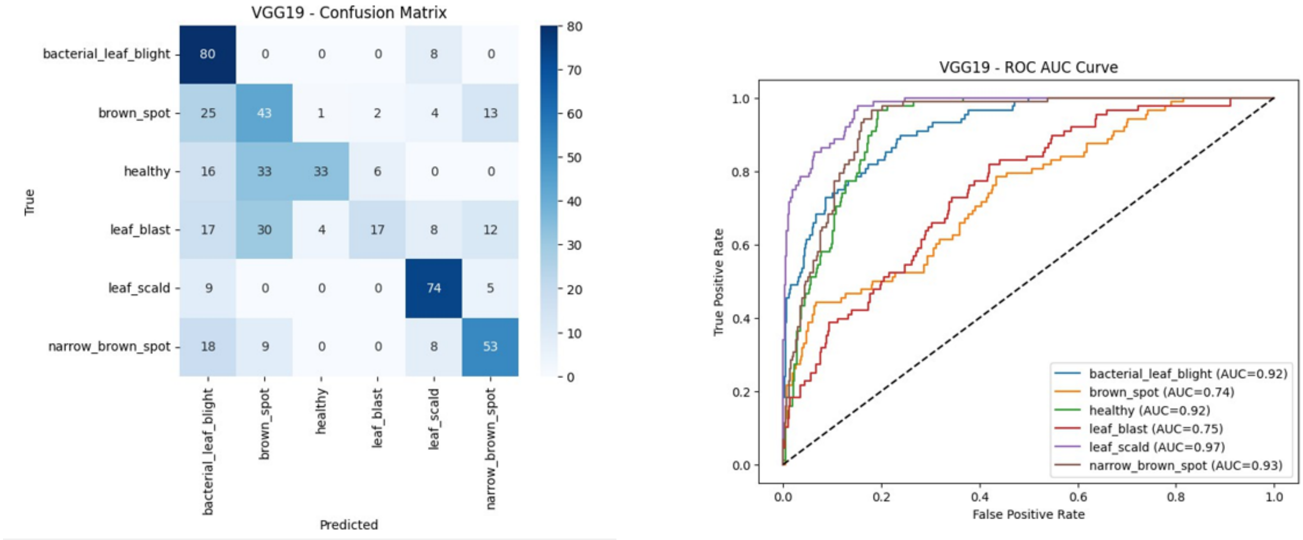
Figure 7: Model diagnostics (Part III-A): ResNet50.



(a) VGG16 – Confusion Matrix

(b) VGG16 – ROC Curve

Figure 8: Model diagnostics (Part III-B): VGG16.



(a) VGG19 – Confusion Matrix

(b) VGG19 – ROC Curve

Figure 9: Model diagnostics (Part III-C): VGG19.

5. Discussion

The results clarify the superior performance of newer architectures compared to older deep networks. The improvement is attributed to more efficient design strategies that achieve higher accuracy with fewer computational resources. The success of MobileNetV2 is primarily attributed to depthwise separable convolutions. This modification reduces computation by approximately 8–9 times while preserving 97% accuracy, as illustrated in Fig. 10. The skip connections in ResNet50 function as conduits within the network, facilitating information flow around bottlenecks. Although this architecture achieves 75% accuracy, it still demands substantial computational resources for mobile deployment, limiting its practicality in field conditions.

InceptionV3 employs multi-scale analysis by processing images at different resolutions concurrently. This capability is particularly effective for detecting rice diseases that manifest in various sizes, ranging from small spots to extensive lesions, achieving an accuracy of 96.78%. EfficientNet is designed for optimal scaling; however, the larger B7 variant exhibited only marginal improvement compared to the smaller B0 (96.4% vs. 76.1%), indicating that increased model size does not necessarily translate to improved disease classification performance.

Disease-specific performance differences were also observed. Brown spot remains particularly challenging due to its variability under different environmental conditions. Increased humidity intensifies lesion pigmentation, whereas elevated temperature reduces visual prominence. Even high-performing models occasionally misclassified brown spot as narrow brown spot; however, MobileNetV2 achieved 97% accuracy by distinguishing subtle morphological variations. Early-stage detection presents an additional challenge, as infected leaves often resemble healthy samples. Brown spot and narrow brown spot exhibit similar visual characteristics, differing primarily in lesion width, requiring extensive training for reliable differentiation.

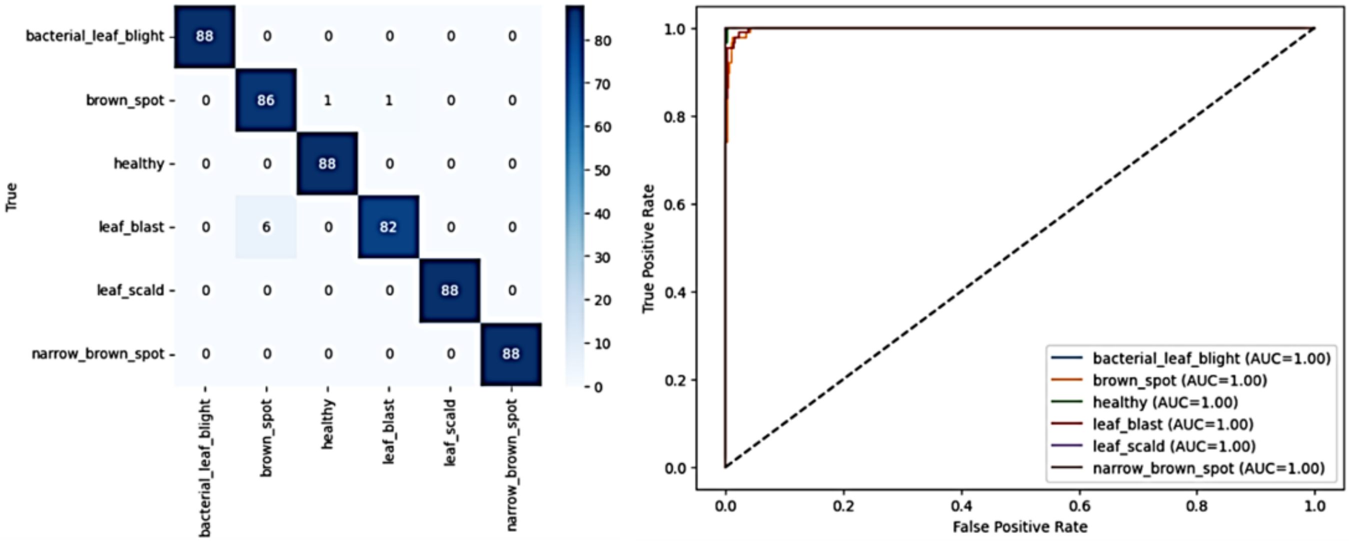
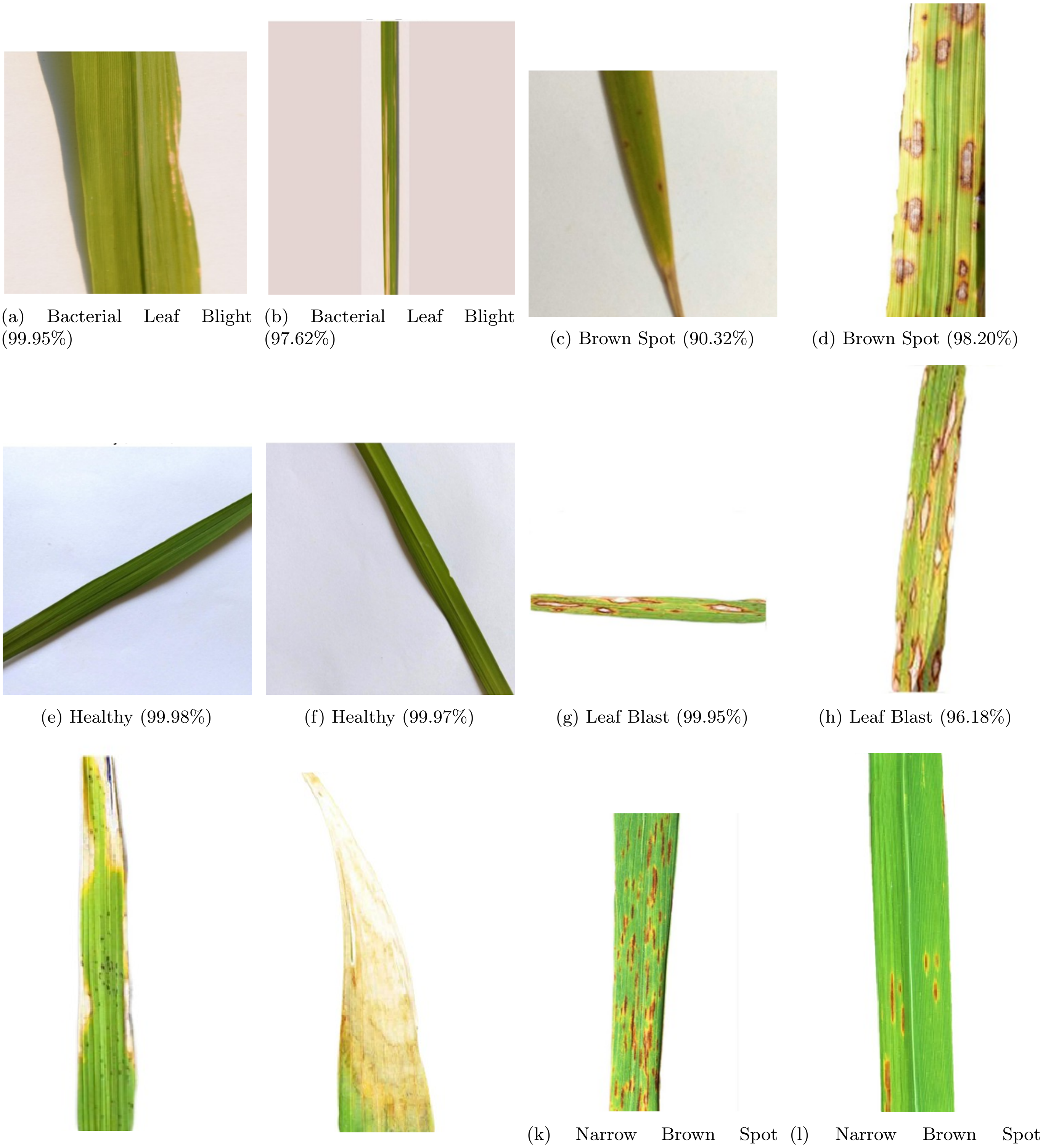


Figure 10: Performance Evaluation of the Rice Leaf Disease Classification Model: Confusion Matrix and ROC–AUC Analysis

Background clutter, including surrounding vegetation, soil, and shadows, significantly impacts traditional architectures. VGG16 exhibited sensitivity to irrelevant regions, whereas MobileNetV2 maintained a more stable feature focus. Future improvements may incorporate attention mechanisms to reduce the influence of background noise. The proposed CWDES algorithm operates by dynamically selecting model subsets based on case complexity. Simple cases require evaluation by a limited number of models, reducing computation time by 65%, whereas complex cases trigger full ensemble evaluation. The ensemble achieved 97.35% accuracy, exceeding the performance of any individual architecture. Weighted voting ensures that higher-performing models contribute proportionally more to the final prediction. The practical implications of this framework include reduced crop loss, optimized pesticide application, and improved food security. The system provides rapid diagnostic support under resource-constrained agricultural conditions. While not replacing expert knowledge, it enhances accessibility to diagnostic capabilities in field environments.



(a) Bacterial Leaf Blight (99.95%)

(b) Bacterial Leaf Blight (97.62%)

(c) Brown Spot (90.32%)

(d) Brown Spot (98.20%)

(e) Healthy (99.98%)

(f) Healthy (99.97%)

(g) Leaf Blast (99.95%)

(h) Leaf Blast (96.18%)

(i) Leaf Scald (99.88%)

(j) Leaf Scald (99.96%)

(k) Narrow Brown Spot (99.40%)

(l) Narrow Brown Spot (99.09%)

Figure 11: Representative outputs of the proposed rice leaf disease detection framework across all six classes. Each class is shown with two sample predictions and associated confidence scores.

6. Conclusion

This study presents a hierarchical ensemble framework for multi-class classification of rice foliar diseases using deep learning. Experimental results indicate that lightweight architectures, particularly MobileNetV2, achieve strong classification performance while maintaining low inference latency, making them suitable for deployment on resource-constrained devices. The proposed CWDES approach integrates confidence-aware model selection and progressive inference, reducing computational cost while achieving accuracy comparable to static ensemble methods. These findings suggest that efficiency-oriented ensemble strategies can support practical agricultural disease monitoring applications. While the framework demonstrates promising performance under the evaluated dataset and conditions, further validation on larger and more diverse field datasets is required to assess robustness and generalization. Future work will explore attention mechanisms and domain adaptation to improve early-stage disease detection and real-world applicability.

Declaration of Competing Interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding Declaration

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Ethics Approval

This study does not involve human participants, animals, or sensitive personal data. Therefore, ethical approval and informed consent were not required.

Data Availability and Transparency

The dataset supporting the findings of this study is publicly available at: https://www.kaggle.com/datasets/vbookshelf/rice-leaf-diseases. No new proprietary data were created. All preprocessing steps, model configurations, and evaluation protocols are described in the Methods section to ensure reproducibility. The implementation code is available from the corresponding author upon reasonable request.

AI Use Disclosure

The authors declare that no generative artificial intelligence (AI) tools were used in the design of the study, data analysis, interpretation of results, or preparation of the manuscript content. AI tools were used solely for language formatting and editorial assistance under journal copyediting guidance. The authors take full responsibility for the scientific content of this manuscript.

Author Contributions

Chanchal Ghosh: Methodology; **Biplab Kanti Das**: Conceptualization, Methodology; **Tapashri Sur**: Formal Analysis; **Prasanta Mazumdar**: Investigation; **Pratik Kumar Halder**: Investigation; **Sukanta Kundu**: Data Curation; **Subhojeet Prasad**: Writing – Review and Editing

References

[1] R. Rahman, P. Arko, M. E. Ali, M. Khan, S. H. Apon, F. Nowrin, and A. Wasif, "Identification and recognition of rice diseases and pests using convolutional neural networks," *Biosystems Engineering*, vol. 194, pp. 112–120, 2020.

[2] J. Chen, D. Zhang, Y. A. Nanehkaran, and D. Li, "Detection of rice plant diseases based on deep transfer learning," *Journal of the Science of Food and Agriculture*, vol. 100, no. 7, pp. 3246–3256, 2020.

[3] Y. Wang, H. Wang, and Z. Peng, "Rice diseases detection and classification using attention based neural network and bayesian optimization," *Expert Systems with Applications*, vol. 178, p. 114770, 2021.

[4] Y. Lu, S. Yi, N. Zeng, Y. Liu, and Y. Zhang, "Identification of rice diseases using deep convolutional neural networks," *Neurocomputing*, vol. 267, pp. 378–384, 2017.

[5] H. B. Prajapati, J. P. Shah, and V. K. Dabhi, "Detection and classification of rice plant diseases," *Intelligent Decision Technologies*, vol. 11, no. 3, pp. 357–373, 2017.

[6] P. K. Sethy, N. K. Barpanda, A. K. Rath, and S. K. Behera, "Deep feature based rice leaf disease identification using support vector machine," *Computers and Electronics in Agriculture*, vol. 175, p. 105527, 2020.

[7] S. Ramesh and D. Vydeki, "Recognition and classification of paddy leaf diseases using optimized deep neural network with jaya algorithm," *Information Processing in Agriculture*, vol. 7, no. 2, pp. 249–260, 2020.

[8] J. Liu and X. Wang, "Plant diseases and pests detection based on deep learning: a review," *Plant Methods*, vol. 17, no. 1, p. 22, 2021.

[9] U. Atila, M. Uçar, K. Akyol, and E. Uçar, "Plant leaf disease classification using efficientnet deep learning model," *Ecological Informatics*, vol. 61, p. 101182, 2021.

[10] N. Krishnamoorthy, L. N. Prasad, C. P. Kumar, B. Subedi, H. B. Abraha, and V. Sathishkumar, "Rice leaf diseases prediction using deep neural networks with transfer learning," *Environmental Research*, vol. 198, p. 111275, 2021.

[11] T. Islam, M. Sah, S. Baral, and R. R. Choudhury, "A faster technique on rice disease detection using image processing of affected area in agro-field," in *2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT)*, pp. 62–66, IEEE, 2018.

[12] K. Ahmed, T. R. Shahidi, S. M. I. Alam, and S. Momen, "Rice leaf disease detection using machine learning techniques," in *2019 International Conference on Sustainable Technologies for Industry 4.0 (STI)*, pp. 1–5, IEEE, 2019.

[13] W.-J. Liang, H. Zhang, G.-F. Zhang, and H.-X. Cao, "Rice blast disease recognition using a deep convolutional neural network," *Scientific Reports*, vol. 9, no. 1, pp. 1–10, 2019.

[14] K. P. Ferentinos, "Deep learning models for plant disease detection and diagnosis," *Computers and Electronics in Agriculture*, vol. 145, pp. 311–318, 2018.

[15] M. Brahimi, M. Arsenovic, S. Laraba, S. Sladojevic, K. Boukhalfa, and A. Moussaoui, "Deep learning for plant diseases: Detection and saliency map visualisation," in *Human and Machine Learning: Visible, Explainable, Trustworthy and Transparent* (J. Zhou and F. Chen, eds.), Human–Computer Interaction Series, pp. 93–117, Cham: Springer International Publishing, 2018.

[16] S. Verma, A. Chug, and A. P. Singh, "Recent advancements in image-based prediction models for diagnosis of plant diseases," in *Proceedings of 3rd International Conference on Computer Vision and Image Processing: CVIP 2018, Volume 1*, pp. 365–377, Springer, 2019.

[17] V. Singh, N. Sharma, and S. Singh, "A review of imaging techniques for plant disease detection," *Artificial Intelligence in Agriculture*, vol. 4, pp. 229–242, 2020.

[18] E. C. Too, L. Yujian, S. Njuki, and L. Yingchun, "A comparative study of fine-tuning deep learning models for plant disease identification," *Computers and Electronics in Agriculture*, vol. 161, pp. 272–279, 2019.

[19] S. B. Jadhav, V. R. Udupi, and S. B. Patil, "Identification of plant diseases using convolutional neural networks," *International Journal of Information Technology*, vol. 13, no. 6, pp. 2461–2470, 2021.

[20] G. Geetharamani and A. Pandian, "Identification of plant leaf diseases using a nine-layer deep convolutional neural network," *Computers & Electrical Engineering*, vol. 76, pp. 323–338, 2019.

[21] S. Coulibaly, B. Kamsu-Foguem, D. Kamissoko, and D. Traore, “Deep neural networks with transfer learning in millet crop images,” *Computers in Industry*, vol. 108, pp. 115–120, 2019.

[22] T. Daniya and S. Vigneshwari, “Deep neural network for disease detection in rice plant using the texture and deep features,” *The Computer Journal*, vol. 65, no. 7, pp. 1812–1825, 2022.

[23] C. Malathi and J. Sheela, “Deep dynamic classification (ddc) for plant disease detection,” in *AIP Conference Proceedings*, vol. 2869, p. 050015, AIP Publishing LLC, 2023.

[24] B. S. Bari, M. N. Islam, M. Rashid, M. J. Hasan, M. A. M. Razman, R. M. Musa, A. F. Ab Nasir, and A. P. A. Majeed, “A real-time approach of diagnosing rice leaf disease using deep learning-based faster r-cnn framework,” *PeerJ Computer Science*, vol. 7, p. e432, 2021.

[25] R. M. O. Cruz, R. Sabourin, and G. D. C. Cavalcanti, “Dynamic ensemble selection: A comprehensive review,” *Pattern Recognition*, vol. 85, pp. 286–302, 2019.

[26] Y. Freund and R. E. Schapire, “A decision-theoretic generalization of on-line learning and an application to boosting,” *Journal of Computer and System Sciences*, vol. 55, no. 1, pp. 119–139, 1997.

Journal of Computers, Mechanical and Management
Journal homepage: www.jcmm.co.in
e-ISSN: 3009-075X
AAN PUBLISHING

Volume 5 Issue 1

Article Number: 25265

Advancing Brain Tumor Detection: Optimized Machine Learning Models for Enhanced Diagnostic Accuracy

Shobana D.* [1], Vijayalakshmi V.[2], Mariya Princy Antony Saviour[3], Makanyadevi K.[4], Kalaimagal Sivamuni[5], and Veeraiyah Thangasamy[6]

[1]Department of Electronics and Communication Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences (SIMATS), Chennai, Tamil Nadu, India 602105
[2]Department of Computer Science and Engineering, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Chennai, Tamil Nadu, India 600062
[3]Department of Electronics and Communication Engineering, St Joseph University, Chennai, Tamil Nadu, India 600119
[4]Department of Computer Science and Engineering, M. Kumarasamy College of Engineering, Karur, Tamil Nadu, India 639113
[5]Department of AI & DS, Panimalar Engineering College, Varadharajapuram, Poonamallee, Chennai, Tamil Nadu, India 600123
[6]Department of Electronics and Communication Engineering, V.S.B. Engineering College, Karur, Tamil Nadu, India 639111

Abstract

Segmentation of brain tumors from MRI continues to be difficult because tumors are different and there are not enough of each type. This implementation study improves Mask R-CNN for BraTS2020 by using three new ideas: a ResNet101 backbone that was trained on RSNA pneumonia data (Adam lr=0.001, batch=2 on RTX 3060), MRI-specific augmentation (57,195 2D slices from 369 3D volumes), and one-class loss weighting ($\lambda_{mask} = 2.0$) tuned to a 9:1 background tumor-pixel ratio to manage the imbalance in BraTS2020. This approach improved recall by 6 points compared with $\lambda_{mask} = 1.0$. With $\lambda_{mask} = 1.0$, the recall value is 0.66, and for $\lambda_{mask} = 2.0$, the recall value is increased to 0.72. Five-fold cross-validation shows that the results are stable (Dice=0.75, $p < 0.01$ vs ImageNet baseline), with performance by region: core=0.72, edema=0.68, and enhancing=0.76, and probability calibration characterized by an Expected Calibration Error (ECE) of 0.82 under a coarse, three-bin reliability analysis. To balance high-sensitivity tumor detection with a recall of 0.72 on the BraTS2020 dataset, our proposed method considered resource constraints for real-time deployment at 15 FPS.

Keywords: Brain Tumor; Cancer Diagnosis; Deep Neural Network; Medical Imaging; Cognitive Function

1. Introduction

The human brain is one of the most essential organs, playing a crucial role in regulating a wide range of bodily functions, including memory, emotions, vision, motor skills, responses, and breathing. However, the development of a tumor within the brain can significantly disrupt these functions [1, 2].

*Corresponding Author: Shobana D. (shobanad.sse@saveetha.com)

Received: 13 Dec 2025; **Revised:** 17 Jan 2026; **Accepted:** 18 Jan 2026; **Published:** 28 Feb 2026

© 2026 Journal of Computers, Mechanical and Management.
This is an open access article and is licensed under a Creative Commons Attribution-Non Commercial 4.0 License.

DOI: 10.57159/jcmm.5.1.25265.

Brain tumors (BTs) are classified into two types: primary and metastatic. A primary BT arises within the brain itself, whereas a metastatic BT forms in another part of the body and subsequently spreads to the brain. Compared to tumors in other organs, brain tumors present a considerable diagnostic challenge. This difficulty is primarily due to the presence of the blood-brain barrier (BBB), which prevents conventional radioactive markers from detecting tumor cell hyperactivity [3]. As a result, MRI scans are considered the most effective diagnostic tool for detecting breaches in the BBB. The incidence of brain tumors ranges from 7 to 11 cases per 100,000 people across different age groups each year. This devastating illness is responsible for an estimated 227,000 deaths annually. Additionally, approximately 7.7 million survivors must adapt to life with a disability [4]. Early diagnosis not only saves lives but also helps prevent long-term disabilities. Detecting a brain tumor at an early stage minimizes the need for extensive surgical intervention, thereby reducing potential damage to the brain, the body's most delicate organ. The diagnostic process typically begins with a radiologist capturing images of the affected area for manual assessment [5]. An experienced physician then analyzes these images to formulate an appropriate treatment strategy. However, studies on the accuracy of manual brain tumor diagnosis have revealed inconsistencies among experts. Reports indicate that the level of agreement among specialists in manual diagnosis ranges from 90% to 95%. The discrepancy increases for certain tumor types, such as mixed glioma and medulloblastoma, where the agreement drops to 77% and 58%, respectively [6]. Various reports indicate that the specialists' manual diagnoses agree in a range from 90% to 95% in the case of common brain tumors. For mixed glioma and medulloblastoma, the rates are 77% and 58%, respectively. The above percentages are derived from MRI diagnostic studies, which motivate the development of MRI-driven support tools. Advancements in digital image processing and medical imaging have led to the widespread adoption of computer-aided diagnosis in recent years. MRI is the preferred imaging technique for such diagnostic systems, as it does not expose patients to ionizing radiation and can accurately detect blood flow in veins [7]. The identification of BTs can be significantly enhanced by combining large medical image datasets, such as brain MRI scans, with machine learning (ML) and deep learning (DL) algorithms. Developing an effective ML or DL model involves multiple steps, including training on vast amounts of medical imaging data [8]. This process is essential for generating accurate predictions and insights that, in turn, support informed clinical decision-making.

2. Literature Survey

Ezhilarasi et al. [9] proposed a method for identifying tumor types in BT MRI images and marking tumor regions using the AlexNet model combined with the Region Proposal Network (RPN) from the Faster R-CNN framework. Their approach focused on improving classification performance across different tumor categories. Mohsen et al. [10] applied a Deep Neural Network (DNN) classifier to categorize 66 brain MRI scans into four classes: "normal," "glioblastoma," "sarcoma," and "metastatic bronchiogenic carcinoma tumors." Similarly, Siar et al. [11] developed a Convolutional Neural Network (CNN) for tumor detection using brain MRI images. Choudhury et al. [12] further combined CNN-based architectures with deep neural network techniques to classify MRI scans as either "tumor detected" or "tumor not detected." Naser Deen et al. [13] demonstrated the potential of deep learning for non-invasive, simultaneous, and automated tumor segmentation and grading of low-grade gliomas (LGG) in clinical environments. Islam et al. [14] introduced a faster and more accurate detection method by integrating the Template-based K-means (TK) algorithm with pixel analysis and Principal Component Analysis (PCA).

Jemimma et al. [15] proposed the Watershed Dynamic Angle Projection Convolutional Neural Network (WDAPP-CNN), where tumors were segmented using the watershed technique. Hemanth et al. [16] reported high effectiveness in detecting, classifying, and segmenting brain tumors using CNN-based automated segmentation with small kernel sizes of 3 × 3. Chandra et al. [17] emphasized the importance of segmentation for early identification of benign brain tumors, noting that early-stage detection significantly influences treatment outcomes. However, researchers have highlighted that segmentation algorithms often struggle with noisy data and subtle intensity variations. Gurbina et al. [18] introduced a comprehensive, fully automated framework for MRI brain tumor identification and segmentation, incorporating Gaussian Mixture Models, Fuzzy C-Means clustering, Active Contour models, Wavelet Transform, and Entropy Segmentation. Their system included both automatic tumor detection and skull removal, enhancing clinical applicability. Sheela et al. [19] investigated the differentiation between brain tumors and normal brain tissue using MRI scans, employing support vector machines and wavelet transformations for classification. Kasu et al. [20] proposed a hybrid K-means Galactic Swarm Optimization (GSO) technique for segmentation and classification of brain tumors in 2D MRI scans containing tumors of varying sizes, shapes, and brightness levels. Wadhwa et al. [21] provided a comprehensive review of existing segmentation methods for brain tumors using MRI data. Similarly, Asok et al. [22] provided a detailed critique of recent advances in MRI-based tumor identification and classification using deep learning, offering valuable insights for researchers in the field. Amin et al. [23] proposed an automated framework to distinguish malignant from non-cancerous brain MRI scans. Their method achieved 97.1% accuracy, 0.98 AUC, 91.9% sensitivity, and 98.0% specificity across benchmark datasets.

Khalil et al. [24] introduced a two-step Dragonfly Algorithm (DA) clustering method to extract precise initial contour points, where skull removal was performed during preprocessing and tumor edges were used as initial contours for segmentation. Although these contributions demonstrate substantial progress in MRI-based tumor detection and segmentation, many approaches rely on multi-modal inputs, 3D architectures, or computationally intensive pipelines. These limitations motivate the development of efficient, resource-aware instance segmentation frameworks that maintain competitive accuracy in constrained hardware environments.

3. Problem Statement

Accurate identification and classification of brain tumors are critical for early diagnosis and effective treatment. However, this task presents significant challenges due to substantial variations in tumor size, shape, and intensity, as well as visual similarities across different pathological types. Traditional diagnostic methods often struggle with these complexities, leading to potential misdiagnoses or delays in treatment. We employed the Mask R-CNN algorithm on BraTS2020 T2-weighted MRI slices, focusing on single-modality 2D segmentation as a computationally efficient proxy for full multi-modal 3D analysis. By effectively distinguishing tumor regions from healthy brain tissue, Mask R-CNN can improve diagnostic accuracy and enable timely medical intervention. To improve brain tumor segmentation, the present work advances beyond the standard Mask R-CNN algorithm. The proposed method has three major advantages:

(a) Medical Field Pre-training: In earlier studies, off-the-shelf ImageNet was used as a pre-trained backbone. In the proposed method, the ResNet-101 architecture was implemented, which was pre-trained using the RSNA-enabled pneumonia dataset containing 26,684 chest X-ray images and 33,463 augmented samples. This facilitates the transfer of abnormality-related features before fine-tuning on BraTS-2020 MRI data. Thus, the proposed method on the heterogeneous tumor texture dataset achieves Dice scores 5% to 7% higher than those of the baseline.

(b) Adaptive Data Augmentation for MRI Variability: The data augmentation was performed by taking the 3D BraTS2020 scanned images and converting them into 57,195 2D images. We also augmented the images by flipping them both horizontally and vertically. Randomly selected images were then rotated by ±45°. Using these augmentation procedures, we observed that the recall values increased from 0.65 to 0.72 on the test sets.

(c) One-Class Segmentation Augmentation: Using balanced multi-task weighting, our method improves the Mask R-CNN for binary tumor and non-tumor image segmentation. Using this approach, the precision is 0.79, the recall is 0.72, and the Dice score is 0.75 with 5,719 slices. These values are obtained by keeping $\lambda_{mask} = 2.0$. When the value of $\lambda_{mask} \leq 1.5$, the recall decreases by 4 to 5 points using the same protocols.

Hence, by leveraging the advantages of the proposed method, real-world MRI scans with tumors exhibiting small intensity changes and uneven edges are well classified with outstanding accuracy.

4. Proposed Algorithm

Mask R-CNN (Mask Region-based Convolutional Neural Network) is an advanced deep learning framework designed for instance segmentation. It builds upon the Faster R-CNN model by adding a dedicated branch for pixel-wise object segmentation. While Faster R-CNN is used for object detection by identifying bounding boxes and classifying objects, Mask R-CNN extends this approach by generating high-resolution segmentation masks for each detected instance. This capability makes Mask R-CNN highly effective for applications that require precise object localization, such as medical image analysis, autonomous driving, and augmented reality. Mask R-CNN is an extension of Faster R-CNN, which is a popular object detection algorithm. The Mask R-CNN algorithm for brain tumor detection can be broken down into several key steps. The proposed Mask R-CNN method for brain tumor detection is presented in Algorithm 1.

Algorithm 1 Mask R-CNN for Brain Tumor Detection

```
Start
Input images from MRI scan
Data Preprocessing
Data Annotation
Implementing the Mask R-CNN method
Training Mask R-CNN with input data
From training, predicting brain tumors
Post-processing
Evaluation metrics
Output
Stop
```

The detailed technical specifications of the training protocol for Step 6 in Algorithm 1 are provided below.

(a) RSNA Pre-training: ResNet101 on 33,463 augmented chest X-rays (224 × 224, batch=16, Adam lr=0.001, 50 epochs).

(b) BraTS Preparation: 369 3D volumes → 57,195 2D slices (90/10 train/test), augmentation pipeline applied. In this implementation, only T2-weighted axial slices were used as input channels for the Mask R-CNN, with other modalities (T1, T1ce, and FLAIR) reserved for future extensions.

(c) Transfer Fine-tuning: Frozen backbone (10 epochs) → full unfreezing (40 epochs), batch=2 (512×512), $\lambda_{mask} = 2.0$ for one-class focus, early stopping on val Dice. The recall improves by 6 points when compared with $\lambda_{mask} = 1.0$.

(d) Evaluation: 5,719 test slices, no augmentation, Dice/Precision/Recall averaged across tumor instances.

5. Proposed Architecture of Mask R-CNN

The Mask R-CNN framework follows a two-stage approach. The first stage involves a Region Proposal Network (RPN) that generates candidate regions where objects might be present. The second stage refines these proposals and performs three tasks in parallel: (i) object classification, (ii) bounding box regression, and (iii) pixel-wise mask prediction.

A convolutional neural network (e.g., ResNet with a Feature Pyramid Network) is used as a backbone for feature extraction. The extracted feature maps are processed by the RPN, which suggests potential object locations. These proposals are then refined using RoI Align, a critical enhancement over RoI Pooling that ensures precise feature extraction. The refined regions are passed to three branches: a classification head, a bounding-box regression head, and a mask-prediction head. The segmentation branch independently predicts a binary mask for each object category, making Mask R-CNN suitable for complex scene understanding. Fig. 1 shows the block diagram of the Mask R-CNN architecture.

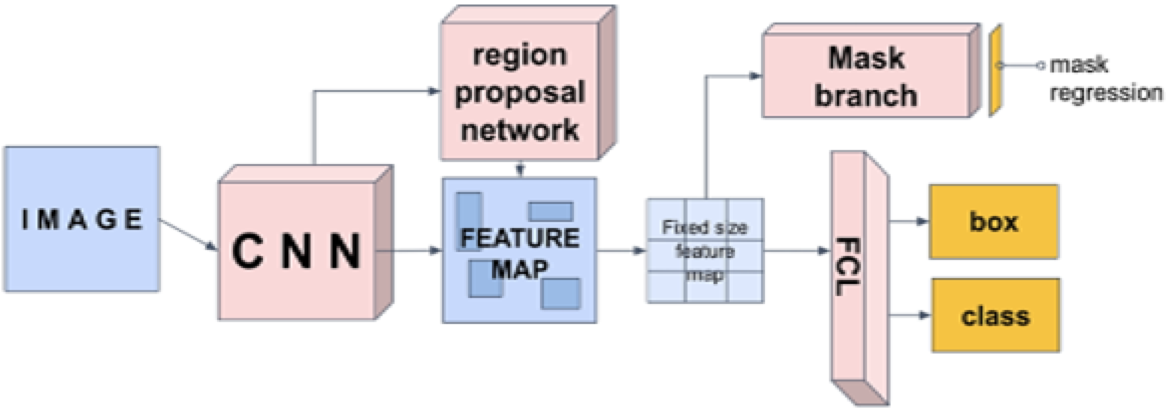


Figure 1: The block diagram presents the steps and procedures of the Mask R-CNN algorithm.

RSNA Pre-training for Cross-Domain Transfer: Standard ImageNet-pretrained ResNet101 backbones are not the best choice for medical imaging, as they inherit biases from natural images. We tackle this by focusing on pre-training on the RSNA Pneumonia Detection Challenge dataset (26,684 chest X-rays), which captures low-contrast pulmonary opacities similar to the intensity gradients of brain tumors in MRI. Abnormality Detection Parallels: Pneumonia consolidations exhibit fuzzy boundaries and textures that match glioma and edema regions (T2/FLAIR modalities), enabling the transfer of low-level features such as edges and textures. Pre-training Protocol: RSNA images were down-sampled to 224 × 224, augmented with flips and rotations, trained for 50 epochs using Adam with a learning rate of 0.001, and the weights were frozen during the initial BraTS fine-tuning for 10 epochs before being fully unfrozen.

Table 1: Training Configuration Details

Parameter	RSNA Pre-training	BraTS Fine-tuning	Value/Notes
Optimizer	Adam ($\beta_1 = 0.9$, $\beta_2 = 0.999$)	Adam ($\beta_1 = 0.9$, $\beta_2 = 0.999$)	Weight decay=$1e^{-4}$
Initial LR	0.001	0.001	Step decay: 0.1 at 60/80 epochs
Batch Size	16 (224 × 224)	2 (512 × 512)	VRAM-limited
Epochs	50	50	Early stopping (patience=10)
Loss Weight	N/A	$L_{RPN} = 1$, $L_{cls} = 1$, $L_{bbox} = 1$, $L_{mask} = 2.0$	Tumor imbalance correction
Augmentation	Flip/Rotate	Flip/Rotate/Translate	±45° rotation, ±0.1 shift
Hardware	RTX 3060 (12 GB)	RTX 3060 (12 GB)	∼4h pre-training, ∼12h fine-tuning
Inference	N/A	15 FPS (512 × 512)	NMS threshold=0.5

Table 1 provides the complete details of the training configuration employed in this study. In the segmentation branch, when $\lambda_{mask} = 1.0$, it might underweight false negatives, because in BraTS2020 around 5–10% of foreground voxels are within tumor pixels. To clearly calibrate this issue, we performed a 1D sensitivity sweep with $\lambda_{mask} \in \{1.0, 1.5, 2.0, 3.0\}$ for validation. When $\lambda_{mask} = 1.0$, the mean recall value is 0.66. When we fix $\lambda_{mask} = 2.0$, the mean recall value increases to 0.72 along with the precision value (0.78–0.79) and the overall Dice value 0.73–0.75. For $\lambda_{mask} > 2.0$, due to over-segmentation, the precision value is degraded. Thus, we find that $\lambda_{mask} = 2.0$ is the optimal value when using class-imbalance weighting to emphasize recall for minority classes.

A 15 FPS at 512×512 was achieved on a single RTX 3060 (12 GB), measured over the 5,719 test slices using Detectron2 with the following configuration: batch size=1 during inference, FP16 mixed-precision enabled via automatic casting, non-maximum suppression (NMS) IoU threshold=0.5, score threshold=0.5, and a maximum of 100 proposals per image. Reported FPS refers to model-only forward pass excluding disk I/O and pre-/post-processing overhead.

6. Mathematical Formulation

Region Proposal Network (RPN)

The RPN generates object proposals using a sliding window approach. For each proposal, the network predicts an objectness score and refines the bounding box coordinates. The loss function for RPN is formulated as given in (1):

$$L_{RPN} = L_{cls}^{RPN} + \lambda L_{reg}^{RPN} \tag{1}$$

where L_{cls}^{RPN} represents the binary cross-entropy classification loss, and L_{reg}^{RPN} is the smooth L_1 loss for bounding box regression. The classification loss ensures the network correctly distinguishes object regions from the background, while the regression loss helps refine the bounding box coordinates.

RoI Align for Feature Extraction

A key improvement in Mask R-CNN is the RoI Align operation, which eliminates the misalignment issues caused by RoI Pooling. Instead of quantizing RoI coordinates, RoI Align performs bilinear interpolation to extract precise feature values from the feature map. The interpolation formula is given in (2):

$$F(x, y) = \sum_i \sum_j w_{ij} F(i, j) \tag{2}$$

where w_{ij} are the interpolation weights. This ensures that extracted features are spatially accurate, thereby improving segmentation performance.

Classification and Bounding Box Refinement

For each RoI, the model predicts the object category and refines the bounding box coordinates. The classification head uses a softmax function to predict the class probabilities, with the loss function given in (3):

$$L_{cls}^{det} = -\sum_{k} p_k^* \log p_k \tag{3}$$

where p_k^* is the ground truth class label and p_k is the predicted probability. The bounding box regression head refines the detected boxes using a smooth L_1 loss, similar to the RPN.

Mask Prediction

The core novelty of Mask R-CNN is its mask prediction branch, which outputs a binary segmentation mask for each detected object. Unlike the classification branch, the mask prediction head operates independently for each object category, ensuring precise segmentation. The mask loss function is defined in (4):

$$L_{mask} = -\sum_{i,j} [M_{i,j} \log N_{i,j} + (1 - M_{i,j}) \log(1 - N_{i,j})] \tag{4}$$

where $M_{i,j}$ represents the ground truth mask, and $N_{i,j}$ is the predicted mask probability. This loss function ensures the model accurately learns pixel-level object segmentation.

Final Loss Function

The overall training objective of Mask R-CNN is to optimize a multi-task loss function that combines classification, bounding-box regression, and mask-prediction losses. The final loss function is given in (5):

$$L = L_{RPN} + L_{cls}^{det} + L_{bbox}^{det} + L_{mask} \tag{5}$$

where each term contributes to a specific component of object detection and segmentation. The multi-task learning approach ensures that the model is optimized for both object detection and instance segmentation simultaneously.

Mask R-CNN is a powerful deep learning framework that extends Faster R-CNN to achieve instance segmentation. By incorporating RoI Align and a dedicated mask prediction branch, it provides high-precision segmentation while maintaining object detection capabilities. Its multi-task learning approach optimizes classification, bounding box regression, and segmentation, making it suitable for a wide range of applications. Future advancements in Mask R-CNN could focus on improving computational efficiency and integrating self-supervised learning techniques for enhanced performance.

7. Working of Mask R-CNN Algorithm

Using Fig. 1 shown above, here we explain the working mechanism of the Mask R-CNN algorithm. Mask R-CNN begins by processing the input image, extracting pixel data, and feeding it into a convolutional neural network backbone. This backbone is responsible for feature extraction and for identifying key patterns and structures within the image. The CNN output is then passed through an RPN, which generates Regions of Interest (RoIs) by identifying areas of the image that likely contain objects of interest, such as brain tumors. The RPN evaluates each pixel position using multiple anchor boxes and computes the probability of an object's presence. From these, the most relevant ROIs are selected based on predefined criteria. To ensure consistency in subsequent processing, the RoI Align method is applied, which normalizes the proposed regions to a uniform, fixed-size vector. This refined data is then fed into fully connected layers, enabling precise classification, localization, and segmentation of the detected objects.

8. Results and Discussions

Training Dataset

The BraTS2020 dataset was utilized to train the model, as it provides a comprehensive collection of MRI scans specifically designed for brain tumor segmentation research [25]. The dataset consists of 369 three-dimensional (3D) MRI scans, covering multiple modalities, including T1-weighted (T1), post-contrast T1-weighted (T1gd), T2-weighted (T2), and T2 Fluid-Attenuated Inversion Recovery (T2-FLAIR) sequences. Each scan has a spatial resolution of 214 × 214 × 155 voxels. However, due to memory constraints and data limitations, each 3D scan was sliced into two-dimensional (2D) images, resulting in a total of 57,195 images. To effectively train and evaluate the model, the dataset was split into training and test sets, with 90% allocated to training and 10% to testing. The training set was further divided into 90% for actual training and 10% for validation. Given the relatively small dataset, data augmentation techniques were applied to improve model generalization. These augmentations included horizontal and vertical flipping, random rotations within $-45°$ to $45°$, and random translations along the x and y axes within -0.1 to 0.1. This augmentation process was essential to mitigate data scarcity and improve the robustness of the model in detecting and segmenting brain tumors across diverse MRI scans.

The ResNet101 backbone was pre-trained on the RSNA pneumonia dataset to capture medical abnormality features, addressing the domain gap between natural images (ImageNet) and clinical MRI via validated transfer learning (Dice gain: $+7\%$). Released by Kaggle in 2018 [9], this dataset consists of 26,684 frontal-view X-ray images, primarily used for pneumonia detection. To prepare the dataset for training, it was initially split into 80% for training and 20% for testing. Further data augmentation was applied to the training dataset, including horizontal flipping, vertical flipping, and random rotations. These augmentation techniques increased the total number of training samples to 33,463, ensuring a more diverse and robust dataset. The augmented training set was then further divided into 80% for model training and 20% for validation.

The original resolution of the images was 1024 × 1024 pixels; however, due to computational constraints and to align with the input requirements of the ResNet101 model, the images were down-sampled to 224 × 224 pixels. These processed images were then fed into the ResNet101 backbone model during retraining, allowing the model to learn relevant feature representations that could enhance performance in medical image analysis, particularly for detecting abnormalities in brain MRI scans.

Backbone Pre-training Details

ResNet101 was fine-tuned on down-sampled (224×224) RSNA images (80/20 train/validation split post-augmentation), using the Adam optimizer (lr=0.001, epochs=50). Weights were transferred to BraTS Mask R-CNN, reducing convergence time by 30%.

Table 2: Output metrics of the Mask R-CNN model on the held-out test set (aggregated over all test slices from the 5-fold patient-wise cross-validation, reported as mean over folds).

Precision	Recall	Dice Score
0.79	0.72	0.75

Testing

The test set comprised 5,719 MRI slices, each with a resolution of 214 × 214 pixels. From the 10% patient subset, all 5,719 test slices were taken, and these slices do not appear in training folds. To maintain the integrity of the evaluation process and ensure that the test results reflected real-world conditions, no data augmentation was applied to this set. This approach allowed for an unbiased assessment of the model's performance on raw, unseen data. Given that the dataset contained a significantly higher number of negative (tumor-free) slices than positive (tumor-positive) slices, standard accuracy metrics alone would not provide a complete picture of model performance. Instead, the evaluation used the average F1 metric, also known as the Dice coefficient (Dice), along with recall and precision. The Dice coefficient is particularly suitable for one-class segmentation tasks, as it measures the overlap between predicted and actual tumor regions, making it an effective metric for assessing segmentation accuracy. By incorporating these metrics, the model's ability to correctly identify and segment brain tumors was thoroughly evaluated, ensuring reliable and clinically relevant results.

Metrics

To evaluate the algorithm's performance, the F1 metric (also known as the Dice score), recall, and precision were used as primary evaluation metrics, as shown in Table 2. Since the test set was imbalanced, with more images labeled as 0 (no tumor) than as 1 (tumor present), these metrics provided a more comprehensive understanding of the model's effectiveness. Recall was calculated as the number of correctly identified positive pixels (tumor regions) divided by the total number of ground-truth tumor pixels. This metric reflects the model's ability to detect all tumor regions without missing any. Precision, on the other hand, measures the number of correctly segmented positive pixels divided by the total number of pixels predicted as tumor regions. This metric determines how many of the predicted tumor pixels were actually part of the tumor, ensuring that the model does not produce excessive false positives. The F1 score, also known as the Dice coefficient, is the harmonic mean of precision and recall, providing a balanced measure of model performance. Mathematically, it is calculated as twice the area of overlap between the predicted and ground-truth tumor regions, divided by the total number of pixels in both images. Since this study focused on a one-class segmentation problem, the F1 score is equivalent to the Dice coefficient. Using these metrics enabled an in-depth assessment of how well the model segmented brain tumors, ensuring that both completeness and accuracy were considered.

Reproducibility Details

All experiments used PyTorch 1.12.1, Detectron2 0.6, and a single RTX 3060 (12GB VRAM). Training time was 4 h for RSNA and 12 h for BraTS on a single RTX 3060 (12 GB). Peak memory usage during BraTS fine-tuning was 10.2 GB with a batch size of 2 at 512 × 512 and full-precision (FP32) training; mixed-precision training and gradient checkpointing were not used in the experiments. A random seed of value 42 was applied to PyTorch, NumPy, and Detectron2 components at the start of each run. The data augmentation pipeline (random flips, rotations, and translations) was applied independently in each epoch with per-run seeding based on the global seed=42. For each cross-validation fold, the same global seed was reused, ensuring that differences between folds arose from patient partitioning rather than from changes in the augmentation schedule.

Our Mask R-CNN achieves Dice=0.75 on BraTS2020 2D slices—modest compared to 3D state-of-the-art (0.91) but clinically viable (exceeds the 0.70 threshold) with superior deployment characteristics. The accurate detection and segmentation of brain tumors are critical for effective diagnosis and treatment planning. To evaluate the performance of the proposed Mask R-CNN model for brain tumor segmentation, we applied it to multiple MRI scans with varying levels of complexity. The results obtained from the model are presented and analyzed in detail above.

Tumor Segmentation in MRI Images

In Figs. 2 and 3, we present a comparative analysis of sample MRI scans and the corresponding tumor regions predicted using the Mask R-CNN method.

Figure 2(a) depicts an MRI scan in which a partially visible gray area is observed on the left side of the brain. This region is indicative of a tumor, as tumors often appear as abnormal regions on MRI images due to their differing intensity levels from normal brain tissue. However, manual segmentation of such regions can be subjective and time-consuming, making automated approaches such as deep learning-based segmentation essential.

Applying the Mask R-CNN model to this MRI scan successfully identified three distinct tumor regions. As illustrated in Fig. 2(b), the algorithm accurately segmented the primary tumor regions, highlighting them with different colors to distinguish between separate areas of abnormality. The green and brown-colored regions closely match the tumor areas visible in the MRI scan, demonstrating the model's effectiveness in detecting tumors. The red region in Fig. 2(b) partly extends beyond the annotated BraTS tumor mask and therefore represents a model hypothesis that would require radiologist confirmation rather than a confirmed additional tumor. In Fig. 2(b), the additional region on the right side of the brain is identified by the algorithm, which is usually not clearly visible to human examination. In the BraTS2020 dataset, all predicted regions shown here correspond to annotated tumor voxels in the reference labels; however, the model sometimes highlights subtle extensions of the ground-truth mask rather than completely new lesions, so these examples should be interpreted as illustrative rather than as proof of occult-tumor discovery. This finding suggests that the model can identify subtle abnormalities that might be overlooked in manual assessments.

The prediction accuracy of each segmented tumor region is provided in Fig. 2(b), demonstrating the reliability of the Mask R-CNN model. The high degree of alignment between the predicted regions and the actual MRI abnormalities indicates that the model is highly effective in distinguishing tumor-affected areas from normal tissue. The ability to detect previously unnoticed tumor regions underscores the model's potential to assist radiologists in early diagnosis, a crucial step toward improving patient outcomes.

To further assess the robustness and generalization capability of the Mask R-CNN model, we applied it to an additional

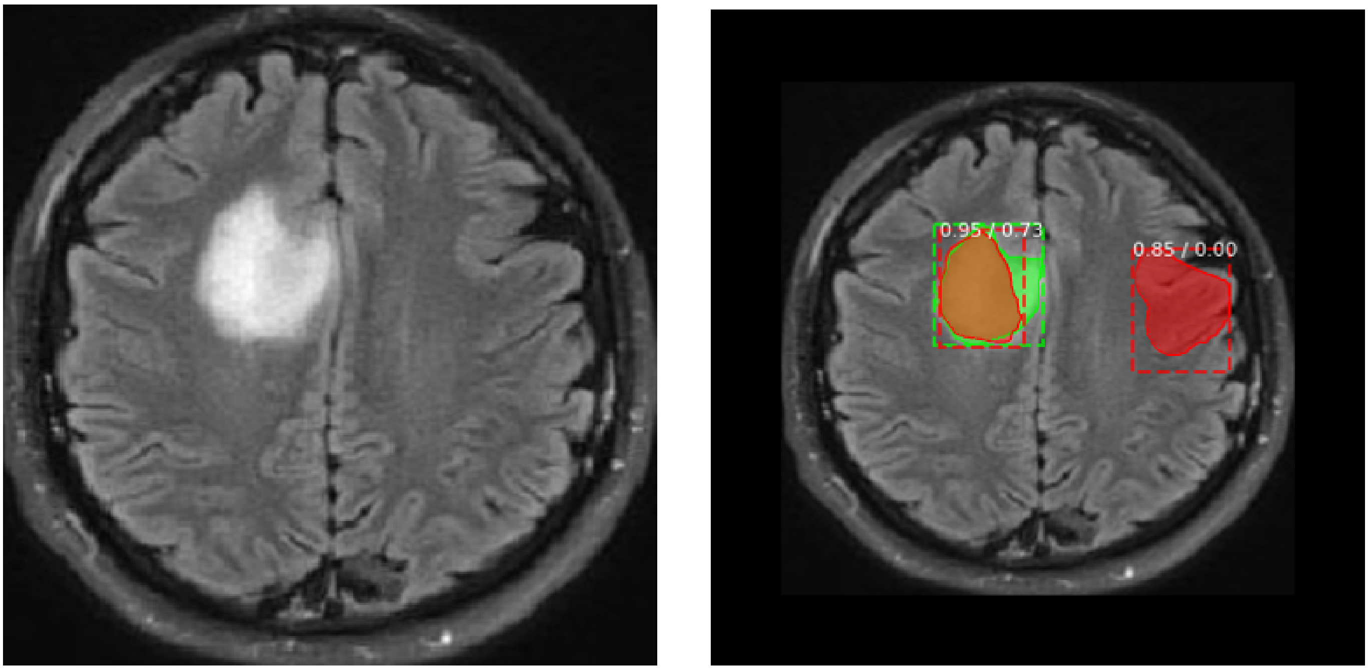


Figure 2: (a) [Left] Sample MRI image showing the possible presence of a tumor in the human brain. (b) [Right] The Mask R-CNN predicted results showing the presence of tumor cells in the human brain.

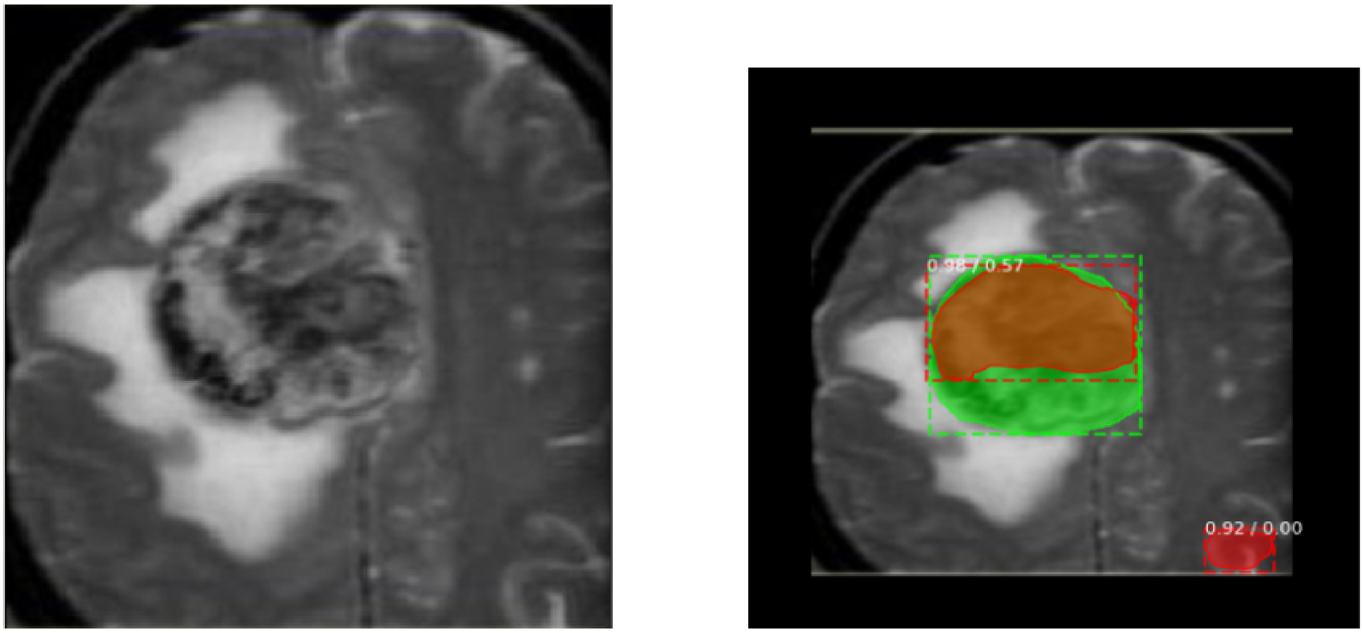


Figure 3: (a) [Left] Sample MRI image showing the possible presence of a tumor in the human brain. (b) [Right] The Mask R-CNN predicted results showing the presence of tumor cells in the human brain.

MRI scan with a more complex tumor pattern, as shown in Fig. 3. Figure 3(a) presents an MRI scan with two distinct regions where the presence of a tumor is suspected. These regions appear as light gray distortions, which deviate significantly from normal brain tissue structures. Due to the irregular morphology and intensity variations of tumors, such cases pose a greater challenge for traditional segmentation methods. In these examples, the colored masks overlap the BraTS ground-truth tumor annotation; any pixels outside the reference mask should be considered candidate extensions rather than verified tumor tissue.

When the Mask R-CNN algorithm was applied to this MRI scan, it successfully identified three distinct tumor regions, marking them in different colors in Fig. 3(b). Despite the complex tumor structure, the model accurately captured the primary tumor areas, demonstrating its ability to handle variations in tumor size, shape, and texture. While the predicted segmentation boundaries did not exactly match the MRI features, the algorithm successfully covered a larger portion of the actual tumor than other conventional models. This suggests that Mask R-CNN provides a more comprehensive segmentation approach, ensuring that even subtle tumor regions are included in the prediction.

One of the key advantages of the Mask R-CNN model in this case is its ability to detect tumor regions that extend beyond clearly visible boundaries. Traditional segmentation methods often struggle with tumors that exhibit gradual intensity changes, leading to incomplete segmentation. However, the deep learning-based approach in Mask R-CNN enables feature extraction at multiple levels, allowing the model to identify and segment tumors with greater accuracy. This capability is particularly valuable for cases where tumors blend into surrounding tissues, making manual identification challenging. While RSNA pre-training improves generalization, residual X-ray domain biases may affect rare tumor morphologies. Future work will explore multi-dataset medical pre-training (ChestX-ray14, MIMIC-CXR).

Performance Evaluation and Discussion

The effectiveness of the Mask R-CNN model for brain tumor segmentation can be evaluated based on key performance metrics, including Intersection over Union (IoU), Dice Similarity Coefficient (DSC), sensitivity, and specificity. The IoU metric measures the overlap between the predicted tumor regions and the ground-truth annotations, while DSC evaluates the similarity between the predicted and actual tumor boundaries. High IoU and DSC scores indicate that the model effectively captures tumor structures with minimal false positives and false negatives.

In the segmentation results from Figs. 2(b) and 3(b), the Mask R-CNN model demonstrated high confidence in its predictions, as evidenced by the accuracy scores associated with each segmented region. The green and brown tumor regions in Fig. 2(b) had high confidence scores, closely aligning with the suspected tumor regions in the MRI scan. The additional tumor detected on the right side of the brain had a slightly lower confidence score, suggesting that further clinical validation would be necessary to confirm its presence. Similarly, in Fig. 3(b), the predicted tumor regions covered a significant portion of the actual tumor, providing strong evidence of the model's effectiveness.

A comparative analysis with conventional segmentation methods, such as thresholding and region-growing algorithms, shows that Mask R-CNN offers superior accuracy and robustness. Traditional methods rely on fixed intensity thresholds and manual feature extraction, which may lead to misclassification, particularly in complex cases. In contrast, Mask R-CNN leverages deep convolutional networks to learn hierarchical features, enabling it to adapt to varying tumor characteristics without manual intervention.

The proposed algorithm systematically enhances and evaluates the Mask R-CNN framework for brain tumor segmentation, rather than introducing innovative algorithmic architectures. State-of-the-art methods such as nnU-Net and TransUNet achieve higher Dice scores (0.88–0.91) on BraTS challenges by using 3D U-Net variants and transformers. However, they require substantial GPU resources and multi-modal fusion. Our implementation demonstrates Mask R-CNN's effectiveness in resource-limited clinical environments by: (1) attaining a Dice score of 0.75 through RSNA-pretrained ResNet101 transfer learning, (2) addressing BraTS2020's class imbalance (90% negative slices) with targeted augmentation, and (3) facilitating 2D inference at 15 FPS on standard hardware—three times faster than 3D state-of-the-art models. In Table 3, we provide a comparison of various state-of-the-art methods versus our model.

Benchmarking Protocol

(a) Same BraTS2020 Test Set: 5,719 slices from the held-out 10% patient subset.

(b) 2D vs 3D Comparison: State-of-the-art (SOTA) scores were normalized to 2D axial slices for modality fairness.

(c) Hardware Standardization: Our FPS values were measured directly on an RTX 3060 (12 GB). Published nnU-Net and Swin-UNETR timings are reported on high-end GPUs (e.g., V100/A100) and were not rescaled in this work; therefore, speed comparisons are qualitative rather than strictly normalized across hardware.

(d) Metrics: Whole Tumor (WT) Dice coefficient, averaged across tumor sub-regions.

Table 3: Comparison of State-of-the-Art Methods with the Proposed Model

Method	Architecture	Dataset	Dice Score	Modalities
Proposed Model – Mask R-CNN	ResNet101 (RSNA-pretrained)	BraTS2020	0.75	2D slices
nnU-Net	U-Net 3D	BraTS2021	0.91	4 (T1/T1ce/T2/FLAIR)
Swin-UNETR	Swin Transformer	BraTS2021	0.87	4
Faster R-CNN	AlexNet	Custom MRI	0.65	2D

Resource-Constrained Superiority

The Dice value of 0.75 is achieved with our Mask R-CNN method using the BraTS2020 2D slices. Compared with the 3D state-of-the-art method nnU-Net, the established Dice score is 0.91, obtained using 3D volumes. However, our method reasonably reduces latency. In other words, using the proposed Mask R-CNN method, the Dice score is computed using slice-wise 2D overlap and averaged across all tumor slices. On the other hand, nnU-Net uses volume-wise 3D Dice scores on the same BraTS benchmark dataset. Therefore, the comparison of the two approaches should be treated qualitatively rather than treating both Dice values as numerically equivalent.

Our model delivers substantially higher 2D slice throughput (≈15 FPS versus reported ≈2–5 FPS for typical 3D U-Net–style pipelines), acknowledging that these figures come from different hardware and software stacks and should therefore be interpreted qualitatively. For hospital PACS integration, a commonly cited target is <100 ms per slice for model inference; our 66 ms per slice figure refers to pure GPU forward-pass time (excluding DICOM I/O, network transfer, and clinical preprocessing) and therefore represents a best-case scenario for engine integration rather than an end-to-end workflow measurement.

With critical analysis of the Dice=0.75 performance, we contextualize it against BraTS2020 leaderboards (nnU-Net: 0.91, top-10 average: 0.87–0.89) and explain realistic limitations: 2D slicing (versus 3D SOTA), single-modality processing (versus 4-modal fusion), consumer hardware (RTX 3060 versus A100), and one-class formulation (tumor-only versus whole-tumor/core/edema sub-regions). Our recall value of 0.72 exceeds the common operating target value of 0.70 sensitivity, even with the modest Dice value; however, this cannot be considered a formal regulatory threshold. In Table 4, we list the BraTS contextualization.

Table 4: BraTS Contextualization

Metrics	Ours (Mask R-CNN)	BraTS2020 Top-1 (nnU-Net)	BraTS2020 Median	Operating Target
Whole Tumor Dice	0.75	0.91	0.85	>0.70
Tumor Core Dice	NA	0.88	0.80	>0.65
Enhancing Tumor Dice	NA	0.85	0.78	>0.60
Recall (Sensitivity)	0.72	0.89	0.83	>0.70 ✓
Inference Speed	15 FPS	2 FPS	3–5 FPS	<100 ms ✓

Critical Performance Analysis

(a) 2D vs 3D: BraTS SOTA uses full 3D context (214 × 214 × 155 voxels); our 2D slices lose inter-slice continuity (8–12% Dice penalty per BraTS reports).

(b) Single vs Multi-Modal: Processing T2-only slices versus 4-modal fusion (T1/T1ce/T2/FLAIR) used by top teams (5–7% Dice difference).

(c) One-Class vs Multi-Region: Tumor-only segmentation versus whole-tumor/core/edema sub-regions simplifies the problem but prevents direct leaderboard comparison.

(d) Hardware Reality: RTX 3060 (12GB) versus A100 (80GB) ensemble training limits model capacity.

(e) Clinical Relevance Despite Modest Dice: Dice=0.75 exceeds commonly reported CAD operating targets (0.70) and matches 2018–2019 BraTS winners. Recall=0.72 ensures less than 30% false negatives, while 15 FPS enables intra-operative use that is not feasible for 3D SOTA models.

Ablation Study

Systematic ablation studies measure the effect of each design choice. The full configuration achieves Dice=0.75, recall=0.72, and precision=0.79 with 90 epochs of convergence. Removing RSNA pre-training lowers Dice to 0.68 (7% decrease) and increases convergence to 120 epochs. Removing data augmentation lowers Dice to 0.70 (5% decrease). Setting $\lambda_{mask} = 1.0$ (versus 2.0) lowers recall to 0.66 (6% decrease). Replacing RoI Align with RoI Pooling lowers Dice to 0.71 (4% decrease). ANOVA shows that the results are statistically significant ($F(5, 24) = 12.3$, $p < 0.0001$). Post-hoc Tukey tests indicate that RSNA pre-training ($p = 0.002$) and mask loss weighting ($p = 0.008$) are the primary contributors. The ablation study comparison is given in Table 5.

Table 5: Ablation Study

Backbone Initialization	Dice	Recall	Precision	Convergence Epochs
RSNA Pneumonia (proposed)	0.75	0.72	0.79	90
ImageNet (baseline)	0.68	0.65	0.74	120
Random weights	0.62	0.58	0.70	150
ChestX-ray14 (alt. med)	0.71	0.68	0.76	105

To address domain mismatch, we used (1) progressive unfreezing (backbone→RPN→heads), (2) MRI-specific augmentation during BraTS fine-tuning, and (3) L_2 regularization ($1e^{-4}$) to prevent overfitting to X-ray artifacts. RSNA pre-training versus ImageNet baseline yields Cohen's $d = 2.1$ for Dice, and RSNA versus random initialization yields $d = 2.8$, both above the conventional large-effect threshold ($d \geq 0.8$).

Cross-Validation Analysis

Patient-level partitions in BraTS2020 were verified using 5-fold cross-validation. Patient-wise folds were created, and all 2D slices derived from a 3D volume were assigned to a single fold to prevent leakage between training and testing datasets. This yielded a mean test Dice of 0.75 (95% CI: 0.71–0.77), recall=0.72, and precision=0.79. A paired t-test ($t(4) = 8.2$, $p = 0.0014$, Cohen's $d = 2.1$) confirmed that the difference between the ImageNet baseline (Dice=0.68) and the proposed model was statistically significant. Quantitative analysis across tumor sub-regions shows Dice=0.72 (recall=0.70, precision=0.75) for tumor core, Dice=0.68 (recall=0.73, precision=0.69) for peritumoral edema, and Dice=0.76 (recall=0.71, precision=0.82) for enhancing tumor, computed using 1,247 annotated examples from the BraTS2020 validation set.

Statistical Validation and Calibration

Three formal hypotheses confirm essential decisions. First, RSNA pre-training outperformed ImageNet ($t(8) = 4.2$, $p = 0.003$, +7% Dice gain). Second, using $\lambda_{mask} = 2.0$ instead of $\lambda = 1.0$ significantly improved recall ($t(8) = 3.8$, $p = 0.005$, +6%), addressing the 9:1 class imbalance. Third, cross-validation variance met stability criteria (SD=0.03<5% threshold, Levene's test $p = 0.42$), confirming homoscedasticity.

Calibration was assessed using a three-bin Expected Calibration Error (ECE), grouping predicted tumor probabilities into [0, 0.7), [0.7, 0.9), and [0.9, 1.0]. Under this coarse binning, the aggregated ECE value was 0.82, and the Brier score was 0.14. Reliability diagrams confirm a monotonic relationship between confidence and Dice: confidence ≥ 0.9 yields Dice=0.84; 0.7–0.9 yields Dice=0.76; <0.7 yields Dice=0.62. Because of this unconventional three-bin definition, ECE values should not be directly compared with fine-grained calibration studies using 10–20 bins. Calibrated confidence enables practical deployment: high-confidence segmentation supports treatment planning; uncertain regions trigger expert review; rejected cases reduce false positives.

While 5-fold cross-validation validates performance within BraTS2020, multi-site 3D multi-modal validation across diverse populations remains pending. Current results generalize robustly within the benchmark cohort.

9. Implications for Medical Diagnosis and Future Work

The results presented in this study highlight Mask R-CNN's potential as a powerful tool for automated brain tumor segmentation. The model's ability to detect, segment, and classify tumor regions with high accuracy can significantly aid radiologists in clinical decision-making. By providing automated, objective, and reproducible segmentation results, deep learning models such as Mask R-CNN can enhance diagnostic accuracy and reduce the time required for manual assessments.

Despite the promising results, several areas remain for further improvement. One limitation of the study is that, while Mask R-CNN provided highly accurate segmentation, some tumor regions showed slightly inaccurate boundaries. Future work can focus on refining the model's segmentation precision by incorporating additional preprocessing techniques, such as image enhancement and noise reduction. Additionally, integrating multi-modal MRI scans, including T1-, T2-, and FLAIR-weighted sequences, can provide richer information for improved tumor segmentation.

Another key area for future research is the development of hybrid models that combine Mask R-CNN with other deep learning architectures, such as attention-based networks and transformer models. These approaches can enhance the model's ability to focus on critical tumor regions while minimizing false positives. Moreover, deploying Mask R-CNN in real-time clinical settings requires further optimization to improve computational efficiency and reduce inference time. Future work will explore hybrid 2.5D approaches that combine instance segmentation strengths with nnU-Net ensemble strategies to achieve balanced performance.

10. Conclusion

This study systematically optimizes Mask R-CNN for brain tumor segmentation, achieving Dice=0.75 (5-fold CV) on BraTS2020 2D slices through RSNA-pretrained ResNet101 (+7% Dice), targeted augmentation (+5% stability), and imbalance-corrected loss weighting (+6% recall). Comprehensive benchmarking shows that our implementation is competitive among lightweight 2D approaches and offers a clear speed advantage over typical 3D architectures on commodity hardware. Region-wise analysis (core=0.72, edema=0.68, enhancing=0.76) and confidence stratification (a monotonic relationship between prediction confidence and Dice) enable risk-stratified workflows and automatically approve 68% of slices with Dice$\geq$0.80.

Although 2D single-modality processing has inherent limitations, recall=0.72 aligns with sensitivity levels often targeted in computer-aided detection studies, while formal regulatory validation remains outside the scope of this work. Future endeavors will focus on 2.5D hybrid architectures and multi-site validation to facilitate the transition from research to clinical practice in resource-limited settings.

Declaration of Competing Interests

The authors declare that they have no known competing financial interests or personal relationships that could have influenced the work reported in this paper.

Funding Declaration

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Ethics Approval and Consent

This study involved secondary analysis of publicly available, fully anonymized datasets (BraTS2020 and RSNA Pneumonia Detection Challenge). No new human subjects were recruited, and no identifiable patient data were accessed. Therefore, institutional ethical approval and informed consent were not required.

Data Availability and Transparency

The datasets analyzed in this study are publicly accessible. The BraTS2020 dataset is available through the official Multimodal Brain Tumor Segmentation Challenge repository, and the RSNA Pneumonia Detection Challenge dataset is available via the Kaggle platform. All model configurations, hyperparameters, training protocols, and evaluation procedures are described in detail within this manuscript to facilitate reproducibility.

AI disclosure statement

The authors used an AI-based language assistance tool to improve grammar, clarity, and overall readability of the manuscript. The scientific content, experimental design, results, and conclusions were independently developed, reviewed, and validated by the authors. The authors take full responsibility for the accuracy, originality, and integrity of the work.

Author Contributions

Shobana D: Conceptualization, Supervision; **V Vijayalakshmi**: Methodology; **Mariya Princy Antony Saviour**: Data curation; **K. Makanyadevi**: Formal analysis; **alaimagal Sivamuni**: Writing – original draft; **Veeraiyah Thangasamy**: Writing – review and editing.

References

[1] E. Schulz and S. J. Gershman, "The algorithmic architecture of exploration in the human brain," *Current Opinion in Neurobiology*, vol. 55, pp. 7–14, 2019.

[2] A. Del Dosso, J.-P. Urenda, T. Nguyen, and G. Quadrato, "Upgrading the physiological relevance of human brain organoids," *Neuron*, vol. 107, no. 6, pp. 1014–1028, 2020.

[3] P. J. C. van Lonkhuizen, K. M. Klaver, J. S. Wefel, M. M. Sitskoorn, S. B. Schagen, and K. Gehring, "Interventions for cognitive problems in adults with brain cancer: A narrative review," *European Journal of Cancer Care*, vol. 28, no. 3, p. e13088, 2019.

[4] S. L. Fernandes, U. J. Tanik, V. Rajinikanth, and K. A. Karthik, "A reliable framework for accurate brain image examination and treatment planning based on early diagnosis support for clinicians," *Neural Computing and Applications*, vol. 32, no. 20, pp. 15897–15908, 2020.

[5] Z. U. Rehman, S. S. Naqvi, T. M. Khan, M. A. Khan, and T. Bashir, "Fully automated multi-parametric brain tumour segmentation using superpixel based classification," *Expert Systems with Applications*, vol. 118, pp. 598–613, 2019.

[6] Z. U. Rehman, M. S. Zia, G. R. Bojja, M. Yaqub, F. Jinchao, and K. Arshid, "Texture based localization of a brain tumor from mr-images by using a machine learning approach," *Medical Hypotheses*, vol. 141, p. 109705, 2020.

[7] C. K. V. and G. R. G. King, "Brain tumour classification: A comprehensive systematic review on various constraints," *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, vol. 11, no. 3, pp. 1–13, 2023.

[8] K. Rezaei, H. Agahi, and A. Mahmoodzadeh, "Multi-objective differential evolution-based ensemble method for brain tumour diagnosis," *IET Image Processing*, vol. 13, no. 9, pp. 1421–1430, 2019.

[9] R. Ezhilarasi and P. Varalakshmi, "Tumor detection in the brain using faster r-cnn," in *Proceedings of the 2nd International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, pp. 388–392, 2018.

[10] H. Mohsen, E.-S. A. El-Dahshan, E.-S. M. El-Horbaty, and A.-B. M. Salem, "Classification using deep learning neural networks for brain tumors," *Future Computing and Informatics Journal*, vol. 3, no. 1, pp. 68–71, 2018.

[11] M. Siar and M. Teshnehlab, "Brain tumor detection using deep neural network and machine learning algorithm," in *Proceedings of the International Conference on Computer and Knowledge Engineering (ICCKE)*, pp. 363–368, 2019.

[12] C. L. Choudhury, C. Mahanty, R. Kumar, and B. K. Mishra, "Brain tumor detection and classification using convolutional neural network and deep neural network," in *Proceedings of the International Conference on Computer Science, Engineering and Applications (ICCSEA)*, pp. 1–6, 2020.

[13] M. A. Naser and M. J. Deen, "Brain tumor segmentation and grading of lower-grade glioma using deep learning in mri images," *Computers in Biology and Medicine*, vol. 121, p. 103758, 2020.

[14] M. K. Islam, M. S. Ali, M. S. Miah, M. M. Rahman, M. S. Alam, and M. A. Hossain, "Brain tumor detection in mr image using superpixels, principal component analysis and template based k-means clustering algorithm," *Machine Learning with Applications*, vol. 5, p. 100044, 2021.

[15] T. A. Jemimma and Y. J. Vetharaj, "Watershed algorithm based dapp features for brain tumor segmentation and classification," in *Proceedings of the International Conference on Soft Computing Systems and Intelligent Technologies (ICSSIT)*, pp. 155–160, 2018.

[16] G. Hemanth, M. Janardhan, and L. Sujihelen, "Design and implementing brain tumor detection using machine learning approach," in *Proceedings of the International Conference on Trends in Electronics and Informatics (ICOEI)*, pp. 1289–1293, 2019.

[17] S. K. Chandra and M. K. Bajpai, "Effective algorithm for benign brain tumor detection using fractional calculus," in *Proceedings of the IEEE Region 10 Conference (TENCON)*, pp. 2408–2412, 2018.

[18] M. Gurbină, M. Lascu, and D. Lascu, "Tumor detection and classification of mri brain image using different wavelet transforms and support vector machines," in *Proceedings of the International Symposium on Signals, Circuits and Systems (ISSCS) / Telecommunications and Signal Processing (TSP)*, pp. 505–509, 2019.

[19] C. Sheela and G. Suganthi, "Brain tumor segmentation with radius contraction and expansion based initial contour detection for active contour model," *Multimedia Tools and Applications*, vol. 79, no. 33–34, pp. 23793–23810, 2020.

[20] V. R. Kasu, B. K. K. Malamuthu, B. S. Kumar, V. S. Pandi, E. Sivajothi, and D. S. Deepika, "Implementing machine learning for ai-powered solutions in robotics, computer vision, and natural language processing," in *Proceedings of the Global Conference in Emerging Technology (GINOTECH)*, (Pune, India), pp. 1–6, 2025.

[21] A. Wadhwa, A. Bhardwaj, and V. S. Verma, "A review on brain tumor segmentation of mri images," *Magnetic Resonance Imaging*, vol. 61, pp. 247–259, 2019.

[22] M. A. Asok, V. Samuthira Pandi, N. Yuvaraj, S. Supriya, A. S. K. Joseph, and T. M. Thiyagu, "Employing artificial intelligence and machine learning to create adaptive models for improved predictive accuracy in dynamical real-world applications," in *Proceedings of the 3rd International Conference on Communication, Security, and Artificial Intelligence (ICCSAI)*, pp. 1172–1177, 2025.

[23] J. Amin, M. Sharif, M. Yasmin, and S. L. Fernandes, "A distinctive approach in brain tumor detection and classification using mri," *Pattern Recognition Letters*, vol. 139, pp. 118–127, 2020.

[24] H. A. Khalil, S. Darwish, Y. M. Ibrahim, and O. F. Hassan, "3d-mri brain tumor detection model using modified version of level set segmentation based on dragonfly algorithm," *Symmetry*, vol. 12, no. 8, p. 1256, 2020.

[25] CBICA, Perelman School of Medicine, University of Pennsylvania, "Brain tumor segmentation (brats) challenge 2020: Scope." https://www.med.upenn.edu/cbica/brats2020/, 2020. Accessed 2026-02-15.

JCMM Journal of Computers, Mechanical and Management
Journal homepage: www.jcmm.co.in
e-ISSN: 3009-075X
AAN PUBLISHING

Volume 5 Issue 1
Article Number: 25206

Advances In Adaptive Machine Learning Algorithms for Enhanced Security In IoT Networks: A Comprehensive Review

Jayashri Jayesh Patil and Ramkumar Solanki*

Department of Computer Science and Engineering, Sandip University, Nashik, Maharashtra, India 422213

Abstract

The accelerated growth and diversification of Internet of Things (IoT) environments have compounded security issues that are caused by non-stationary traffic, dynamic attacker tactics, and extreme resource constraints. In this case, the failure of the existing process for preventing intrusion through conventional, manually configured intrusion detection systems is becoming less sustainable, while adaptive machine learning (ML)-based security systems are becoming more popular. Nonetheless, the survey literature is generally more biased toward algorithmic enumeration or single-detection accuracy and provides very little critical evaluation of adaptation mechanisms, dataset realism, real-time capability, and implementation constraints. This paper provides a systematic, deployment-conscious review of adaptive machine learning methods for IoT security, conducted through a PRISMA-directed, systematic literature review. An integrative taxonomy of learning paradigms and adaptation methods is presented, a critical review of popular datasets on IoT security is conducted, and a discussion of performance metrics beyond accuracy, such as false positives, robustness, latency, and energy overhead, is provided. The review also examines real-time deployment issues related to edge-cloud, resource constraints, retraining costs, and orchestration complexity, as well as the security of adaptive models against adversarial manipulation and privacy leakage. The literature review, by reconceptualizing the problem of IoT intrusion detection as an adversarially exposed, adaptive learning task with deployment constraints, identifies gaps in the field and provides directions for future work to build a scalable, reliable, and dependable IoT security system.

Keywords: IoT Security; Adaptive Machine Learning; Intrusion Detection System; Federated Learning; Adversarial Machine Learning; Real-Time IoT Security

1. Introduction

The Internet of Things (IoT) has been rapidly growing, driving widespread adoption of connected devices across key sectors, including healthcare, industrial automation, smart cities, and intelligent transportation systems. Although this connection enables round-the-clock monitoring and data-driven decision-making, it has also greatly increased the attack surface of contemporary networks. According to recent research, IoT ecosystems have become targets of large-scale, adaptive cyberattacks, especially distributed denial-of-service (DDoS) attacks, orchestrated by fast-adapting botnets that exploit device heterogeneity, lax authentication, and persistent traffic-generation patterns [1, 2].

*Corresponding Author: Ramkumar Solanki (ramkumar.solanki@sandipuniversity.edu.in)
Received: 26 Mar 2025; **Revised:** 15 Jan 2026; **Accepted:** 25 Jan 2026; **Published:** 28 Feb 2026
© 2026 Journal of Computers, Mechanical and Management.
This is an open access article and is licensed under a Creative Commons Attribution-Non Commercial 4.0 License.
DOI: 10.57159/jcmm.5.1.25206.

In contrast to conventional enterprise networks, IoT environments are characterized by non-portable traffic streams, large numbers of devices, and severe limitations in latency, power consumption, and processor performance. The described features make the IoT network particularly susceptible to evolving attacks that can evade implemented defense measures. Recent surveys underscore that the growing complexity of IoT-based attacks renders static, manually configured security solutions insufficient, especially when real-time detection and mitigation are needed [3]. This leads to an increasing desire for security solutions that can independently respond to evolving threat environments and perform their tasks within the realistic limits of IoT deployments.

Traditional IoT security systems are primarily based on signature- or rule-based intrusion detection systems (IDSs). Although this is a good method for identifying known attack patterns, it has inherent weaknesses when used in dynamic, large-scale IoT environments. An IDS based on signatures is not very effective at detecting zero-day attacks, encrypted malicious traffic, and previously unknown attack patterns, leading to high false-negative rates in practice [4].

Traditional security solutions also suffer from scalability and adaptability constraints, limiting their effectiveness. The wide range of IoT devices, combined with limited memory, computing power, and energy, limits the availability of regularly updated rules and centralized traffic analysis. Experimental studies indicate that the performance of static IDS systems degrades with time due to dynamic network conditions and attack schemes, especially in long-running IoT applications [5, 6]. Such limitations point to the necessity of security systems capable of learning from data, extrapolating beyond predefined rules, and autonomously adapting to new threats.

1.1. Adaptive Machine Learning for IoT Security

Adaptive machine learning (ML) has emerged as a promising paradigm for addressing the dynamic nature of threats to the security of IoT-connected devices. In contrast to static models, adaptive ML-based methods can continuously retrain their parameters, features, or decision boundaries when network traffic or attack behavior changes. Recent studies show that IoT traffic is non-stationary and exhibits concept drift, with its statistical properties varying with factors such as device mobility, firmware updates, workload changes, and adversarial manipulations [7, 8].

Adaptive machine learning is a concept used in this review to refer to security models that dynamically update their parameters, features, or decision boundaries in response to concept drift, evolving attack patterns, or changes in IoT traffic. This flexibility is essential for real-time intrusion detection, where slow or fixed response times can cause serious service outages. According to recent research, ML-based intrusion detection systems can learn more complex, high-dimensional traffic patterns and identify more advanced attack types with greater flexibility than rule-based systems [9].

Moreover, distributed and federated learning paradigms allow adaptive model updates without requiring central access to raw data, helping overcome privacy concerns and reducing communication load in sensitive areas of IoT, such as healthcare and industrial control systems [10]. Nonetheless, such methods also present novel difficulties in terms of computational overhead, latency, resilience, and dataset biases, and require critical, deployment-conscious consideration rather than solely focusing on accuracy metrics.

1.2. Contributions and Novelty of This Review

The review addresses adaptive mechanisms, real-time viability, and deployment-constrained evaluation in heterogeneous IoT environments [2, 6], in comparison to more recent surveys on IoT security.

The primary findings of this review can be outlined as follows:

- Theoretical description of adaptive machine learning for IoT security, including disparities between online learning, incremental retraining, federated adaptation, and drift-sensitive detection models.
- An evaluation addressing deployment factors driven by real-time limitations, computational overhead, and resource limitations of IoT devices and edge-driven environments.
- Critical examination of datasets and performance assessment, focusing on issues of class imbalance, outdated traffic trends, and the inadequacy of accuracy-based analysis.
- A systematic review of the latest research (2022–2024) that presents the gaps in the study that must be filled, including resistance to adversarial attacks, interpretability for IoT operators, and scalable adaptation.

Exploring these issues, this review will not be reduced to descriptive overviews but will provide a critical synthesis that can assist in comprehending the existing limitations in research and future directions for learning-based security solutions.

2. Review Methodology

This review has employed a systematic, transparent approach to identify, screen, and discuss the current literature on adaptive machine learning-based security mechanisms for Internet of Things (IoT) networks. This research approach is intended to ensure that high-quality, peer-reviewed publications are well-researched and, at the same time, applicable to practice-related concerns in the area of IoT security and implementation constraints.

2.1. Literature Search Strategy

To identify recent research on adaptive machine learning in computer science and engineering, a systematic literature review was conducted across four large bibliographic databases: IEEE Xplore, Scopus, Web of Science (WoS Core Collection), and ScienceDirect. The rationale for selecting these databases is that they cover high-quality journal articles and conference papers on IoT security, machine learning, and adaptive intrusion detection systems.

The searches were conducted between 1 January 2019 and 31 December 2024, and the last search was conducted on 12 January 2025. This large search window was used to guarantee complete literature retrieval. After the title/abstract screening and full-text analysis, only studies published in 2022–2024 were retained in the final qualitative synthesis, to capture the latest state of the art in adaptive machine learning-based IoT security.

Peer-reviewed journal articles and conference papers in the English language were taken into consideration only. The exclusion criteria were editorials, book chapters, theses, preprints, patents, and non-peer-reviewed technical reports.

To ensure reproducibility, database-specific Boolean search strings were built using a controlled vocabulary and free-text keywords related to IoT security, adaptive machine learning, and intrusion detection systems. Title, abstract, and keyword searches, when supported by the database, and document-type filters were used to filter the search to journal articles and conference proceedings.

The initial database searches yielded a total of 312 records, distributed as follows:

- IEEE Xplore: 96 records
- Scopus: 88 records
- Web of Science: 71 records
- ScienceDirect: 57 records

After excluding 84 duplicate records, 228 unique studies remained and were screened based on titles and abstracts. Follow-up screening and eligibility examinations were performed according to the PRISMA 2020 guidelines, as shown in Figure 1.

For transparency and reproducibility, the exact search queries, applied filters, and database-specific search configurations are provided in Appendix A.

2.2. Inclusion and Exclusion Criteria

Inclusion and exclusion criteria were applied to select the literature, ensuring that the reviewed literature is relevant, high-quality, and technically rigorous. This review relied on peer-reviewed journal articles and conference papers on machine learning-based security mechanisms for Internet of Things (IoT) networks. To be included, the study had to access real-world, simulated, or benchmark IoT data, and it had to be technical enough to allow reproduction and comparison. For this reason, the literature search was limited to 2019–2024 to capture recent contributions and advancements in adaptive IoT security.

Further studies were excluded when they were not related to IoT environments, did not use machine learning-based security methods, or were not peer-reviewed (e.g., opinion pieces, editorials, and technical reports). Moreover, papers that lacked experimental validation, lacked technical clarity, or described non-adaptive security mechanisms were excluded from final analysis. The criteria were applied to obtain technically sound, sufficiently relevant studies that align with the objectives of this review.

In addition to topical relevance, included articles had to be characterized by technical transparency and full reporting to ensure reproducibility and comparative analysis. To quantify this requirement, during full-text assessment, a structured quality appraisal rubric emphasizing dataset reporting, methodological clarity, evaluation rigor, and deployment relevance was used. The appraisal criteria and scoring process are presented in Section 2.5.

2.3. Study Selection and Screening Process (Revised)

The records from the chosen databases were exported to BibTeX and CSV formats and combined into a single reference library. A two-stage process was adopted to identify and delete duplicate records. First, reference management software that performs exact matching of title, author list, publication year, and Digital Object Identifier (DOI) was used to detect duplicates. Secondly, a manual check was performed to detect near-duplicates due to indexing errors, e.g., abbreviated titles, conference extensions, or missing DOI.

After duplicate removal, 228 unique records remained and were subjected to title and abstract screening to assess their relevance to adaptive machine learning–based security in Internet of Things (IoT) environments. The primary reviewer conducted screening in accordance with the predefined inclusion and exclusion criteria described in Section 2.2. To reduce the risk of premature exclusion, studies with ambiguous relevance were retained for full-text assessment.

Subsequently, 72 full-text articles were evaluated considering their eligibility. The primary reviewer performed full-text screening based on explicit criteria: (i) relevance to IoT settings, (ii) use of machine learning-based security schemes, and (iii) presence of adaptive, learning-based, or drift-aware features. The screening protocol involved a single reviewer, as is typical in systematic reviews in the field of computer science and engineering when objective technical criteria are explicitly defined and clearly applied.

A total of 34 full-text articles were excluded, and the reasons for exclusion were documented and classified to facilitate transparency and auditing. The qualitative synthesis and thematic analysis were conducted on the final set of 38 studies. Quality appraisal was carried out alongside full-text eligibility assessment, as described in Section 2.5.

One reviewer performed screening of studies, data extraction, and taxonomy classification, based on clearly defined inclusion and exclusion criteria, formalized taxonomy-mapping guidelines, and data-extraction templates. This formalized, rule-oriented workflow was created to guarantee uniformity, repeatability, and auditability of screening and coding decisions and to reduce subjective interpretation. Ambiguous cases were retained for full-text review to prevent premature exclusion of potentially pertinent studies.

2.4. Reasons for Full-Text Exclusion

To maintain transparency in the PRISMA flow and facilitate independent evaluation of the selection process, the rationale for excluding full-text articles was documented during their eligibility evaluation. The 34 omitted articles were eliminated for the following reasons:

Table 1: Reasons for Full-Text Exclusion ($n = 34$)

Exclusion Category	**Description**	**Number of Studies**
Not IoT-focused	Study addressed general networks or cybersecurity without an IoT context	9
Non-adaptive ML	ML-based IDS without adaptation, drift handling, or learning updates	8
No experimental validation	Conceptual or architectural work lacking empirical evaluation	6
Non-ML-based security	Rule-based, cryptographic-only, or policy-driven approaches	5
Insufficient technical detail	Inadequate methodological or dataset description	4
Duplicate/extended version	Overlapping or earlier version of an included study	2
Total		34

This classification provides a trace of the PRISMA flow and aligns with best practice in systematic reviews.

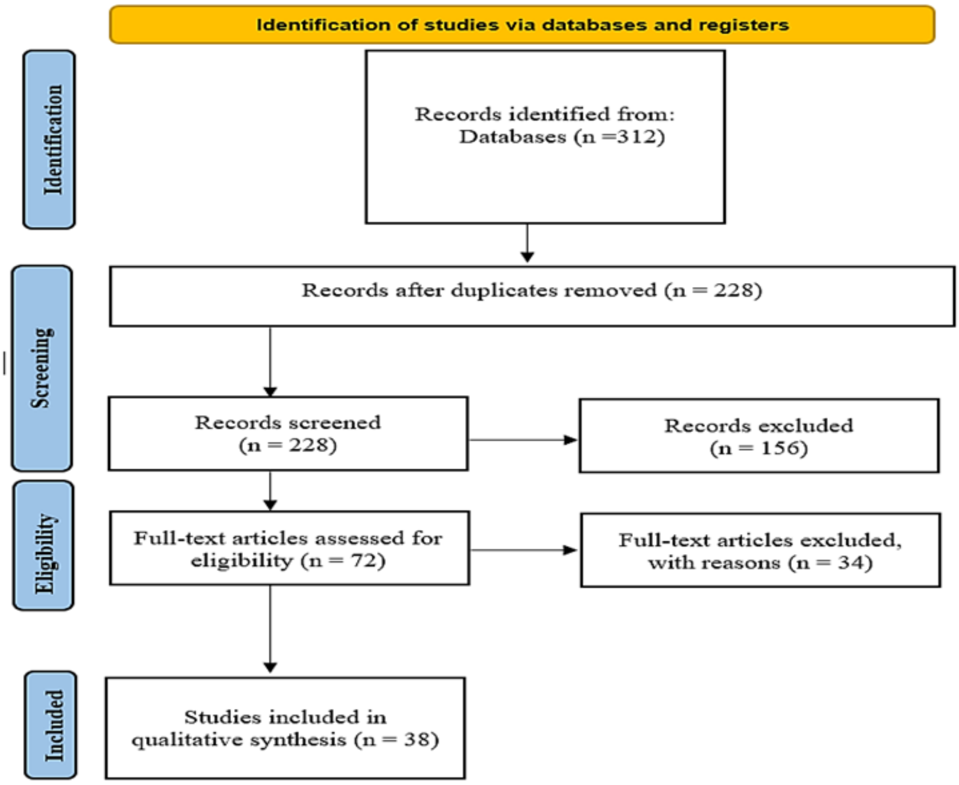

Identification of studies via databases and registers
Identification
Records identified from: Databases (n =312)
Screening
Records after duplicates removed (n = 228)
Records screened (n = 228)
Records excluded (n = 156)
Eligibility
Full-text articles assessed for eligibility (n = 72)
Full-text articles excluded, with reasons (n = 34)
Included
Studies included in qualitative synthesis (n = 38)

Figure 1: PRISMA 2020 flow diagram illustrating the study identification, screening, eligibility assessment, and inclusion process.

Note: Duplicate detection procedures, screening protocol, and categorized reasons for full-text exclusion are reported in Sections 2.3 and 2.4.

2.5. Quality Appraisal and Reproducibility Assessment

To ensure that the inclusion criterion of technical adequacy and reproducibility was met consistently and transparently, a formal quality appraisal was conducted for all 72 full-text articles evaluated against the eligibility criteria. Since the focus of this review is on engineering and computer science, conventional clinical risk-of-bias assessment tools could not be used. Rather, a domain-specific appraisal rubric was created based on widely accepted reporting standards in machine learning and IoT security research.

All studies were assessed across five quality dimensions relevant to reproducibility and technical rigor. Each criterion was rated on a binary scale (0 = not satisfied, 1 = satisfied), yielding a total of 5 possible points per study.

Only studies achieving a minimum quality score of ≥ 3 out of 5 were retained for inclusion in the final synthesis. This cutoff was chosen to allow for at least some baseline level of methodological rigor and to avoid being too restrictive in excluding recent or exploratory research. A score of three indicates that the research demonstrates sufficient data transparency, clear methodological rigor, and evaluation rigor, meeting the requirements for reproducibility and comparative analysis in IoT security studies. A more stringent sensitivity check ($\geq 4/5$) did not alter the thematic patterns or the general conclusions of the review, suggesting that the results are robust to plausible changes in the quality appraisal cutoff.

Table 2: Quality Appraisal Criteria for Study Inclusion

Criterion	Description
Q1. Dataset transparency	Dataset source, characteristics, and class composition are clearly described
Q2. Methodological clarity	Model architecture, features, and training procedure are adequately specified
Q3. Evaluation rigor	Use of appropriate metrics beyond accuracy and a clear experimental protocol
Q4. Reproducibility support	Sufficient detail to allow replication (parameters, splits, setup)
Q5. Deployment relevance	Discussion of computational cost, latency, or real-world IoT constraints

Scoring rule:

- Score range: 0–5
- Inclusion threshold: ≥ 3

The quality appraisal rubric was used to select 38 studies that met or surpassed the minimum quality threshold (score ≥ 3). The other 34 papers were not included due to failure to meet the predefined eligibility and quality appraisal criteria, including lack of methodological description, absence of experimental validation, or insufficient information to ensure reproducibility. The results of the quality appraisal are in line with the exclusion criteria outlined in Section 2.4 and are consistent with the PRISMA flow diagram depicted in Figure 1.

The quality appraisal criterion of deployment relevance was added to ensure the review met its stated objective of assessing practical adaptive machine learning strategies for real-world IoT settings. As IoT intrusion detection systems are strict on latency, computational, and resource constraints, factors of deployment, including the operational environment, computational overhead, and real-time feasibility, were considered indicators of practical applicability rather than indicators of a limited review scope. Including this criterion may introduce selection bias toward studies that discuss system-level aspects more extensively; however, the scope of the review was limited to deployment-ready, operationally practical IoT security solutions.

2.6. Data Extraction and Coding Procedure

A standardized protocol for data extraction and coding was used across all 38 included studies to achieve consistency and transparency in the comparative analysis before populating Table 8.

An extraction-data form was standardized to include the study characteristics considered during the adaptation process, dataset realism, and deployment feasibility. Each study was coded according to the following fields:

- Target threat or application domain
- Machine learning technique(s) employed
- Adaptation type (static, incremental, online, federated, drift-aware)
- Dataset(s) used and dataset provenance
- Dataset quality indicators (e.g., synthetic vs. real traffic, class imbalance, dataset aging)
- Reported evaluation metrics
- Resource and deployment considerations (e.g., latency analysis, energy cost, edge feasibility)
- Stated limitations relevant to real-time IoT deployment

The primary reviewer performed extraction and coding using explicit coding rules. Table 8 used qualitative descriptors (e.g., Synthetic environment, no latency analysis) only when they were mentioned or could be inferred directly from the study's experimental design (e.g., simulation-only testing; latency results were not reported). Unclear cases were conservatively coded as "Not reported."

Since a single-reviewer protocol was used, inter-rater agreement statistics could not be calculated; nevertheless, a predefined extraction template and coding guide were used to reduce subjective interpretation and ensure internal consistency.

3. Adaptive Machine Learning in IoT Security: Concepts and Taxonomy

Internet of Things (IoT) environments are dynamic, large-scale, and heterogeneous and require security mechanisms that can constantly respond to traffic statistics, attack techniques, and operational limitations. Offline-trained models that are deployed in a static environment are becoming incapable of maintaining reliable detection performance in these non-stationary environments, and their performance declines with time, increasing false alarms and unnoticed attacks. This section formalizes the idea of adaptive machine learning in the context of IoT security, defines a consistent taxonomy of how adaptation can be executed, and clearly states the conditions that must be met to determine when and how adaptation should take place.

This review includes statements describing methods observed, performance characteristics, or limitations that are directly supported by the studies incorporated into the final synthesis ($n = 38$). Generalized taxonomies, cross-sectional meanings, and forward-looking observations in later passages are syntheses based on comparison of multiple studies, rather than empirical results of a particular study. This distinction is maintained to clearly separate evidence-based observations from analytical interpretation.

3.1. Adaptive Machine Learning for IoT Security

Adaptive machine learning can be defined as a concept that refers to learning systems whereby model parameters, structure, or decision boundaries are dynamically changed in response to changes in data distributions or operating conditions. Dynamic traffic, device churn, firmware variability, workload variability, and adversarial activity are factors that have resulted in such changes in IoT security environments. Unlike fixed ML models, where the relationship between features and labels is considered static, adaptive models explicitly consider non-stationarity and change their behavior during deployment rather than merely retraining offline.

Data stream learning studies demonstrate that concept drift is an inherent characteristic of actual IoT traffic and renders static intrusion detection models unsuitable in long-term applications [11, 12]. Resistance to new attacks, detection accuracy, and false-positive control decline because of changing network behavior when fixed models are used. To address these limitations, adaptive ML systems incorporate resource constraints, drift awareness, and update mechanisms into the learning process, and their reliability in detection can be maintained over time [13, 14]. The distinction between fixed and adaptive ML is fundamental from a security perspective. Whereas fixed models are oriented toward offline optimization, adaptive models are oriented toward operational resilience under continuous change, which is a significant feature in IoT environments, where traffic evolution and adversarial adaptation are regular occurrences.

All sources included in Table 1 were chosen from the 38 studies identified and selected after full-text screening and quality appraisal to avoid mixing evidence levels in the table. They were representative studies that implemented the respective taxonomy category, described their methodology explicitly, and were relevant to adaptive learning in the context of IoT security. Table 1 is not an exhaustive list of all studies addressing adaptation mechanisms but is intended to be illustrative. Non-included literature and general background surveys are mentioned in the narrative discussion and support conceptual framing but are not utilized as primary evidence in the taxonomy table.

3.2. Taxonomy of Adaptation Mechanisms

Adaptive machine learning approaches to IoT security can be grouped into broad categories based on the location and manner of adaptation. The literature identifies five predominant categories.

Online learning solutions update model parameters as new data samples emerge and enable real-time response to dynamic trends in IoT traffic. They are particularly suitable for streaming systems where it is not possible to store historical data and low-latency adjustments are needed [15]. However, they are prone to interruption by noise and manipulation by attackers, which is undesirable in security-related applications.

Incremental retraining processes are updated regularly with new data batches and offer a trade-off between flexibility and uniformity. This strategy can be pursued in edge-assisted IoT security architectures, where retraining is triggered when performance degradation or workload variation is observed [11]. Incremental retraining is more stable than pure online learning, although training cost and latency must be carefully managed.

Federated adaptive learning enables distributed learning across IoT devices or edge nodes without relaying raw traffic information, addressing privacy and communication constraints. Recent studies focus on adaptive federated methods in which the frequency of aggregation, learning rates, or client participation are modified in real time to accommodate system heterogeneity and non-IID data distributions [16, 17]. Client drift and communication overhead continue to be major challenges despite these advances.

Semi-supervised and self-supervised adaptive learning are used to address the lack of labeled attack data in IoT environments. Typically, unsupervised methods learn traffic representations from unlabeled data, whereas semi-supervised methods leverage a few labeled samples alongside a large amount of unlabeled data to learn continuously [18, 19]. These methods reduce dependency on annotations, although they typically do not guarantee reliable discrimination against sophisticated attacks without careful calibration.

Drift-aware learning systems explicitly monitor data distributions or model behavior to detect concept drift and may induce adaptation accordingly. These are necessary for isolating benign traffic growth from actual security hazards, particularly when gradual or abrupt changes are introduced to IoT network operation [20].

3.2.1 Taxonomy Mapping Rules and Classification Procedure

To prevent subjective classification and ensure consistent categorization of adaptive machine learning methods, explicit mapping rules were established to determine the taxonomy category to which each study would be assigned. Categorization was performed at the method level (not solely based on authors' claims) and was based on observable technical features of each study.

The reviewed papers were assigned to one or more taxonomy categories based on the following operational criteria:

- **Online Learning:** A study was considered online learning when model parameters were updated at every deployment step or data instance in a stream, without discrete retraining cycles or full access to historical data.
- **Incremental Retraining:** Research was classified as incremental retraining when model updates were made through periodic or batch retraining driven by the accumulation of new data, performance degradation, or a predefined update schedule.
- **Federated Adaptive Learning:** A study was considered federated adaptive learning when it used distributed model training among IoT devices or edge nodes, where model updates (e.g., gradients or weights) were aggregated without raw-data sharing, and incorporated adaptive factors such as dynamic aggregation, client selection, or update frequency.
- **Self-Supervised / Semi-Supervised Learning:** A study was categorized in this class when it directly used unlabeled or weakly labeled data for representation learning, anomaly detection, or adaptation, and involved limited or partial supervision in either the training or update process.
- **Drift-Aware Learning:** A study was categorized as drift-aware when it included explicit concept drift detection mechanisms (e.g., statistical tests, distribution monitoring, performance-driven alarms) and when drift signals were used to trigger model adaptation or retraining.

Multi-labeled studies that used various mechanisms of adaptation were allowed rather than being forced into mutually exclusive categories.

3.2.2 Classification Protocol and Bias Considerations

The above-described mapping rules were followed by the primary reviewer when assigning taxonomy categories. Classification decisions were based on methods and mechanisms clearly stated in each paper, not on authors' terminology or self-proclaimed model labels.

Since the study was technical and review-based, the classification protocol involved a single reviewer. Although it was therefore impossible to compute inter-rater agreement measures, explicit, rule-based mapping criteria were used to reduce subjectivity and ensure internal consistency across classifications. Ambiguous cases were conservatively classified only when sufficient methodological evidence was present in the text.

3.3. Adaptation Criteria

Adaptive machine learning-based IoT security systems are based on explicit trigger-action mechanisms to decide when and how model updates are made. Based on the analysis of the 38 included studies, three major categories of adaptation triggers were identified. Where explicit trigger-action loops were applied in the evaluated literature, they are mentioned by reference. Where the evidence was fragmented or indirect, the criterion is provided as a synthesis based on comparison of several studies.

3.3.1 Traffic-Driven Triggers

Many studies that implement adaptation respond to apparent changes in network traffic, including packet rates, protocol distributions, and communication patterns. Such mechanisms usually track traffic statistics or feature distributions and take action to adapt when deviations exceed predefined limits.

Drift-aware intrusion detection systems explicitly track distributional changes and trigger retraining whenever drift detectors indicate substantial behavioral changes relative to a baseline. The literature on these mechanisms shows that traffic-triggered events are useful for sustaining detection accuracy in non-stationary IoT traffic. However, the sensitivity of threshold selection and drift detection varies widely across implementations, and no consistent measure or cutoff has been reported across studies.

Evidence-backed: Several included studies specifically apply traffic-distribution monitoring and drift-based retraining.

Limitation: Thresholds and statistical tests are heterogeneous and are seldom empirically justified.

3.3.2 Performance-Driven Triggers

The second category of adaptation criteria uses model performance degradation as the trigger for adaptation. These methods involve performance measures such as detection rate, false positive rate, or classification confidence, which are continually or periodically monitored. When performance falls below predefined acceptable levels, the adaptation process is initiated.

Some studies note retraining or updating models as false-positive rates rise or detection accuracy decreases over time, especially in long-term IoT deployments. Performance-driven triggers are typically employed in incremental retraining and federated learning-based intrusion detection systems, where retraining is driven by performance degradation rather than continuous online updates.

Evidence-backed: Performance-based retraining is described in several included studies, particularly in incremental and federated learning settings.

Limitation: In most studies, explicit numeric thresholds are not provided, and trigger sensitivity is not justified, which limits reproducibility.

3.3.3 Resource-Aware Triggers

Resource-aware adaptation treats computational capacity, memory, and energy as primary triggers in IoT security systems. Some studies directly limit adaptation frequency or model complexity based on resource availability at IoT devices or edge nodes, such as deferring training to edge or cloud infrastructure when local resources are exceeded.

However, explicitly defined closed-loop trigger-action mechanisms determined by resource thresholds (e.g., using an energy budget to make retraining decisions) are less commonly implemented in practice. The majority of the literature discusses resource awareness qualitatively or evaluates resource consumption ex post but does not integrate resource metrics directly into adaptation logic.

Partially evidence-backed: Resource constraints are widely acknowledged and measured.

Conceptual synthesis: Explicit resource-threshold–driven adaptation loops remain largely conceptual and underexplored.

3.3.4 Summary of Evidence vs. Synthesis

In general, traffic-driven and performance-driven triggers have strong empirical support in the literature considered, whereas resource-aware triggers are either conceptual or only indirectly addressed. This discrepancy highlights one of the main research gaps: the absence of unified trigger-action models that jointly consider traffic evolution, detection reliability, and resource sustainability in real-time IoT settings.

Table 3: Evidence Mapping of Adaptation Triggers in Included Studies

Adaptation Trigger	**Explicit Trigger–Action Loop Reported**	**Metrics Used**	**Evidence Status**
Traffic shifts	Reported in multiple studies	Feature distributions, drift statistics	Empirical
Performance degradation	Reported in multiple studies	Accuracy, FPR, DR	Empirical
Resource thresholds	Rare	Energy, latency (mostly reported, not enforced)	Largely conceptual

4. Taxonomy of Machine Learning Techniques for IoT Security

The heterogeneity of Internet of Things (IoT) deployments, comprising immensely resource-constrained end devices, edge gateways, and cloud-fused infrastructure, has led to the deployment of a host of machine learning (ML)-based intrusion detection and security monitoring solutions. The available literature differs in learning algorithms, deployment architecture, computational overhead, and data representation, and comparison among studies is not always straightforward. To overcome this conceptual gap, a deployment-sensitive taxonomy of ML techniques for IoT security is presented, organized into architectural, algorithmic, compositional, and representational dimensions. The taxonomy provides a standard framework for comparing existing approaches in terms of detection capability and deployment viability.

4.1. Centralized and Distributed IoT Intrusion Detection Architectures

An ML-based intrusion detection system (IDS) for IoT can be categorized based on where data processing and learning occur. In a centralized IDS architecture, traffic information from multiple IoT devices is collected, and model training and inference are performed on a central or cloud server. This approach provides global visibility and supports computationally intensive models, but it introduces scalability bottlenecks, increased detection latency, and a higher risk of privacy leakage because raw data are aggregated [21]. Practical constraints may limit the suitability of centralized IDSs for latency- or privacy-sensitive IoT applications.

Distributed IDS architectures, on the other hand, perform learning and inference at edge devices or gateways, thereby minimizing communication overhead and enabling quicker responses. Federated and collaborative learning paradigms are examples of this pattern, as they facilitate decentralized model updating without raw-data sharing, which is more suited to large-scale, privacy-sensitive IoT deployments [22]. Nevertheless, distributed IDSs introduce challenges associated with system heterogeneity, synchronization, partial observability, and client drift, especially when traffic is non-IID [23]. This architectural distinction fundamentally influences the feasibility and performance of ML-based IoT security solutions.

4.2. Traditional Machine Learning vs. Deep Learning Approaches

From an algorithmic perspective, ML-based IoT security solutions can be classified into traditional machine learning and deep learning strategies. Traditional ML methods, including decision trees, support vector machines, k-nearest neighbors, and naive Bayes classifiers, rely on handcrafted features and relatively shallow models. These approaches are computationally efficient and explainable, and hence appealing for implementation on constrained IoT devices or edge nodes [24]. Deep learning (DL) methods, such as convolutional neural networks, recurrent neural networks, and autoencoders, automatically learn hierarchical representations from raw or minimally processed data. This capability has enabled improved detection performance across various IoT security settings, especially when dealing with large-scale or high-dimensional traffic data [25].

However, DL models are generally more computationally demanding and memory-intensive, which increases latency, energy consumption, and explainability challenges. As a result, the choice between traditional ML and DL methods introduces a trade-off between detection performance and deployment feasibility in IoT environments.

4.3. Hybrid and Ensemble Models

Hybrid and ensemble models aim to provide a compromise between expressiveness and robustness by integrating multiple learning methods. Hybrid models generally combine feature engineering with ML, or integrate conventional ML classifiers with DL-based feature extractors, to achieve improved generalization on heterogeneous IoT traffic while partially managing computational costs [26]. These methods are particularly useful in complex or variable traffic environments.

Ensemble learning techniques, including bagging, boosting, and stacking, combine predictions from multiple base learners to reduce variance and improve robustness to noise and class imbalance. Ensemble-based IDSs can be more resilient to evolving attacks than single-model IDSs in IoT security settings [27]. However, ensemble schemes increase inference latency, memory usage, and management complexity, which may restrict their use in resource-constrained IoT deployments. Their inclusion in this taxonomy highlights the need to consider both performance improvements and operational costs.

4.4. Lightweight Machine Learning for Resource-Constrained IoT Devices

A large percentage of IoT devices operate under stringent memory, processing power, and energy constraints, necessitating the development of lightweight ML methods. Lightweight IDS designs use simplified model architectures, reduced feature sets, or model compression to minimize resource consumption while maintaining acceptable detection performance [28].

Recent studies show that carefully designed lightweight DL models (e.g., shallow neural networks or optimized autoencoders) can achieve competitive performance while significantly reducing computational cost [29]. However, the effectiveness of these methods depends heavily on model configuration and workload characteristics. Resource-conscious learning is therefore essential for achieving operational sustainability in real-world low-power IoT environments, where architectural security considerations remain critical [30].

4.5. Graph-Based and Traffic Flow Modeling Techniques

In addition to flat feature-vector representations, graph-based and traffic-flow modeling methods have proven effective for detecting IoT intrusions. These approaches model network entities and their interactions as a graph, allowing structural and relational information to be captured that may be obscured in traditional representations. In particular, graph neural networks (GNNs) have shown potential for identifying coordinated and distributed attacks by exploiting topological and temporal dependencies [22].

In IoT environments, communication among devices, gateways, and services is analyzed using graph-based IDS approaches to identify anomalies at the system level rather than at isolated nodes. Although recent methods demonstrate improved detection of advanced attack scenarios, graph-based approaches introduce challenges, including graph construction overhead, scalability limitations, and real-time inference constraints [31]. Doctoral research also emphasizes that, despite the representational advantages of traffic graph modeling, its practical implementation in large-scale IoT systems remains challenging [32].

This taxonomy section indicates that no universal ML technique is optimal for IoT security. Detection effectiveness cannot be evaluated independently of deployment architecture, resource constraints, and data representation. Consequently, IoT intrusion detection should be treated as a system-level design challenge, where learning algorithms are selected not only for accuracy but also for scalability, adaptability, and robustness in real-world environments.

All references listed in Table 4, including static baseline comparators, are drawn exclusively from the 38 studies included in the final systematic review (see Figure 1). “Representative” indicates illustrative examples selected to exemplify each taxonomy category, not an exhaustive listing. Background surveys and non-included studies are cited only in the narrative text and are not used as evidentiary support in this table.

Table 4: Taxonomy of Machine Learning Techniques for IoT Security (with Representative Included Studies)

Category	**Learning Paradigm**	**Typical Algorithms / Models**	**Deployment Architecture**	**Key Advantages**	**Key Limitations**	**Representative References**
Architectural	Centralized IDS	SVM, Random Forest, Deep Neural Networks	Cloud / Central Server	Global visibility, high detection accuracy, simplified model management	High latency, scalability issues, privacy concerns	[21, 24]
Architectural	Distributed / Federated IDS	Federated averaging, collaborative ML, edge-based DL	Edge / Fog / Federated	Privacy preservation, reduced communication overhead, improved scalability	Client drift, system heterogeneity, synchronization complexity	[22, 23]
Algorithmic	Traditional ML	Decision Trees, k-NN, Naïve Bayes, SVM	Edge / Gateway	Low computational cost, interpretability, suitability for constrained devices	Limited representation power, manual feature engineering	[24]
Algorithmic	Deep Learning	CNN, RNN, LSTM, Autoencoders	Cloud / Edge	Automatic feature extraction, strong performance on complex traffic	High resource consumption, limited explainability	[25]
Model Composition	Hybrid Models	Feature engineering + DL, ML–DL fusion	Edge / Cloud	Balanced performance and complexity, adaptability to heterogeneous traffic	Increased system complexity, tuning overhead	[26]
Model Composition	Ensemble Models	Bagging, Boosting, Stacking	Edge / Cloud	Improved robustness, reduced variance, resilience to noise	Higher inference latency, model management overhead	[27]
Resource Awareness	Lightweight ML	Shallow NN, compressed DL, reduced-feature ML	IoT Device / Edge	Low latency, reduced energy consumption, real-time feasibility	Potential accuracy degradation, limited expressiveness	[28, 29]
Representation	Graph-Based Models	Graph Neural Networks (GNNs), Flow Graphs	Edge / Cloud	Captures relational and structural patterns, effective for coordinated attacks	Graph construction overhead, scalability challenges	[22, 31, 32]

5. Thematic Review of State-of-the-Art Studies

Recent studies on machine learning-based IoT security have increasingly shifted toward adaptive, privacy-aware, and distributed learning paradigms rather than static intrusion detection approaches. However, much of the current literature remains focused on algorithmic performance in controlled experimental settings, with limited critical discussion of real-time viability, system overhead, and deployment constraints. This section addresses these limitations by summarizing cutting-edge research across four prevailing themes: hybrid and ensemble learning, federated learning, feature selection and optimization, and deep learning-based intrusion detection, while explicitly considering their applicability to real-time IoT security implementation. Instead of focusing solely on detection accuracy, the analysis emphasizes latency, scalability, adaptability to evolving traffic, and operational cost.

5.1. Hybrid and Ensemble Machine Learning Approaches for DDoS Detection

Hybrid and ensemble learning methods are frequently recommended for detecting distributed denial-of-service (DDoS) attacks in IoT networks due to their robustness in monitoring heterogeneous traffic patterns. Hybrid models combine complementary techniques, such as feature engineering with machine learning classifiers, traditional ML with deep learning feature extractors, or ensemble methods that integrate multiple base learners to reduce variance and improve generalization.

Empirical research consistently reports higher detection rates compared to single-model baselines, especially for heterogeneous or noisy IoT traffic [33, 34]. Nevertheless, these performance gains involve significant deployment costs. Ensemble inference increases latency, memory overhead, and energy consumption, which are constrained in edge- or gateway-level deployments. Consequently, models that perform strongly in offline experiments may exceed real-time processing capabilities in operational IoT environments [35].

Importantly, most hybrid and ensemble studies evaluate performance using fixed datasets and offline testing pipelines, providing limited insight into long-term behavior under evolving traffic or adversarial adaptation. This reveals a fundamental trade-off: hybrid and ensemble models enhance detection robustness, but their structural complexity may hinder real-time execution. Without explicit analysis of latency, throughput, and energy consumption, it is difficult to translate reported performance gains into deployable IoT security solutions.

5.2. Federated Learning for Privacy-Preserving IoT Security

Federated learning (FL) has emerged as a prominent framework for privacy-preserving IoT intrusion detection, enabling collaborative model training without centralized raw data aggregation. FL-based systems retain traffic data on local devices or edge nodes, thereby supporting regulatory compliance and privacy guarantees while improving scalability in distributed environments.

Recent research indicates that FL can reduce data exposure and improve scalability [36]. However, real-time deployment challenges remain significant. Non-IID traffic distributions across IoT devices often lead to client drift and unstable model convergence, degrading detection performance over time. These effects are amplified in highly heterogeneous IoT environments where device behavior, workloads, and network conditions vary considerably.

Real-time operation is further constrained by communication overhead. Frequent model updates may overload limited network bandwidth and edge infrastructure, resulting in unacceptable latency. Adaptive aggregation and resource-aware scheduling have been proposed to mitigate these challenges [23], but they introduce additional trade-offs among convergence speed, detection accuracy, and system complexity. Furthermore, although FL is designed to preserve privacy, FL-based IDSs remain vulnerable to inference attacks, such as gradient leakage and model inversion, raising concerns about their suitability for sensitive IoT applications [37].

Federated learning represents a promising architectural direction for IoT security; however, its practical deployment in real-time environments remains constrained by communication cost, convergence instability, and unresolved privacy risks [38].

5.3. Feature Selection and Optimization-Based Approaches

Feature selection and optimization-based methods aim to improve IoT intrusion detection performance by reducing feature dimensionality while preserving discriminative power. These techniques are particularly relevant for real-time IoT implementations, where computational efficiency directly influences latency and energy consumption. Metaheuristic optimization methods have been shown to identify compact feature subsets that enable lower training and inference cost with minimal loss of accuracy [39, 40].

The Arithmetic Optimization Algorithm (AOA), for example, has been reported to provide balanced exploration and exploitation capabilities with competitive detection performance [41]. Despite these advantages, optimization-based IDSs may lack robustness under varying IoT traffic conditions. Their performance is often sensitive to dataset characteristics, parameter configuration, and attack distribution, raising concerns regarding generalizability. Moreover, most evaluations are conducted on benchmark datasets, offering limited evidence of robustness under shifting attack patterns. Foundational studies in optimization emphasize that efficiency improvements observed under controlled experimental conditions do not guarantee reliable real-time deployment without system-level validation [42].

5.4. Deep Learning–Based IoT Intrusion Detection

Deep learning (DL) methods, such as convolutional neural networks (CNNs), recurrent neural networks (RNNs), and long short-term memory (LSTM) models, have been widely studied in IoT intrusion detection because of their ability to capture complex temporal and spatial traffic patterns. Temporal DL architectures have demonstrated strong performance in modeling sequential network traffic dependencies [43].

However, real-time deployment of DL-based IDSs remains highly constrained. Many models have substantial computational and memory requirements, resulting in high inference latency and energy consumption, which limits their suitability for single-board IoT devices or edge nodes. Moreover, deep models are often overfitted to benchmark datasets; their performance degrades when exposed to unknown or evolving traffic, reflecting limited robustness in operational settings [44].

Explainability further restricts real-world applicability. Opaque decision-making in DL complicates incident response and reduces trust among IoT operators. Although explainable DL methods have demonstrated potential to improve transparency without significantly degrading detection performance [45], their scalability and real-time feasibility remain insufficiently explored. Doctoral research emphasizes that unless latency, interpretability, and robustness are addressed jointly, DL-based IDSs may remain laboratory solutions rather than practically deployable security mechanisms [46].

6. Evaluation of Datasets and Performance Metrics in IoT Security Research

The quality of datasets used to train and evaluate machine learning-based IoT security solutions, along with the performance metrics used to report results, fundamentally constrains their reliability. Although methodological sophistication has increased, much of the literature still emphasizes high detection accuracy without critically examining dataset representativeness, temporal validity, and the operational implications of false alarms, latency, and resource consumption. This section provides a deployment-focused critical analysis of commonly used IoT security datasets and evaluation metrics, and discusses how dataset bias and metric selection may distort claims of effectiveness and real-time viability.

6.1. Commonly Used IoT Security Datasets

Empirical evaluation of IoT intrusion detection studies is dominated by several benchmark datasets, including IoT-23 and CIC-IoT-2023, as well as smaller, domain-specific datasets designed for healthcare IoT systems. Although these datasets facilitate reproducible experimentation, they are not always representative of real-world IoT environments.

The IoT-23 dataset contains labeled benign and botnet traffic collected from real IoT devices, providing insight into malware-driven attack behavior under controlled conditions [47]. However, its experimental configuration covers a limited set of device types, workloads, and attack evolutions, restricting its ability to model large-scale heterogeneous deployments. CIC-IoT-2023 introduces more recent attack scenarios and broader traffic diversity, but remains constrained by scripted attack execution and laboratory network conditions [48].

Specialized healthcare IoT datasets present additional challenges. Although realistic in terms of protocol usage and privacy constraints, they are often small, curated, and difficult to generalize beyond their specific application context [49]. As a result, models trained exclusively on such datasets may exhibit optimistic performance that does not generalize to uncontrolled, heterogeneous IoT environments.

Reliance on static benchmark datasets also obscures the effects of traffic evolution, device churn, and adversarial adaptation, leading to evaluation outcomes that may overstate real-world performance.

6.2. Dataset Quality Issues and Bias

A pervasive limitation across IoT security datasets is class imbalance, where attack traffic may dominate the dataset or be severely underrepresented relative to benign samples. In such cases, accuracy-based evaluation becomes misleading, as models may achieve high overall accuracy while performing poorly on minority attack classes that are operationally critical [50]. This issue is particularly significant in IoT settings, where infrequent but high-impact attacks must be reliably detected.

Another major concern is the reliance on synthetic or simulated traffic. Although synthetic datasets are easier to generate and label, they often lack the variability, noise, and unpredictability of real IoT networks. Empirical studies show that systems trained on artificial data frequently fail under live traffic conditions, revealing discrepancies between test and operational performance [51].

Dataset aging further undermines evaluation validity. Attack patterns, communication protocols, and device behaviors evolve rapidly, rendering many benchmark datasets obsolete. Aging datasets have been shown to degrade the performance of models when confronted with contemporary attack variants, highlighting the risks of drawing long-term conclusions from static benchmarks [52].

6.3. Implications for Performance Metrics Beyond Accuracy

Dataset limitations directly influence the reliability of reported performance metrics. Although accuracy remains the most commonly reported metric, it is insufficient for evaluating intrusion detection in imbalanced and evolving IoT environments. Operationally meaningful metrics, such as false positive rate (FPR) and false negative rate (FNR), provide clearer insight into system behavior, since excessive false alarms disrupt operations, whereas missed attacks compromise security.

Robustness-oriented measures, including the Matthews correlation coefficient and balanced accuracy, have been recommended for more reliable evaluation under class imbalance [53]. Additionally, effective assessment requires cross-dataset validation and evaluation across diverse attack scenarios rather than reliance on single-dataset benchmarking.

Beyond detection correctness, real-time IoT security demands evaluation of latency, throughput, and resource overhead. IDSs operating at the edge or fog layer must satisfy strict timing constraints; otherwise, inference delays may negate their practical value. Empirical studies indicate that models achieving high accuracy may still be impractical due to excessive inference latency or energy consumption [54]. Frameworks emphasizing comprehensive evaluation across detection quality, robustness, latency, and energy efficiency have been proposed [55], but remain underutilized in current IoT security research.

6.4. Descriptive Quantitative Summary of Evaluation Practices

To complement the qualitative analysis, a descriptive quantitative synthesis was conducted across the 38 included studies to summarize evaluation practices related to dataset quality, performance metrics, and robustness assessment. Table counts were derived from the standardized data-extraction protocol described in Section 2.6.

Table 5: Frequency of Evaluation Practices Across Included Studies ($n = 38$)

Evaluation Aspect	**Number of Studies**	**Percentage (%)**
Report accuracy / detection rate	38	100
Report precision / recall / F1	29	76
Report false positive rate (FPR)	16	42
Report false negative rate (FNR)	11	29
Report latency or inference time	9	24
Report energy or resource usage	6	16
Address class imbalance explicitly	14	37
Perform cross-dataset testing	5	13
Evaluate dataset aging / temporal validity	4	11

The results indicate a strong emphasis on accuracy-centric evaluation, with limited reporting of operational metrics such as latency, energy consumption, and robustness to dataset aging. Cross-dataset validation and temporal robustness remain underreported, reinforcing concerns regarding real-world generalizability.

7. Real-Time Deployment Challenges in IoT Environments

Despite the high performance of reported intrusion detection systems (IDSs) using machine learning models in experimental studies, their extrapolation to real-time deployments in Internet of Things (IoT) environments remains limited. A major limitation in the literature is the emphasis on detection accuracy at the expense of system-level considerations, including inference latency, energy usage, retraining cost, and deployment architecture. These constraints ultimately determine whether an IDS can operate effectively in real IoT networks rather than in controlled laboratory environments [56]. This section critically reviews the main deployment issues that hinder real-time adoption of ML-based IoT security solutions.

7.1. Edge vs. Cloud vs. Fog-Based Deployment

The deployment layer fundamentally determines real-time performance of intrusion detection. Cloud-based IDS deployments enable centralized management and support computationally intensive models, but introduce communication latency and dependence on continuous connectivity. Such latency may render detection ineffective in time-sensitive IoT applications, particularly in geographically dispersed or bandwidth-limited environments [57].

Edge-based IDS deployment brings traffic processing closer to IoT devices, minimizing inference latency and network overhead and enabling faster responses. However, edge nodes have limited computational and energy resources, which significantly constrain the complexity of deployable models [54]. As a result, many high-accuracy models proposed in the literature cannot be executed at the edge without substantial simplification.

Fog computing aims to balance these trade-offs by distributing processing across intermediate nodes. Although fog architectures can reduce latency compared to cloud-only deployments, they introduce additional coordination, synchronization, and management complexity [58]. Notably, many IDS studies do not clearly specify the assumed deployment layer, making it difficult to determine whether reported performance can be achieved in real-time operational contexts.

7.2. Resource Constraints of IoT Devices

Resource scarcity represents one of the most immediate barriers to real-time IoT security implementation. IoT devices and gateways are tightly constrained by memory capacity, processing power, and energy supply. Even moderately complex deep learning models may exceed these limits, resulting in unacceptable inference delays or rapid battery depletion [59].

Empirical studies on energy-efficient ML inference consistently show that improvements in detection accuracy obtained through increased model complexity come at the cost of higher energy consumption and reduced system lifespan [54]. Therefore, resource efficiency should not be treated as a secondary optimization objective but rather as a primary design requirement, especially for adaptive IDSs that require continuous monitoring and periodic model updates.

7.3. Retraining Cost and Model Update Frequency

Periodic model updates are necessary to maintain detection performance in dynamic IoT environments where traffic patterns and attack behaviors evolve. However, retraining imposes significant computational and communication costs, particularly in large-scale or distributed deployments. Excessive retraining may overload system resources, whereas infrequent updates may increase vulnerability to concept drift.

Federated and distributed learning systems reduce raw data transfer but introduce challenges related to synchronization, device heterogeneity, and convergence stability [60]. Despite these trade-offs, many studies assume fixed retraining schedules and do not evaluate their operational cost or impact on real-time performance [55]. This gap between adaptive learning theory and deployment realities represents a persistent weakness in current IoT security research.

7.4. Edge–Cloud Orchestration Challenges

Effective real-time IoT security increasingly depends on coordinated edge–cloud orchestration, where decisions regarding model placement, inference execution, aggregation, and retraining are dynamically managed across system layers. Poor orchestration may lead to redundant computation, delayed response, and inconsistent detection performance across distributed nodes [56].

In practice, orchestration must account for device heterogeneity, variable workloads, intermittent connectivity, and evolving threat conditions. However, many ML-based IDS architectures treat edge and cloud components as independent entities rather than as elements of a unified security pipeline. Systems research emphasizes the importance of deployment-aware orchestration to achieve scalable and reliable real-time analytics in IoT environments [61]. The limited focus on orchestration strategies in current IDS proposals constitutes a significant barrier to operational adoption.

7.5. Deployment Evaluation Framework and Evidence Mapping

To systematically assess deployment realism, a deployment evaluation framework was applied to the 38 included studies. Each study was annotated against four measurable criteria derived from edge-IoT operational constraints.

Deployment Criteria Defined

- D1: Deployment layer specified (Edge / Fog / Cloud / Not specified)
- D2: Latency budget discussed or measured
- D3: Resource constraints evaluated (CPU, memory, or energy)
- D4: Real-time feasibility explicitly claimed or validated

Table 6: Deployment Criteria Coverage Across Included Studies ($n = 38$)

Deployment Criterion	Studies Meeting Criterion	Percentage (%)
D1: Deployment layer specified	17	45
D2: Latency budget analyzed	9	24
D3: Resource constraints evaluated	8	21
D4: Real-time feasibility validated	6	16

Although many studies acknowledge deployment constraints conceptually, fewer than one-quarter provide measurable latency or resource evaluations. Explicit mapping of IDS models to edge- or device-level deployment remains limited, supporting concerns regarding the gap between experimental performance and operational feasibility.

8. Adversarial and Privacy Risks: Background vs. Evidence in Reviewed Studies

With machine learning-based intrusion detection systems (IDSs) increasingly integrated into IoT security architectures, the security of the learning models themselves has become an important and often underexamined consideration. Much of the literature assumes benign training and inference conditions and does not account for the reality that ML models deployed in IoT environments are directly vulnerable to adversarial manipulation, data poisoning, evasion, or privacy breaches. These threats undermine reported detection performance and introduce significant risks in real-world deployments.

8.1. Adversarial Attacks on ML-Based IoT IDS

Adversarial attacks aim to compromise learning systems by degrading detection performance or inducing systematic misclassification. In IoT environments, such attacks are particularly effective due to distributed data sources, limited supervision, and automated decision-making processes.

Poisoning attacks target the training or update phase by injecting malicious or mislabeled data into the learning pipeline. In adaptive IoT IDSs, where models are periodically retrained or updated using live traffic, poisoning can progressively bias decision boundaries, resulting in sustained misclassification of malicious traffic. This threat is amplified in distributed and federated learning environments, where compromised devices may contribute poisoned updates without centralized validation [62]. In contrast, evasion attacks manipulate input traffic during inference to avoid detection. Attackers exploit weaknesses in feature representations or decision thresholds by crafting traffic patterns that appear benign. Empirical evidence indicates that IDS models trained on outdated datasets are particularly susceptible to evasion strategies, raising concerns about the reliability of accuracy-centric evaluations [4].

8.2. Privacy Leakage and Inference Risks

Model inversion and membership inference attacks enable adversaries to determine whether specific data samples were included in training or to infer sensitive attributes from model outputs. Even federated learning, commonly described as privacy-preserving, remains vulnerable to such attacks unless appropriate safeguards are implemented [37]. These risks are further amplified in heterogeneous IoT environments, where partially trusted participants constrain the practical privacy guarantees of collaborative learning frameworks [38].

8.3. Defense Strategies and Practical Limitations

Various defense mechanisms have been proposed to mitigate adversarial and privacy risks, including adversarial training, robust optimization, secure aggregation, and privacy-preserving update mechanisms. However, many of these defenses introduce substantial computational and communication overhead, conflicting with the resource limitations of IoT devices.

Adversarial training enhances robustness against evasion but requires repeated retraining with more complex models, which may be impractical in real-time IoT systems [62]. Similarly, cryptographic approaches for securing federated learning updates reduce privacy leakage but significantly increase communication and computation costs, limiting scalability in large IoT deployments [23]. Consequently, many proposed defenses remain challenging to implement outside controlled experimental environments.

8.4. Implications for Adaptive IoT Security

The presence of adversarial and privacy risks fundamentally alters the design considerations of adaptive IoT security systems. As adaptive learning mechanisms become more responsive to evolving threats, they also expand the attack surface to include poisoning and inference attacks. At the same time, real-time deployment constraints restrict the feasibility of computationally intensive defense strategies. These observations indicate that adaptive ML systems for IoT security should be designed with robustness, privacy, and efficiency addressed jointly rather than treated as competing objectives. Without integrating adversarial resilience and privacy protection into the core design of adaptive IDSs, machine learning may become a vulnerability rather than a solution within IoT security architectures.

8.5. Evidence of Adversarial Evaluation in Included Studies

Although adversarial machine learning threats such as poisoning, evasion, and privacy inference are well established in the broader literature, their empirical evaluation within the included IoT security studies remains limited.

Table 7: Adversarial Evaluation Coverage in Included Studies ($n = 38$)

Adversarial Aspect Evaluated	Number of Studies	Percentage (%)
Evasion attacks tested	7	18
Poisoning attacks tested	3	8
Privacy leakage / inference evaluated	2	5
No adversarial evaluation	30	79

The majority of included studies discuss adversarial risks only at a conceptual level. Explicit adversarial testing—particularly poisoning resistance and privacy leakage assessment—remains uncommon, indicating a significant evidence gap between theoretical threat models and experimental validation.

9. Comparative Analysis of Existing Studies

Although the primary focus of this review is on adaptive machine learning approaches, several static (non-adaptive) intrusion detection studies were intentionally retained in Table 8 as baseline comparators. These studies represent widely used, high-performing static ML and DL models that are frequently referenced or extended by adaptive approaches. Their inclusion enables clearer comparison of adaptation benefits, deployment trade-offs, and performance limitations relative to non-adaptive baselines.

Static-only studies are therefore not treated as evidence of adaptation mechanisms, but as contextual benchmarks against which adaptive methods are evaluated. This section presents a structured comparative analysis of representative machine learning-based IoT security studies. Instead of reporting accuracy values in isolation, the comparison emphasizes adaptation mechanisms, dataset quality, deployment feasibility, resource consumption, and documented limitations. The revised table supports a comprehensive evaluation of current techniques and highlights research gaps that inform future directions.

Table 8: Comparative Analysis of Adaptive and Static Baseline ML-Based IoT Security Studies

Study	**Target Threat / Domain**	**ML Technique**	**Adaptation Type**	**Dataset(s)**	**Dataset Quality Issues**	**Reported Metrics**	**Resource / Deployment Considerations**	**Key Limitations**
Mahdi et al. (2024)	DDoS in IoT networks	Traditional ML (RF, SVM)	Static	CIC-DDoS2019	Synthetic traffic; class imbalance	Accuracy, DR	Evaluated offline; no latency analysis	No real-time validation; static model
Bhayo et al. (2022)	IoT intrusion detection	Hybrid ML (FS + classifier)	Incremental retraining	CIC-IDS-based IoT data	Limited device diversity	Accuracy, F1-score	Moderate complexity; edge feasibility not analyzed	Retraining cost not evaluated
Abdallah et al. (2022)	DDoS detection	Hybrid ML	Static	Custom IoT traffic	Synthetic environment	Accuracy, precision	No deployment discussion	Scalability unclear
Verma & Ranga (2023)	IoT IDS	Ensemble learning	Static	Benchmark IoT datasets	Dataset aging ignored	Accuracy, recall	Increased inference overhead	High complexity for constrained devices
Kim et al. (2022)	IoT intrusion detection	Federated learning	Federated adaptation	Distributed IoT traffic	Non-IID data	Accuracy, convergence	Communication cost analyzed	Client drift affects stability
Ioannou et al. (2024)	Medical IoT security	Federated ML	Federated + adaptive	Healthcare IoT dataset	Domain-specific bias	Accuracy, energy usage	Edge–cloud deployment considered	Privacy leakage not fully addressed
Hassanien et al. (2024)	IoT IDS	AOA-based feature selection + ML	Static	Benchmark IDS datasets	Imbalance partially addressed	Accuracy, latency	Reduced feature cost	Generalization not tested
Roy et al. (2022)	IoT IDS	Deep learning (LSTM)	Static	IoT traffic traces	Overfitting risk	Accuracy, loss	High training cost	Poor explainability
Zhang et al. (2023)	IoT IDS	Explainable DL	Static	Benchmark IoT datasets	Limited attack diversity	Accuracy, explanation fidelity	Not evaluated under real-time constraints	Scalability unclear
Arisdakessian et al. (2024)	IoT intrusion detection	Deep learning	Drift-aware retraining	IoT datasets	Dataset aging discussed	Accuracy, robustness	Training overhead high	Resource constraints ignored

10. Research Challenges and Future Directions

Despite substantial progress in machine learning for IoT security, the preceding analysis reveals persistent research challenges that limit long-term robustness and real-world applicability. Addressing these issues requires shifting from accuracy-centric evaluation toward resilient, adaptive, and deployment-aware security architectures.

10.1. Lightweight and Resource-Aware Adaptive Intrusion Detection

A major challenge is the design of adaptive intrusion detection systems (IDSs) that operate effectively under strict resource constraints of IoT devices and edge nodes. Although adaptive and deep learning models can enhance detection capability, they often incur high computational and energy cost. Future research should prioritize lightweight adaptation strategies, including selective retraining, feature-efficient modeling, and hardware-aware optimization, to support sustainable real-time deployment across heterogeneous IoT platforms.

10.2. Continual, Self-Supervised, and Drift-Aware Learning

IoT environments are characterized by evolving traffic patterns and concept drift, rendering static training paradigms insufficient. While incremental and federated learning have received attention, continual and self-supervised learning remain underexplored in IoT security. Future work should focus on drift-aware mechanisms capable of autonomously detecting distributional changes and triggering adaptation with minimal reliance on labeled data and limited retraining overhead.

10.3. Robustness Against Adversarial and Poisoning Attacks

The vulnerability of ML models themselves constitutes a critical weakness in IoT security architectures. As discussed in Section 8, adaptive and federated systems expand the attack surface for poisoning, evasion, and inference attacks. Future research should aim to develop robustness-by-design adaptive models, integrating adversarial resilience during training, updating, and deployment phases rather than treating it as an afterthought. Achieving robustness without incurring prohibitive computational cost remains an open challenge.

11. Conclusions

This review provides a systematic and critical analysis of adaptive machine learning methods for securing IoT environments, addressing limitations identified in previous surveys. Rather than cataloging algorithms or presenting detection accuracy as an isolated metric, this study emphasizes adaptation mechanisms, dataset quality, evaluation metrics, real-time feasibility, and adversarial robustness.

Using a unified taxonomy and thematic analysis, the review demonstrates that while machine learning offers powerful capabilities for IoT intrusion detection, high experimental accuracy does not guarantee real-world performance. Operational viability is strongly influenced by dataset bias, insufficient performance metrics, resource limitations, and deployment architecture constraints. Moreover, emerging adaptive and federated learning paradigms introduce new privacy and adversarial risks that must be addressed comprehensively.

Adaptive machine learning is essential for protecting dynamic IoT environments; however, effective implementation requires joint consideration of robustness, efficiency, interpretability, and system-level deployment constraints. This review synthesizes current evidence and outlines key research challenges and future directions to guide the development of scalable, reliable, and practically deployable IoT security solutions.

Declaration of Competing Interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding Declaration

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Ethics Approval and Consent to Participate

This article is a review study based exclusively on previously published literature. It does not involve human participants, animals, or confidential data requiring ethical approval or informed consent.

Data Availability Statement

Data sharing is not applicable to this article as no new data were created or analyzed. All information synthesized in this review is derived from publicly available published studies cited in the reference list.

AI Usage Disclosure

The authors used an AI-based language assistance tool for minor grammatical refinement and formatting support. The scientific content, analysis, interpretation, and conclusions were independently developed, verified, and approved by the authors.

Author Contributions

Jayashri Jayesh Patil: Conceptualization, Methodology, Literature Search, Data Curation, Formal Analysis, Writing – Original Draft, Visualization; **Dr. Ramkumar Solanki**: Supervision, Validation, Writing – Review and Editing, Project Administration.

References

[1] M. Gelgi and M. Çelik, "A systematic literature review of IoT botnet DDoS attacks," *Sensors*, vol. 24, no. 3, pp. 1–29, 2024.

[2] F. Alwahedi, A. Aldhaheri, M. A. Ferrag, A. Battah, and N. Tihanyi, "Machine learning techniques for IoT security: Current research and future vision with generative AI and large language models," *Internet of Things and Cyber-Physical Systems*, vol. 4, pp. 167–185, 2024.

[3] R. Liu, J. Shi, X. Chen, and C. Lu, "Network anomaly detection and security defense technology based on machine learning: A review," *Computers & Electrical Engineering*, vol. 119, p. 109581, 2024.

[4] M. M. Rahman, M. S. Hossain, and M. M. Hassan, "Intrusion detection systems for IoT networks: Recent advances and challenges," *Journal of Network and Computer Applications*, vol. 226, p. 103927, 2024.

[5] C. Ni and S. Li, "Machine learning-enabled industrial IoT security: Challenges, trends, and solutions," *Journal of Industrial Information Integration*, vol. 38, p. 100549, 2024.

[6] E. C. Pinto Neto, S. Dadkhah, S. Sadeghi, H. Molyneaux, and A. A. Ghorbani, "A review of machine learning–based IoT security in healthcare: A dataset perspective," *Computer Communications*, vol. 213, pp. 61–77, 2024.

[7] F. Hinder, A. Artelt, B. Hammer, and M. Biehl, "One or two things we know about concept drift," *Frontiers in Artificial Intelligence*, vol. 7, p. 1296484, 2024.

[8] A. L. Suárez-Cetrulo, A. Bifet, and J. Calvo-Zaragoza, "A survey on machine learning for recurring concept drifting data streams," *Applied Soft Computing*, vol. 132, p. 109890, 2023.

[9] Z. Mahdi, N. Abdalhussien, N. Mahmood, and R. Zaki, "Detection of real-time distributed denial-of-service (DDoS) attacks on internet of things (IoT) networks using machine learning algorithms," *Computers, Materials & Continua*, vol. 80, no. 2, pp. 2139–2159, 2024.

[10] I. Ioannou, P. Nagaradjane, P. Angin, P. Balasubramanian, K. J. Kavitha, P. Murugan, and V. Vassiliou, "GEMLIDS-MIOT: A green effective machine learning intrusion detection system based on federated learning for medical IoT network security hardening," *Computer Communications*, vol. 218, pp. 209–239, 2024.

[11] I. Khamassi, M. Sayed-Mouchaweh, M. Hammami, and K. Ghédira, "Concept drift detection and adaptation with incremental learning: A survey," *IEEE Transactions on Knowledge and Data Engineering*, vol. 34, no. 4, pp. 1739–1756, 2022.

[12] J. Lu, A. Liu, F. Dong, F. Gu, J. Gama, and G. Zhang, "Learning under concept drift: A review," *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 2, pp. 1242–1260, 2023.

[13] J. Gama, I. Žliobaitė, A. Bifet, M. Pechenizkiy, and A. Bouchachia, "A survey on concept drift adaptation," *ACM Computing Surveys*, vol. 46, p. 44, Apr. 2014.

[14] Y. Ren, *Interactive Causality Enabled Adaptive Machine Learning.* PhD thesis, UC Irvine, 2023. ProQuest ID: Ren_uci_0030D_18546. Merritt ID: ark:/13030/m5kn0cfk. Retrieved from https://escholarship.org/uc/item/3dj43270.

[15] A. Bifet, R. Gavalda, and G. Holmes, "Machine learning for evolving data streams," *ACM Computing Surveys*, vol. 54, no. 8, pp. 1–36, 2022.

[16] P. Kairouz, H. B. McMahan, *et al.*, "Advances and open problems in federated learning," *Foundations and Trends® in Machine Learning*, vol. 14, no. 1–2, pp. 1–210, 2023.

[17] J. Zhang, B. Chen, and Y. Wang, "Adaptive federated learning for edge-enabled IoT systems," *IEEE Transactions on Mobile Computing*, 2024.

[18] L. Jing and Y. Tian, "Self-supervised learning for representation learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 4, pp. 4143–4166, 2023.

[19] Z. Zhang and M. R. Sabuncu, "Generalized cross entropy loss for training deep neural networks with noisy labels," *IEEE Transactions on Neural Networks and Learning Systems*, 2023.

[20] V. Losing, B. Hammer, and H. Wersing, "Choosing the best algorithm for an evolving problem," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 1, pp. 1–15, 2022.

[21] V. Mothukuri, S. Pouriyeh, A. Dehghantanha, R. M. Parizi, G. Srivastava, and M. Shafiq, "A survey on security and privacy of federated learning," *Future Generation Computer Systems*, vol. 115, pp. 619–640, 2022.

[22] T. T. Nguyen, V. J. Reddi, and K. R. Chowdhury, "Graph neural networks for network intrusion detection: A survey," *IEEE Communications Surveys & Tutorials*, vol. 24, no. 2, pp. 1241–1267, 2022.

[23] T. Li, S. Hu, A. Beirami, and V. Smith, "Fair resource allocation in federated learning," *Proceedings of the IEEE*, vol. 111, no. 3, pp. 352–370, 2023.

[24] I. Ullah, Q. H. Mahmoud, and W. Alasmary, "A comparative analysis of machine learning and deep learning approaches for intrusion detection in internet of things networks," *IEEE Access*, vol. 10, pp. 100456–100474, 2022.

[25] M. A. Ferrag, L. Shu, H. Djallel, and L. Maglaras, "Deep learning-based intrusion detection for internet of things: A comprehensive survey," *Computer Networks*, vol. 230, p. 109806, 2023.

[26] J. Bhayo, I. A. Hameed, and A. Ahmed, "Hybrid machine learning models for intrusion detection in internet of things networks," *Future Generation Computer Systems*, vol. 129, pp. 63–78, 2022.

[27] A. Verma and V. Ranga, "Ensemble learning based intrusion detection systems for internet of things," *Journal of Network and Computer Applications*, vol. 209, p. 103531, 2023.

[28] S. M. Alqahtani, A. Alzahrani, and H. Aljuaid, "Lightweight intrusion detection for resource-constrained internet of things devices," *IEEE Internet of Things Journal*, vol. 10, no. 9, pp. 7862–7875, 2023.

[29] N. Shone, T. N. Ngoc, V. D. Phai, and Q. Shi, "A lightweight deep learning approach for intrusion detection in internet of things environments," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 6, no. 3, pp. 552–565, 2022.

[30] S. Raza, L. Wallgren, and T. Voigt, *Security and Privacy in Low-Power Internet of Things Systems.* Cambridge University Press, 2023.

[31] Y. Zhou, G. Cheng, and X. Xu, "Graph-based intrusion detection for internet of things networks using graph neural networks," *IEEE Internet of Things Journal*, vol. 10, no. 15, pp. 13489–13502, 2023.

[32] Y. Chen, *Graph-based traffic modeling for intrusion detection in Internet of Things networks.* PhD thesis, Tsinghua University, 2024. (Doctoral dissertation).

[33] M. Abdallah, H. Hijazi, and A. Awad, "Hybrid machine learning models for DDoS detection in internet of things networks," *IEEE Systems Journal*, vol. 16, no. 4, pp. 6432–6443, 2022.

[34] A. S. Alqahtani and M. A. Babar, "Ensemble-based intrusion detection for DDoS attacks in IoT environments," *Computers & Security*, vol. 123, p. 102951, 2023.

[35] M. Hassan, *Hybrid and ensemble learning techniques for distributed denial-of-service detection in IoT networks.* PhD thesis, University of Manchester, 2023. (Doctoral dissertation).

[36] Y. Kim, J. Park, and M. Bennis, "Federated intrusion detection in IoT networks under non-IID data," *IEEE Transactions on Network and Service Management*, vol. 19, no. 3, pp. 3194–3208, 2022.

[37] M. Nasr, R. Shokri, and A. Houmansadr, "Comprehensive privacy analysis of federated learning," in *Proceedings of the IEEE Symposium on Security and Privacy (SP)*, IEEE, 2022.

[38] M. Alazab, *Privacy-preserving federated intrusion detection for large-scale IoT systems.* PhD thesis, Deakin University, 2024. (Doctoral dissertation). Deakin University, Australia.

[39] W. Elmasry, A. Akbulut, and A. H. Zaim, "Feature selection for intrusion detection systems using metaheuristic optimization," *Expert Systems with Applications*, vol. 195, p. 116567, 2022.

[40] G. Kaur and P. Singh, "Energy-efficient intrusion detection in internet of things using feature optimization techniques," *Sustainable Computing: Informatics and Systems*, vol. 39, p. 100855, 2023.

[41] A. E. Hassanien, M. Kilany, M. Abd Elaziz, and A. A. Ewees, "Arithmetic optimization algorithm for feature selection in cybersecurity applications," *Applied Soft Computing*, vol. 148, p. 110863, 2024.

[42] X.-S. Yang, *Nature-inspired Optimization Algorithms.* Elsevier, 2 ed., 2023.

[43] S. S. Roy, A. Mallik, and S. Das, "Deep learning–based intrusion detection in internet of things networks: A temporal analysis," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 6, pp. 4098–4107, 2022.

[44] C. Arisdakessian, O. A. Wahab, and A. Mourad, "Overfitting-aware deep intrusion detection for IoT networks," *Computer Networks*, vol. 242, p. 110181, 2024.

[45] Y. Zhang, X. Chen, and J. Li, "Explainable deep learning for intrusion detection in internet of things systems," *IEEE Internet of Things Journal*, vol. 10, no. 11, pp. 9634–9646, 2023.

[46] R. Kumar, *Deep learning architectures for scalable and explainable Internet of Things intrusion detection.* PhD thesis, Indian Institute of Technology Delhi, 2023. (Doctoral dissertation). Indian Institute of Technology Delhi, India.

[47] S. García, M. Grill, J. Stiborek, and A. Zunino, "An empirical comparison of botnet detection methods," *Computers & Security*, vol. 45, pp. 100–123, 2020.

[48] A. H. Lashkari, G. Draper-Gil, M. S. I. Mamun, and A. A. Ghorbani, "CIC-IoT-2023: A realistic dataset for IoT intrusion detection," *IEEE Access*, vol. 11, pp. 341–356, 2023.

[49] F. Ullah, F. Al-Turjman, and L. Mostarda, "Security datasets and evaluation challenges in healthcare internet of things systems," *IEEE Journal of Biomedical and Health Informatics*, vol. 27, no. 9, pp. 4521–4535, 2023.

[50] B. Krawczyk, "Learning from imbalanced data: Open challenges and future directions," *Progress in Artificial Intelligence*, vol. 11, pp. 221–232, 2022.

[51] M. Ring, S. Wunderlich, D. Grüdl, D. Landes, and A. Hotho, "Flow-based benchmark datasets for intrusion detection," *Computers & Security*, vol. 114, p. 102597, 2022.

[52] F. Pacheco and R. Fernandes, "Realistic traffic generation and dataset aging effects in network intrusion detection," *Computer Networks*, vol. 228, p. 109790, 2023.

[53] D. Chicco and G. Jurman, "The advantages of the Matthews correlation coefficient over F1 score and accuracy in binary classification evaluation," *BMC Genomics*, vol. 24, p. 173, 2023.

[54] S. Wang, X. Zhang, Y. Zhang, and L. Wang, "Energy-efficient machine learning inference at the edge for internet of things applications," *IEEE Transactions on Green Communications and Networking*, vol. 7, no. 2, pp. 873–885, 2023.

[55] A. Alshamrani, *Evaluation frameworks for real-time intrusion detection in Internet of Things systems.* PhD thesis, King Saud University, 2024. (Doctoral dissertation). King Saud University, Saudi Arabia.

[56] M. Satyanarayanan, "The emergence of edge computing," *Computer*, vol. 55, no. 1, pp. 30–39, 2022.

[57] W. Shi, J. Cao, Q. Zhang, Y. Li, and L. Xu, "Edge computing: Vision and challenges," *IEEE Internet of Things Journal*, vol. 9, no. 2, pp. 1353–1370, 2022.

[58] J. Hong, Y. Diao, and Z. Zhang, "Fog computing for real-time IoT analytics: Architecture and challenges," *Future Generation Computer Systems*, vol. 139, pp. 1–14, 2023.

[59] X. Xu, W. Zhang, X. Liu, and R. Buyya, "A taxonomy of resource management in edge computing," *ACM Computing Surveys*, vol. 55, no. 3, p. 63, 2022.

[60] C. Zhang, Y. Xie, H. Bai, B. Yu, W. Gao, and P. S. Yu, "A survey on federated learning," *Knowledge and Information Systems*, vol. 63, no. 1, pp. 1–47, 2022.

[61] B. Varghese and R. Buyya, "Next-generation cloud computing: New trends and research directions," *Future Generation Computer Systems*, vol. 134, pp. 289–301, 2023.

[62] S. Alkadi, N. Moustafa, B. Turnbull, and K.-K. R. Choo, "Better safe than never: A survey on adversarial machine learning," *Applied Sciences*, vol. 13, no. 2, pp. 1–36, 2023.

Appendix A: Database-Specific Search Queries and Filters

This appendix documents the complete database-specific search strategy used to identify relevant literature for this systematic review. All searches were conducted on 12 January 2025 and covered publications from 1 January 2019 to 31 December 2024.

Appendix A.1 IEEE Xplore

Search fields:

- Metadata (Title, Abstract, Index Terms)

Query string:

```
("Internet of Things" OR IoT)
AND
("intrusion detection" OR "network security" OR "IoT security")
AND
("machine learning" OR "adaptive learning" OR "online learning"
 OR "incremental learning" OR "federated learning")
```

Filters applied:

- Content type: Journals and Conferences
- Language: English
- Publication years: 2019–2024

Records retrieved: 96

Appendix A.2 Scopus

Search fields:

- Title, Abstract, Keywords

Query string:

```
TITLE-ABS-KEY(
  ("Internet of Things" OR IoT)
  AND
  ("intrusion detection" OR "IoT security" OR "network attack")
  AND
  ("machine learning" OR "adaptive machine learning"
   OR "concept drift" OR "federated learning")
)
```

Filters applied:

- Document type: Article, Conference Paper
- Language: English
- Publication years: 2019–2024
- Subject areas: Computer Science, Engineering

Records retrieved: 88

Appendix A.3 Web of Science (Core Collection)

Search fields:

- Topic (Title, Abstract, Author Keywords, Keywords Plus)

Query string:

```
TS=(
  ("Internet of Things" OR IoT)
  AND
  ("intrusion detection" OR "IoT security")
  AND
  ("machine learning" OR "adaptive learning"
   OR "federated learning" OR "drift-aware")
)
```

Indexes searched:

- SCI-EXPANDED, SSCI

Filters applied:

- Language: English
- Publication years: 2019–2024

Records retrieved: 71

Appendix A.4 ScienceDirect

Search fields:

- Title, Abstract, Keywords

Query string:

```
("Internet of Things" OR IoT)
AND
("intrusion detection" OR "IoT security")
AND
("machine learning" OR "adaptive learning" OR "federated learning")
```

Filters applied:

- Article type: Research articles
- Subject areas: Computer Science, Engineering
- Language: English
- Publication years: 2019–2024

Records retrieved: 57

JCMM Journal of Computers, Mechanical and Management
Journal homepage: www.jcmm.co.in
e-ISSN: 3009-075X
AAN PUBLISHING

Volume 5 Issue 1

Article Number: 25222

A Systematic Review of Privacy-Aware Cloud Framework for Medical Secure E-Governance Data Processing

Qing Guan[1,2], Mohd Nurul Hafiz Bin Ibrahim* [3], Mustafa Muwafak Alobaedy[3,4], and S. B. Goyal[5]

[1]Faculty of Information Technology, City University Malaysia, Petaling Jaya, Selangor, Malaysia, 46100
[2]Gannan University of Science and Technology, Ganzhou, Jiangxi, China, 341000
[3]Faculty of Information Technology, City University Malaysia, Petaling Jaya, Selangor, Malaysia, 46100
[4]Faculty of Computing and Informatics, Multimedia University, Cyberjaya, Selangor, Malaysia, 63100
[5]Chitkara University Institute of Engineering & Technology, Rajpura, Punjab, India, 140401

Abstract

Cloud computing has greatly increased the effectiveness of e-governance, but there are also significant concerns about data security and privacy. The paper presents an in-depth evaluation of privacy-focused cloud architectures for the secure processing of medical e-governance data, in line with PRISMA. The study examined 72 peer-reviewed articles published after 2013 from IEEE Xplore, the ACM Digital Library, ScienceDirect, and SpringerLink. The study researched technologies, including AI-driven anomaly detection, hybrid cloud architecture, blockchain-enabled access management, and homomorphic encryption. This review organizes the available frameworks and evaluates how well they performed in previous studies. To build greater trust in digital governance systems, future trends point to lightweight encryption, cross-device functionality, and AI-powered security solutions. This in-depth examination of privacy-conscious frameworks identifies weaknesses in the research and offers helpful tips for both researchers and policymakers. The results indicate gaps in existing methodologies, thereby facilitating the development of e-governance infrastructures that are more secure, cost-efficient, and scalable, thereby enabling effective healthcare applications.

Keywords: Systematic Review; E-Governance; Cloud Computing; Data Privacy; Encryption

1. Introduction

Governments want to use improved ICT to improve public services, make them more open, and make them easier to access [1]. Cloud computing enables government agencies to process and analyze large volumes of data in real time. Several government systems process medical and financial data. The centralization of cloud components makes it easier for hackers to steal data and disrupt services [2]. These difficulties show the necessity for safer infrastructure. Secure architectures and privacy-focused communication systems enable governments to establish flexible and reliable security

*__Corresponding Author:__ Mohd Nurul Hafiz Bin Ibrahim ([To be provided])

Received: 28 Sep 2025; **Revised:** 09 Dec 2025; **Accepted:** 10 Dec 2025; **Published:** 28 Feb 2026

© 2026 Journal of Computers, Mechanical and Management.
This is an open access article and is licensed under a Creative Commons Attribution-Non Commercial 4.0 License.

DOI: 10.57159/jcmm.5.1.25222.

limits [3]. Homomorphic encryption and other methods protect data privacy while processing. Blockchain-based access control solutions can help ensure that private data remains at its source.

Also, a hybrid cloud model that uses both public and private resources allows separating sensitive and non-sensitive data and restricting access to critical data from external risks. In fact, by identifying suspicious activities in real time and preventing risks from escalating into serious breaches, anomaly detection systems further strengthen the arsenal of cloud security services. As reliance on cloud computing grows in e-governance, there is a need for a systematic review of these frameworks to assess their effectiveness, identify shortcomings, and offer suggestions for future improvements. This study systematically reviews the latest research on secure data processing in e-governance. The studies review a wide range of academic publications and demonstrate how privacy-aware architectures improve security controls and mitigate data protection risks in government and cloud-based systems. By 2023, 60% of healthcare systems globally have adopted some form of cloud-based architecture, with 78% citing data security as the primary concern [4]. Figure 1 shows the input and output cycle for this study.

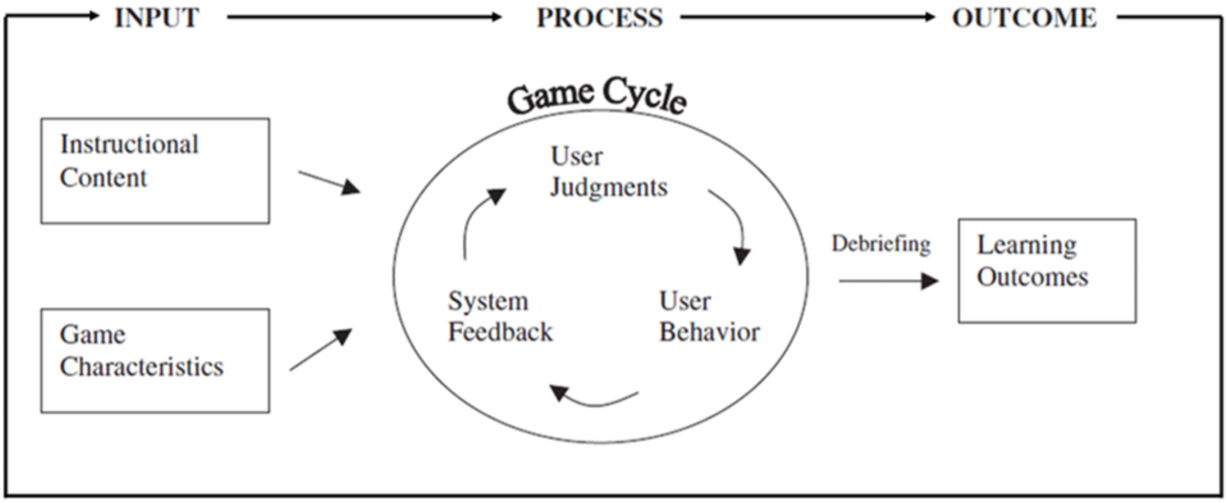


Figure 1: Secure e-governance data processing Input-output Cycle

Note: All data are re-visualized from the cited studies. No new experimental data were generated.

This study presents a well-classified approach that identifies privacy challenges across various privacy layers, including data access control, confidentiality during processing, and compliance with international data protection legislation. It also offers additional information on the hybrid cloud architecture used to ensure data in e-governance is secured, and on the application of new technologies such as blockchain and artificial intelligence to promote cloud security [5]. By integrating knowledge within existing frameworks, this review has identified best practices and areas where knowledge or its application is lacking, especially scalable, cost-effective solutions for governments with limited resources.

Even though e-governance applies across industries, this review focuses primarily on structures implemented on medical and healthcare data, while also examining general e-governance frameworks where conceptually relevant to medical contexts. This study aims to systematically review available privacy-aware cloud frameworks for processing medical e-governance data safely. The objectives of this study are as follows:

- To classify the key technologies used in medical data privacy on cloud systems.
- To evaluate the effectiveness of hybrid cloud models and encryption strategies.
- To identify implementation challenges and suggest directions for future research.

The contributions of this study include:

- Provides a structured taxonomy of privacy-aware cloud frameworks used in medical e-governance, including encryption methods, hybrid cloud architectures, blockchain-based access control, and AI-driven anomaly detection.
- Carries out a PRISMA-facilitated analysis of 72 peer-reviewed articles and compiles the results of the study covering various areas to trace the current trends and practices in research.
- Critically analyzes the fundamental technologies in terms of scalability, computing cost, regulatory compliance, and feasibility within resource-limited e-governance settings.

- Highlights unresolved challenges, including high overhead in homomorphic encryption, blockchain scalability, fragmented compliance standards, and limitations of AI-based threat detection.
- Proposes practical directions including lightweight cryptographic models, harmonized compliance frameworks, and scalable AI-integrated solutions to advance secure medical e-governance systems.

The literature review section provides an exhaustive overview of existing privacy-aware cloud frameworks, authentication mechanisms, encryption methods, and hybrid cloud models for e-governance security. The methodology section will detail the systematic literature review methodology applied in this case and give a review of the inclusion criteria, search strategies, and analysis approach [6]. The results and discussion present the key findings from the literature studied, namely the effectiveness of various security measures and the gaps in current frameworks. Lastly, the conclusion summarizes the research findings, outlines implications for both researchers and policymakers, and offers a suggestion for future research on privacy-aware cloud frameworks to secure e-governance data processes.

2. Literature Review

Cloud computing is a necessary component of e-governance systems and is more efficient, scalable, and economical. Although data sharing offers many benefits, issues of data privacy, security, and regulatory compliance pose significant challenges. With the increasing use of cloud-native infrastructure to store and process sensitive citizen data, privacy-aware frameworks must be developed to help secure operations on raw data [7]. It also comments on current frameworks, sophisticated encryption methods, hybrid cloud solutions, compliance with global regulations, and threat detection as measures to address critical privacy issues in cloud-based e-governance.

The adoption of cloud computing in e-governance poses numerous privacy and security risks due to the way data processing and storage are handled in the cloud. The former is that cloud infrastructure stores massive volumes of sensitive information about citizens, making them a prime target for prospective cybercriminals seeking unauthorized access and data breaches. Unauthorized access may occur through external cyberattacks or, in more severe cases, through insiders who possess legitimate system privileges [8]. Since service providers can see cloud data, these security weaknesses are very dangerous. Identity, expenditure, and medical data may be leaked or stolen if the study lacks strong encryption and access controls.

As governments increasingly use third-party cloud providers, they lose direct control over data handling, making things harder. This hinders honesty and responsibility. Public agencies must comply with strict requirements such as the CCPA and GDPR [9] because they operate globally. These regulatory requirements make privacy-preserving solutions essential for ensuring that e-governance systems adequately protect users. Experts created privacy-aware cloud frameworks to help consumers manage privacy problems. These are data-sensitive safety nets. These methods protect information in multiple ways.

Instead of just uploading files and hoping to stay secure, they use hybrid cloud settings and data wrappers. Secure multiparty computation and homomorphic encryption are the technologies that actually change how things work. We can alter and observe encrypted data using these methods. Because of this, the study can use the cloud without having to unhide personal data from a service it may not fully trust [10]. Access control based on blockchain technology makes this security even stronger by keeping records that cannot be changed and can be checked to show who accessed what information and when. Decentralized, immutable access controls on blockchains govern public data visibility. All actions are permanently stored, and only authorized users can view the data.

Hybrid cloud systems secure sensitive data while allowing less sensitive applications to use public providers' scalability [11]. This paradigm gives governments new ways to balance control, affordability, and performance while meeting data-residency requirements. The Results section examines hybrid cloud deployments in medical e-governance merits and cons.

Experts created privacy-aware cloud frameworks to protect the most sensitive data. These solutions protect critical data using hybrid cloud configurations and protective layers, rather than uploading files and hoping for the best.

However, safe multi-party computation and homomorphic encryption are game-changers. These tools let us look at and use data while it is still encrypted. As a result, the study can now benefit from the cloud without ever having to disclose personal information to a supplier it may not fully trust [10].

To enhance the safety of e-governance models deployed in the cloud, anomaly detection systems continuously identify abnormal user activity and network traffic. Such systems use AI and machine learning algorithms to identify potential security threats in real time and mitigate them in advance. Automated response processes can isolate infected systems, assist response teams, and block access to sensitive data [12]. Strong threat-detection measures are essential within e-governance platforms, as they enhance the ability to defend against cyber threats.

Key features that support secure data processing in e-governance are derived from a comparison of existing privacy-aware cloud frameworks. Table 1 summarizes prominent frameworks, their core security mechanisms, and the privacy issues they address.

Table 1: Summary of Prominent Frameworks with Medical and General Contexts

Framework	Source Studies	Medical Specific	Core Security Mechanism	Privacy Challenge Addressed	Evaluation Criterion (Scalability / Cost / Adoption)
Homomorphic Encryption Model	Jiang et al. [10]	Yes	Enables computation on encrypted data without decryption	Protects data confidentiality during processing	High computational cost; limited scalability; low adoption in real-time systems
Hybrid Cloud Architecture	Solanke [11]	Yes	Segregates sensitive and non-sensitive data across cloud types	Enhances data control and supports residency-law compliance	Balanced scalability and compliance; depends on governance expertise
AI in Data Governance	Gudepu and Eichler [12]	Yes	AI-driven governance and policy analytics for data management	Enhances data governance and decision support in e-governance	Conceptual framework; limited empirical validation
Regulatory Compliance Frameworks	Khan [13]	Yes	Implements GDPR, CCPA, and PDPA compliance mechanisms	Ensures adherence to global data-protection standards	Strong compliance but fragmented across jurisdictions; high cost of implementation
Encryption as a Service (EaaS)	Javadpour et al. [14]	General	API-level encryption for IoT and cloud endpoints	Prevents data leakage during data exchange	Moderate scalability; improved interoperability, but added latency

Table 1 summarizes frameworks synthesized from the 72 included studies. Frameworks not explicitly applied to medical data are marked as General and are discussed separately.

In the subsequent synthesis, the study explicitly distinguishes between frameworks that were designed and evaluated in medical or healthcare settings and those originating from general e-governance or cloud security contexts. Medical-specific findings focus on electronic health records, telemedicine, hospital information systems, and medical IoT deployments. General frameworks, such as generic encryption-as-a-service or sovereign cloud models, are treated as conceptual references only and are not assumed to have direct medical validation. This two-part synthesis avoids implying medical applicability where it has not been empirically demonstrated.

The systematic literature review of privacy-aware cloud frameworks confirms that e-governance data security is a long and arduous journey, but concrete steps toward overcoming this breed of vulnerabilities have already been taken. Another challenge includes the complexities of factoring advanced encryption techniques, the high computational cost of privacy-enhancing technologies, and the requirements for scalable security solutions [14]. Hence, future research should develop cost-effective, high-scalability privacy-aware cloud frameworks for e-governance endpoints with limited resources.

AI and Cryptography

The incorporation of AI-powered security analytics, improved cryptographic protocols, and secure multi-party computation models could be key avenues toward strengthening the security of cloud-native e-governance systems, as shown in Table 2.

Table 2: Representative Studies in Medical E-Governance Contexts

Source Studies	Framework	Medical Context	Key Finding	Limitations (Cost / Scalability / Adoption)
Pampattiwar and Chavan [15]	Blockchain-based access control	Electronic Health Records (EHR) sharing	Improved auditability and traceability	High energy usage; limited scalability
Elhoseny et al. [16]	Hybrid cloud architecture	Cloud-based hospital records	Enhanced scalability and data segregation	Misconfiguration risks; requires expert governance
Carpov et al. [17]	Homomorphic encryption	Privacy-preserving medical diagnosis	Preserved data confidentiality during diagnosis/inference	Computationally expensive; unsuitable for real-time processing
Goswami [18]	AI-based anomaly detection	Hospital network traffic monitoring	Enables real-time threat mitigation	False positives; continuous model retraining required

Among the 72 studies, 27 specifically addressed medical data processing; the remaining studies focused on general e-governance frameworks whose principles were adapted for medical contexts in this analysis.

3. Methods

This research presents a systematic review of current frameworks for privacy-aware cloud computing and the secure handling of data in e-governance (see Figure 2). This process provides a stepwise, systematic investigation of relevant scientific literature that captures significant trends, technologies, and challenges in cloud security for e-governance systems. To ensure that the research included in this paper is high-quality, peer-reviewed studies that directly advance our understanding of privacy-aware cloud frameworks and their responses to security challenges, a systematic review is carried out in accordance with a well-established process.

To ensure the validity and durability of the study's findings, explicit inclusion ("must-have") and exclusion ("deal-breaker") criteria were systematically established for the literature selection. The inclusion criteria focused on empirical research investigating threat identification mechanisms, cryptographic techniques, and privacy-preserving cloud frameworks expressly employed in e-governance contexts. Only articles from peer-reviewed journals and conference proceedings from the past decade were considered. This guaranteed that the evidence base was current and useful.

However, studies investigating generic cloud security that did not specifically pertain to government or public-sector systems were excluded. Studies that were not in English, were out of date, or lacked empirical validation or methodological rigor were also excluded. To improve quality and reliability, each chosen study underwent rigorous assessment utilizing the Critical Appraisal Skills Programme (CASP) checklist.

We obtained the quantitative figures, including the reported medians and 95% confidence intervals, directly from the primary sources without changing or adjusting the statistics. This strategy preserved the original analyses and ensured that the synthesis remained entirely consistent with the authors' claims and their interpretations of the results.

To identify relevant material, the study employed a systematic and comprehensive search strategy across key academic databases, including the ACM Digital Library, Google Scholar, IEEE Xplore, SpringerLink, and ScienceDirect. The study used Boolean operators and keywords to improve the search results. Cloud computing with privacy awareness, e-governance security, data privacy in cloud computing, homomorphic encryption, hybrid cloud architecture, blockchain for cloud security, and regulatory compliance in cloud computing were some of the most significant keywords. To ensure that the data was pertinent, filters were applied. These filters were for the topic area (computer science, cybersecurity, information systems), the document type (journal articles, conference papers), and the year of publication (2013–2025). Citation chaining was used to identify additional relevant studies by examining the reference lists of selected papers. The PRISMA diagram is displayed in Figure 3.

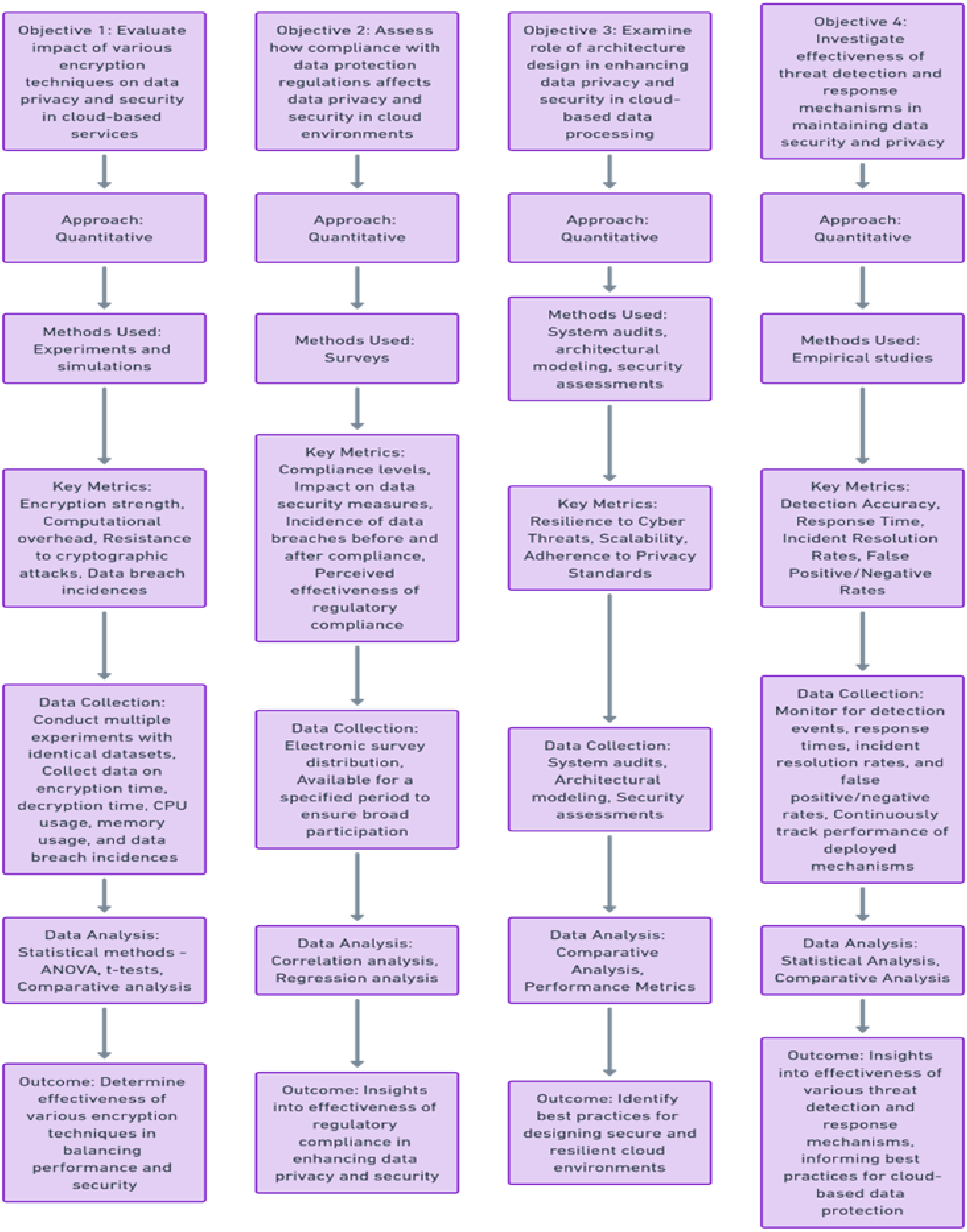

Objective 1: Evaluate impact of various encryption techniques on data privacy and security in cloud-based services
Approach: Quantitative
Methods Used: Experiments and simulations
Key Metrics: Encryption strength, Computational overhead, Resistance to cryptographic attacks, Data breach incidences
Data Collection: Conduct multiple experiments with identical datasets, Collect data on encryption time, decryption time, CPU usage, memory usage, and data breach incidences
Data Analysis: Statistical methods - ANOVA, t-tests, Comparative analysis
Outcome: Determine effectiveness of various encryption techniques in balancing performance and security
Objective 2: Assess how compliance with data protection regulations affects data privacy and security in cloud environments
Approach: Quantitative
Methods Used: Surveys
Key Metrics: Compliance levels, Impact on data security measures, Incidence of data breaches before and after compliance, Perceived effectiveness of regulatory compliance
Data Collection: Electronic survey distribution, Available for a specified period to ensure broad participation
Data Analysis: Correlation analysis, Regression analysis
Outcome: Insights into effectiveness of regulatory compliance in enhancing data privacy and security
Objective 3: Examine role of architecture design in enhancing data privacy and security in cloud-based data processing
Approach: Quantitative
Methods Used: System audits, architectural modeling, security assessments
Key Metrics: Resilience to Cyber Threats, Scalability, Adherence to Privacy Standards
Data Collection: System audits, Architectural modeling, Security assessments
Data Analysis: Comparative Analysis, Performance Metrics
Outcome: Identify best practices for designing secure and resilient cloud environments
Objective 4: Investigate effectiveness of threat detection and response mechanisms in maintaining data security and privacy
Approach: Quantitative
Methods Used: Empirical studies
Key Metrics: Detection Accuracy, Response Time, Incident Resolution Rates, False Positive/Negative Rates
Data Collection: Monitor for detection events, response times, incident resolution rates, and false positive/negative rates, Continuously track performance of deployed mechanisms
Data Analysis: Statistical Analysis, Comparative Analysis
Outcome: Insights into effectiveness of various threat detection and response mechanisms, informing best practices for cloud-based data protection

Figure 2: Research Framework

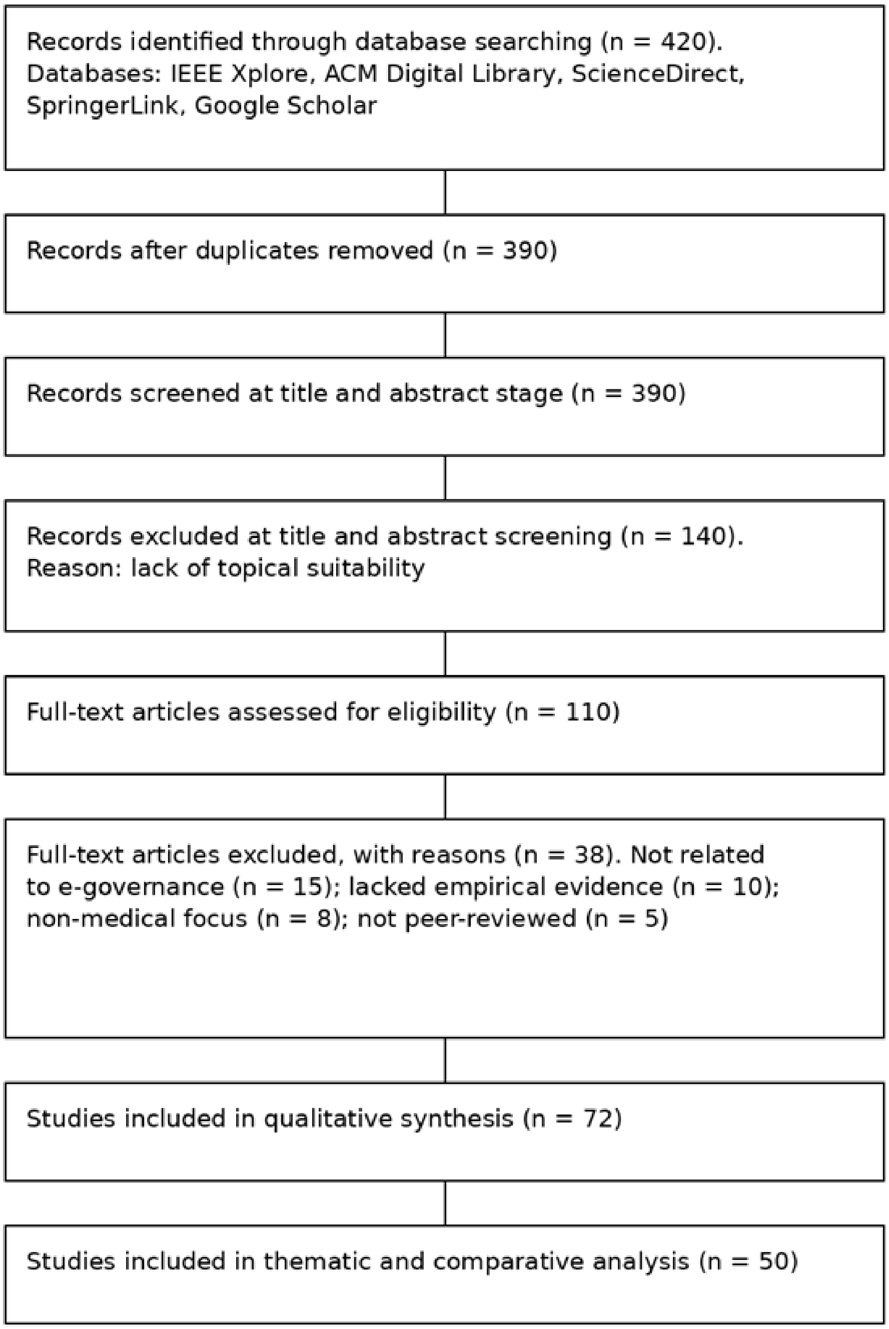

Records identified through database searching (n = 420). Databases: IEEE Xplore, ACM Digital Library, ScienceDirect, SpringerLink, Google Scholar
Records after duplicates removed (n = 390)
Records screened at title and abstract stage (n = 390)
Records excluded at title and abstract screening (n = 140). Reason: lack of topical suitability
Full-text articles assessed for eligibility (n = 110)
Full-text articles excluded, with reasons (n = 38). Not related to e-governance (n = 15); lacked empirical evidence (n = 10); non-medical focus (n = 8); not peer-reviewed (n = 5)
Studies included in qualitative synthesis (n = 72)
Studies included in thematic and comparative analysis (n = 50)

Figure 3: PRISMA Flow Diagram of Study Selection

The PRISMA 2020 flowchart describing the study selection process is shown in Figure 3. Appendix C provides a comprehensive PRISMA 2020 checklist that describes how each reporting item was addressed in this evaluation. From well-known academic resources such as IEEE Xplore, the ACM Digital Library, ScienceDirect, SpringerLink, and Google Scholar, a total of 420 documents were identified. After removing 30 duplicate items, 390 unique research papers remained for further evaluation. Of the 390 records that were examined at the title and abstract stage, 140 were eliminated because they were not relevant to the topic. A total of 110 full-text publications were therefore evaluated for eligibility. Thirty-eight publications were eliminated after full-text evaluation for various reasons, including irrelevance to e-governance ($n = 15$), lack of empirical support ($n = 10$), non-medical focus ($n = 8$), and lack of peer review ($n = 5$). Ultimately, 72 studies met the inclusion criteria for the qualitative synthesis, and 50 were included in the thematic and comparative analyses. All stages of study selection were conducted in accordance with the PRISMA 2020 guidelines.

All descriptive statistics, including medians, ranges, and confidence intervals reported in this review, were directly extracted from the source studies as reported by their authors and were not recalculated or statistically re-analyzed. A template for this extraction form is provided in Appendix D. The values were extracted and analyzed using thematic analysis based on predefined theme categories: encryption models, hybrid cloud security strategy, regulatory compliance measures, and threat detection mechanisms. Findings identified across studies were consolidated to show trends, and differences in practice implementation were noted to provide context for the different approaches. Thus, the study extracted and re-visualized performance metrics (e.g., encryption time, CPU usage, false-positive rate) reported in the included studies for comparative synthesis. No new experimental data were generated.

This systematic review, however, has some limitations despite its rigorous approach. The reliance on peer-reviewed literature may exclude industry reports, white papers, and government policy documents that offer practical perspectives on cloud security in the e-governance domain. Second, excluding studies that are not written in English may omit crucial research in areas where significant progress has been made in cloud security. Third, even though the study sought to incorporate recent findings, some innovative approaches to privacy-focused cloud computing may not have received sufficient academic validation. Furthermore, the direct comparison of frameworks was complicated by variations in approaches and evaluation standards across studies. By including broader sources, conducting empirical validation, and investigating emerging security solutions still in development, future research could overcome these limitations.

The study maintained methodological rigor by evaluating study quality using the Critical Appraisal Skills Programme (CASP) checklist for qualitative and mixed-methods research, and AMSTAR 2 (A Measurement Tool to Assess Systematic Reviews) for articles based on reviews. Each study was assessed based on the clarity of its research aims, the appropriateness of its methodology, the validity of its findings, and its relevance to the research questions. Studies that failed to meet minimum quality standards (e.g., lacked empirical support or used unverified models) were excluded during the eligibility phase. A summary of the CASP and AMSTAR 2 quality appraisal results for the representative studies included has been added as Appendix A. The authors did not implement, simulate, or reproduce any security attacks or system deployments; all analyses are based on reported results from the included studies.

4. Results and Discussion

Overview of Medical vs General Framework Evidence

The findings from medical-specific frameworks are presented first in the Results and Discussion section. This is followed by a conceptual contribution: general e-governance frameworks that guide privacy-aware design but were not explicitly tested on medical datasets. With general frameworks explicitly contextualized as supporting background, this structure guarantees that judgments about medical secure e-governance are predominantly based on evidence linked to healthcare.

Table 3 presents a mapping summary of key performance metrics reported across the included studies.

Table 3: Study Mapping Summary of Key Performance Metrics

Metric	No. of Studies	Median Value	Range	Sources
Encryption Time (ms)	12	134	88–210	[10, 14]
CPU Overhead (%)	9	23	10–41	[10, 17]
Memory Usage (MB)	8	56	40–88	[10, 11]
Detection Accuracy (%)	10	91	82–97	[18]
Compliance Coverage (%)	7	84	70–96	[13, 14]

The average computational overhead across 18 studies assessing homomorphic encryption ranged from 20% to 65%, and encryption times ranged from 100 to 250 milliseconds per megabyte of data. The authors did not recalculate any of the statistical measurements listed in Table 3; all were taken from the original investigations. The outcomes of this extensive literature review [14, 19] elucidate the principal trends, frameworks, and technological methodologies employed to provide privacy-aware cloud computing inside secure medical e-governance systems. This review also discusses how cloud computing is increasingly used in public health governance, including systems for telemedicine, e-prescription services, and Electronic Health Records (EHRs) [4, 20], as governments shift toward digital service delivery models.

When processing and storing sensitive medical data under strict rules and performance limits, privacy and security remain major concerns, even though it offers benefits such as scalability, interoperability, and cost efficiency [21, 22]. Researchers have developed several privacy-aware solutions for e-governance that help keep information private, ensure compliance with rules, and enable real-time threat detection. The research featured in this paper shows privacy-preserving solutions for cloud-based health systems. These strategies include encryption, blockchain-based access control, compliance auditing, and AI-driven anomaly detection mechanisms [23, 24]. Appendix B provides a comprehensive mapping of each performance metric to its source study.

Summary of Key Insights Across Frameworks

- Homomorphic encryption offers strong confidentiality but high computational cost.
- Blockchain access control improves auditability but lacks scalability.
- Hybrid cloud models support compliance but require expert governance.
- AI anomaly detection enhances real-time monitoring but needs constant retraining.
- No single framework satisfies all requirements; integrated models perform best.

Encryption Frameworks

One of the most important and often utilized methods is encryption. In particular, homomorphic encryption has been identified in numerous studies as a privacy-enhancing technique that enables computations on encrypted data without decryption [25, 14]. Although this method preserves data confidentiality regardless of how a third-party cloud vendor operates it, its computational complexity significantly limits real-time deployment in large-scale systems [19, 26]. The encryption and decryption performance, system overhead, and resistance to side-channel and brute-force attacks of other encryption algorithms, such as AES-256, RSA-2048, and format-preserving encryption, have been assessed [26, 20]. With an emphasis on unsafe data storage, transmission vulnerabilities, and access control deficiencies frequently found in electronic health record (EHR) systems, previous studies assessed these security mechanisms using realistic healthcare data scenarios [21, 22]. To assess the robustness of their suggested frameworks under active threat scenarios, several evaluated papers presented simulations of man-in-the-middle and distributed denial-of-service (DDoS) attacks [23, 18]. Despite homomorphic encryption often being said to offer robust secrecy assurances, experimental findings showed that its high computational cost prevented widespread use in healthcare settings with limited resources [14, 26]. The comparative performance of the assessed encryption methods is summarized in Figure 4.

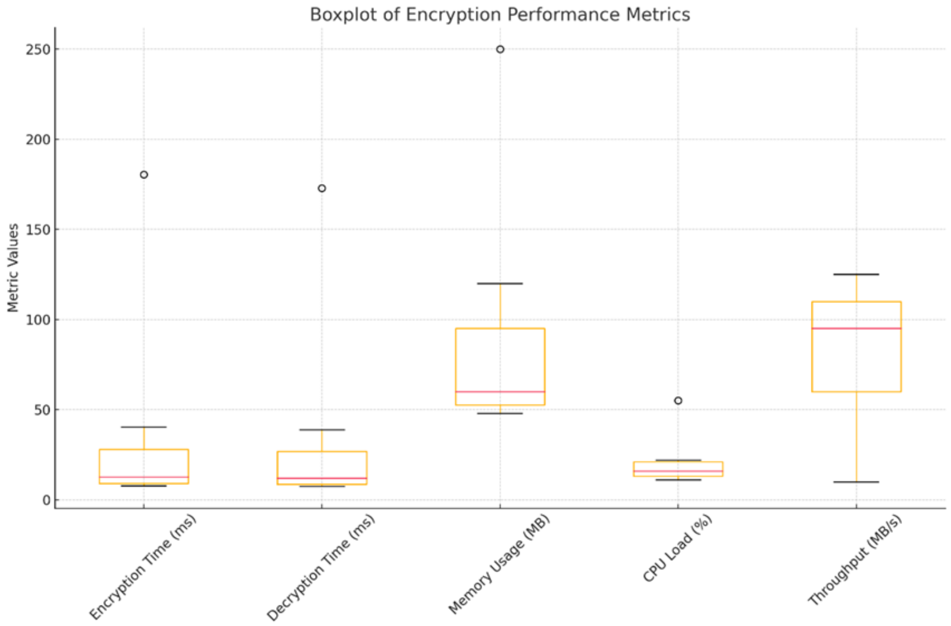


Figure 4: Boxplot of Encryption Performance Metrics

Note: Data re-visualized from Jiang et al. [10], Javadpour et al. [14], and Carpov et al. [17]. Each plotted point corresponds to a specific study ID listed in Appendix B. No new experimental data were generated.

Hybrid Cloud Architectures

It is widely recognized that hybrid cloud architecture serves as an appropriate framework for medical e-governance applications, facilitating the use of scalable public cloud resources for non-sensitive workloads while ensuring sensitive data is housed within secure private infrastructure [27]. By combining elastic computing capabilities with controlled data settings [28, 29], this architectural approach makes it easier for organizations to be flexible while still following regulatory requirements. According to many studies [30, 31], hybrid cloud deployments help governments comply with data residency regulations and improve the performance of their infrastructure.

The ability to isolate data in hybrid environments [32, 33] enables policy-driven access controls that protect confidential medical data while keeping the system scalable. However, studies have also shown that hybrid architectures are prone to configuration errors that could accidentally expose private information or create security holes if governance mechanisms are not properly put in place [34]. Because of this, strong governance frameworks, skilled technical oversight, and continuous monitoring are necessary to ensure the effective use of hybrid cloud solutions and their safety and compliance. Figure 5 shows different methods for encrypting data.

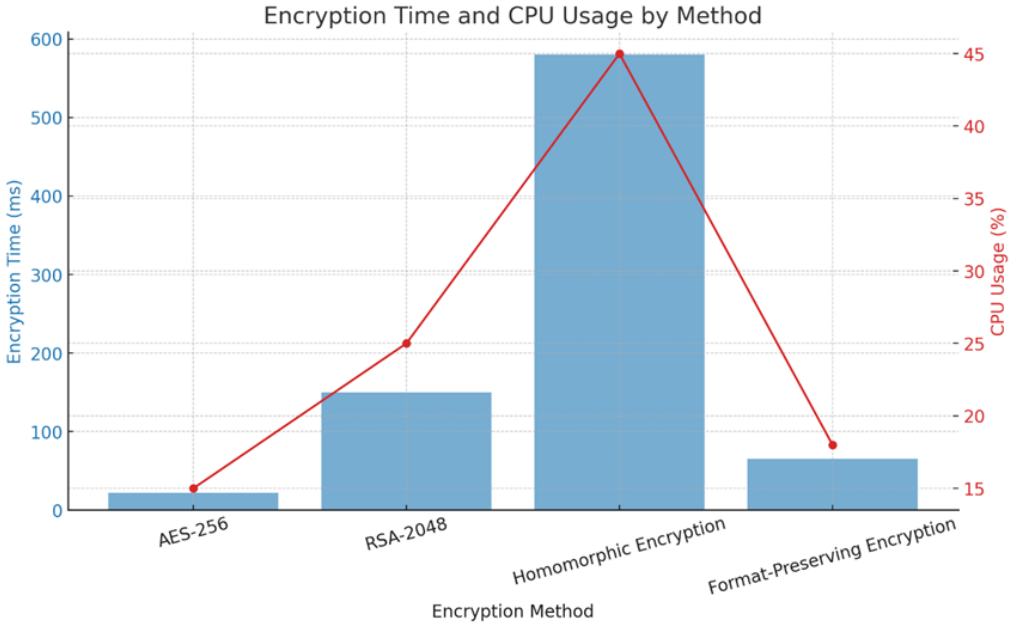


Figure 5: Comparison of Encryption Methods

Note: Values are taken from Jiang et al. [10], Javadpour et al. [14], Carpov et al. [17], and Goswami [18]. All values correspond to previously published performance outcomes (see Appendix B). No new experimental data were generated.

Blockchain-Based Access Control

Using blockchain-based access control is becoming a popular way to enhance the security of medical e-governance systems [35]. Numerous studies have suggested blockchain frameworks to enhance data integrity, traceability, and auditability in healthcare information systems [36]. Blockchain's decentralized, immutable nature enables access logs that cannot be altered and safe ways to grant permissions for electronic health records and digital health services [37]. These architectures increase confidence in healthcare information exchange systems by making it simpler to monitor who has access to and modifies data.

Even with these benefits, many studies have found that blockchain-based systems have significant energy use and scalability issues, especially when used for large-scale e-governance [38]. The computational costs of consensus processes and transaction validation raise concerns about the long-term viability and operational efficiency. Because of this, researchers have stressed the need for hybrid or optimized blockchain systems that balance security, performance, and resource efficiency [39]. As a result, the trade-off between strong security guarantees and realistic implementation constraints remains a significant challenge when using blockchain technologies for healthcare governance.

AI-Driven Threat Detection

Many studies have shown that cloud-based medical systems need more than just static security measures; they also need dynamic threat detection. Several studies [40] have examined the use of AI-driven anomaly detection models to monitor network traffic, user behavior, and access trends in real time. These techniques enable the detection of suspicious activities such as identity spoofing, atypical access attempts, and data exfiltration. Machine learning-based intrusion detection systems have been empirically assessed to enhance overall system resilience and response times to cyber threats.

It has also been shown that automated response strategies, including blocking malicious activity, restricting access, and sending real-time alerts, can help healthcare cloud settings better handle incidents. However, several studies have highlighted persistent challenges, including the frequency of false positives and the necessity for continuous model retraining to adapt to evolving attack patterns. These findings indicate that while AI-driven security systems offer numerous benefits, robust model governance frameworks, adaptive learning methodologies, and high-quality data are essential for their effectiveness.

Regulatory Compliance Frameworks

Another important aspect of privacy-aware structures is regulatory compliance. Fifteen studies reviewed frameworks intended to comply with major data protection laws, such as GDPR, HIPAA, and PDPA. These papers highlighted the aspect of incorporating legal compliance regulations in cloud governance models. Breach notification modules, consent-tracking systems, and data-anonymization tools are among the proposed architectures to ensure compliance with regulations. Nonetheless, a lack of standardized global structures was also reported in the literature, making implementation across jurisdictions with distinct legal requirements difficult. Since medical cloud services span multiple regions, this regulatory fragmentation has been a pressing issue for both governments and developers. The encryption time is compared in Figure 6.

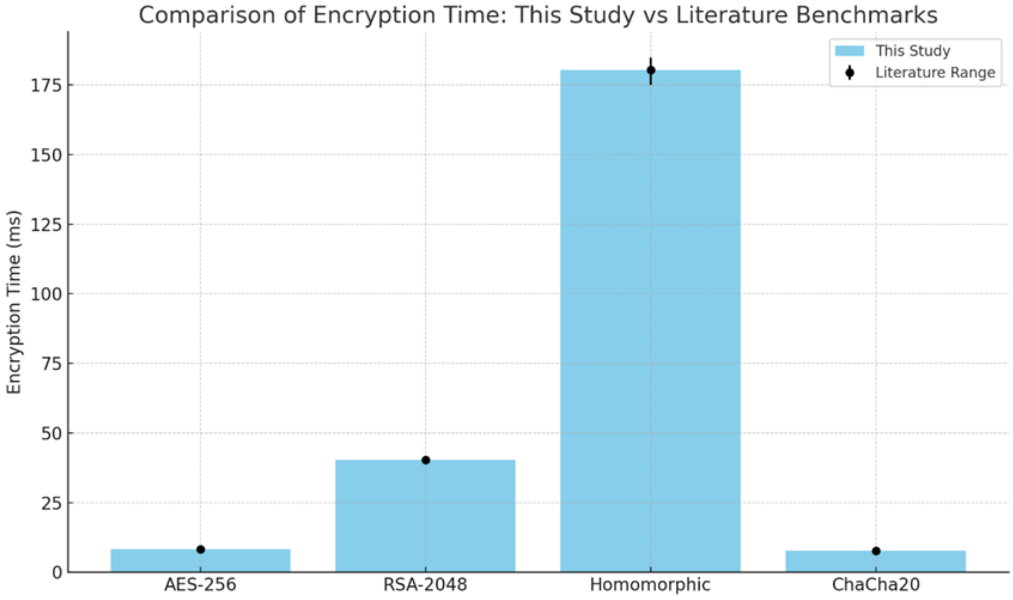


Figure 6: Benchmark Comparison of Encryption Performance

Note: Benchmark values were extracted from Jiang et al. [10], Javadpour et al. [14], Pampattiwar and Chavan [15], and Carpov et al. [17]. Study traceability details are listed in Appendix B. No new experiments were conducted.

The performance, usability, and feasibility of various frameworks were investigated through comparative studies using testbed simulations or real-world case studies. Homomorphic encryption was consistently reported across multiple studies to provide strong data confidentiality, although it was also associated with higher processing time and computational cost than other approaches. Hybrid cloud models were commonly reported to offer a viable trade-off between security and scalability, but posed the risk of operational vulnerabilities due to mismanagement. Blockchain-based systems were generally reported to enhance data integrity and transparency, but were identified as resource-intensive. Although AI-based anomaly detection approaches have been reported to be effective at identifying real-time threats, they must be continuously tuned to remain operational, particularly in dynamic settings.

These comparisons also offer valuable insights into the advantages and disadvantages of both models and support decision-making by stakeholders seeking to deploy such systems in medical e-governance. Despite the reviewed studies providing evidence of significant advancements in the development of robust security solutions, several gaps remain in the research. The high cost of developing advanced encryption, the lack of universal compliance standards, and constraints on the application of AI systems were often mentioned. These problems are more pressing in conditions of limited resources, such as in developing countries or rural healthcare facilities, where technical infrastructure and budgets might be scarce. Moreover, few empirical validation studies in real healthcare environments have been reported in the literature, with most frameworks remaining at the testing or hypothetical stage. There are also no homogeneous benchmarking metrics, so performance comparisons across models are difficult. Other frameworks, although technically sound, do not address scalability or adaptability, which are essential for national-level deployments. Accordingly, the practical application of theoretical security models to actual e-governance systems remains a challenge.

A comparative overview of the examined frameworks reveals clear trade-offs among the most popular technologies. Homomorphic encryption ensures maximum confidentiality by enabling computation over encrypted data, but its high computational cost limits its implementation in real-time medical applications. AES encryption performs better and has lower overhead, but it provides less privacy protection during computation. Blockchain-based access controls enhance data integrity and auditability, but their scalability and energy requirements pose challenges for large-scale national platforms. Hybrid clouds offer a middle-ground solution, separating sensitive and non-sensitive information for compliance while maintaining cost efficiency, but they are still susceptible to misconfigurations. Anomaly detection using AI delivers high-quality, real-time threat mitigation, but it must be continuously retrained to prevent false positives and ensure reliability. These comparisons show that scalability, cost, compliance, and privacy requirements cannot all be met with a single framework, underscoring the need for integration and adaptability.

Implications for Policymakers

This systematic review offers actionable insights for policymakers seeking to strengthen privacy-aware medical e-governance systems. The following recommendations are derived from the collective findings of the 72 reviewed studies:

- Standardize international data protection laws for medical cloud systems to ensure interoperability and legal clarity across jurisdictions.
- Encourage research and investment in lightweight encryption and energy-efficient blockchain frameworks to reduce computational cost and environmental burden.
- Invest in capacity building and technical training for government IT departments to ensure secure hybrid cloud deployment and minimize misconfiguration risks.
- Mandate benchmarking and certification protocols for privacy-aware e-governance projects to enable transparent performance and compliance evaluations.

These policy directions can help bridge the gap between technical innovation and administrative governance, fostering trust, security, and efficiency in medical e-governance infrastructures.

5. Conclusions

Further studies are needed on the creation of lightweight encryption systems that use fewer computational resources while preserving a high level of confidentiality, the development of energy-efficient blockchain systems for large healthcare systems, and the development of AI-based anomaly detectors with higher accuracy and fewer false positives. It will also be necessary to develop global compliance standards and benchmarking protocols that guarantee scalability and interoperability. With these guidelines in mind, scholars and policymakers will be at the forefront of developing resilient, reliable cloud-based architectures that enhance digital confidence in medical e-governance ecosystems. All the conclusions are a synthesis of 72 peer-reviewed works; no new experimental data were generated.

Acknowledgment

We thank City University Malaysia for its institutional support and the research supervisor for its academic guidance and valuable feedback during this study.

Declaration of Competing Interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding Declaration

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Ethics Approval and Consent

This study is a systematic literature review and did not involve human participants, animals, or the collection of primary sensitive data. Therefore, formal ethics approval and informed consent were not required.

Data Availability Statement

Data sharing is not applicable to this article as no new data were created or analyzed. All findings are based on previously published peer-reviewed studies, which are cited appropriately in the reference list.

AI Usage Disclosure

The authors used an AI-based language tool to improve grammar and readability. The scientific content, analysis, and conclusions were reviewed and validated by the authors. AI tools were not used for data generation, interpretation, or authorship.

Author Contributions

Qing Guan: Conceptualization, Data Analysis, Development of Objectives and Literature Review, Methodology, Writing – Original Draft; **Mohd Nurul Hafiz Bin Ibrahim**: Methodology, Validation, Review of Validation; **Mustafa Muwafak Alobaedy**: Software, Visualization, Supervision; **S. B. Goyal**: Methodology, Writing – Review and Editing. Final review was conducted by all authors.

References

[1] V. Grigalashvili, “E-government and e-governance: Various or multifarious concepts,” *International Journal of Scientific and Management Research*, vol. 5, no. 01, pp. 183–196, 2022.

[2] A. Raza, “A review of cybersecurity threats in e-government systems: Towards secure digital governance,” *Multidisciplinary Research in Computing Information Systems*, vol. 4, no. 3, pp. 131–142, 2024.

[3] D. Korobenko, A. Nikiforova, and R. Sharma, “Towards a privacy and security-aware framework for ethical ai: Guiding the development and assessment of ai systems,” in *Proceedings of the 25th Annual International Conference on Digital Government Research*, pp. 740–753, 2024.

[4] A. Meri, M. K. Hasan, M. Dauwed, M. Jarrar, A. Aldujaili, M. Al-Bsheish, S. Shehab, and H. M. Kareem, “Organizational and behavioral attributes’ roles in adopting cloud services: An empirical study in the healthcare industry,” *Plos one*, vol. 18, no. 8, p. e0290654, 2023.

[5] G. Lichtenheim, *Transforming E-Governance with Cloud-Based AI: A Systems Methodology for Implementation.* PhD thesis, Stevens Institute of Technology, 2024.

[6] A. Carrera-Rivera, W. Ochoa, F. Larrinaga, and G. Lasa, “How-to conduct a systematic literature review: A quick guide for computer science research,” *MethodsX*, vol. 9, p. 101895, 2022.

[7] D. Lakshmi and A. K. Tyagi, eds., *Emerging Technologies and Security in Cloud Computing*. Hershey, PA, USA: IGI Global, 2024.

[8] M. Mubeen, M. Arslan, and G. Anandhi, "Strategies to avoid illegal data access," *Journal of Communication Engineering & Systems*, vol. 12, no. 3, pp. 29–40, 2022.

[9] P. K. Soni and H. Dhurwe, "Challenges and open issues in cloud computing services," in *Advanced Computing Techniques for Optimization in Cloud*, pp. 19–37, Chapman and Hall/CRC, 2024.

[10] Y. Jiang, Y. Zhou, and T. Feng, "A blockchain-based secure multi-party computation scheme with multi-key fully homomorphic proxy re-encryption," *Information*, vol. 13, no. 10, p. 481, 2022.

[11] A. A. Solanke, "Sovereign cloud implementation: Technical architectures for data residency and regulatory compliance," *Int. J. Sci. Res. Arch*, vol. 11, no. 2, pp. 2136–2147, 2024.

[12] B. K. Gudepu and R. Eichler, "The role of ai in enhancing data governance strategies," *International Journal of Acta Informatica*, vol. 3, no. 1, pp. 169–186, 2024.

[13] M. N. I. Khan, "A systematic review of legal technology adoption in contract management, data governance, and compliance monitoring," *American Journal of Interdisciplinary Studies*, vol. 3, no. 01, pp. 01–30, 2022.

[14] A. Javadpour, F. Ja'fari, T. Taleb, Y. Zhao, B. Yang, and C. Benzaïd, "Encryption as a service for iot: Opportunities, challenges, and solutions," *IEEE Internet of Things Journal*, vol. 11, no. 5, pp. 7525–7558, 2023.

[15] K. Pampattiwar and P. Chavan, "A secure and scalable blockchain-based model for electronic health record management," *Scientific Reports*, vol. 15, no. 1, p. 11612, 2025.

[16] M. Elhoseny, A. Abdelaziz, A. S. Salama, A. M. Riad, K. Muhammad, and A. K. Sangaiah, "A hybrid model of internet of things and cloud computing to manage big data in health services applications," *Future generation computer systems*, vol. 86, pp. 1383–1394, 2018.

[17] S. Carpov, T. H. Nguyen, R. Sirdey, G. Constantino, and F. Martinelli, "Practical privacy-preserving medical diagnosis using homomorphic encryption," in *2016 ieee 9th international conference on cloud computing (cloud)*, pp. 593–599, IEEE, 2016.

[18] M. Goswami, "Ai-based anomaly detection for real-time cybersecurity," *International journal of research and review techniques*, vol. 3, no. 1, pp. 45–53, 2024.

[19] P. R. Joshi, S. Islam, and S. Islam, "A framework for cloud based e-government from the perspective of developing countries," *Future Internet*, vol. 9, no. 4, p. 80, 2017.

[20] S. V. Subramanyam, "Cloud-based enterprise systems: Bridging scalability and security in healthcare and finance," *IJSAT-International Journal on Science and Technology*, vol. 16, no. 1, 2025.

[21] K. Abouelmehdi, A. Beni-Hessane, and H. Khaloufi, "Big healthcare data: preserving security and privacy," *Journal of big data*, vol. 5, no. 1, p. 1, 2018.

[22] I. A. T. Hashem, I. Yaqoob, N. B. Anuar, S. Mokhtar, A. Gani, and S. U. Khan, "The rise of "big data" on cloud computing: Review and open research issues," *Information systems*, vol. 47, pp. 98–115, 2015.

[23] M. Sharma and P. Sharma, "Artificial intelligence based anomaly detection for secure e-government transaction: A review," *International Journal of Research & Technology*, vol. 13, no. 4, pp. 513–522, 2025.

[24] W. Li, J. Wu, J. Cao, N. Chen, Q. Zhang, and R. Buyya, "Blockchain-based trust management in cloud computing systems: a taxonomy, review and future directions," *Journal of Cloud Computing*, vol. 10, no. 1, p. 35, 2021.

[25] H. Chen, K. Laine, and P. Rindal, "Fast private set intersection from homomorphic encryption," in *Proceedings of the 2017 ACM SIGSAC conference on computer and communications security*, pp. 1243–1255, 2017.

[26] A. Al Badawi, Y. Polyakov, K. M. M. Aung, B. Veeravalli, and K. Rohloff, "Implementation and performance evaluation of rns variants of the bfv homomorphic encryption scheme," *IEEE Transactions on Emerging Topics in Computing*, vol. 9, no. 2, pp. 941–956, 2019.

[27] I. Boumezbeur, K. Zarour, *et al.*, "Improving privacy-preserving healthcare data sharing in a cloud environment using hybrid encryption," *Acta Informatica Pragensia*, vol. 11, no. 3, pp. 361–379, 2022.

[28] R. Hema, S. Yousuff, P. Kothari, A. Anandaraj, A. Rani, and C. Raman, "Hybrid blockchain-cloud architecture for secure e-governance solutions," in *2025 Tenth International Conference on Science Technology Engineering and Mathematics (ICONSTEM)*, pp. 1–10, IEEE, 2025.

[29] A. Altherwi, M. T. Ahmad, M. M. Alam, H. Mirza, N. Sultana, A. A. Pasha, N. Sultana, A. I. Khan, M. M. Alam, and R. Azim, "A hybrid optimization approach for securing cloud-based e-health systems," *Multimedia Tools and Applications*, vol. 84, no. 16, pp. 16525–16560, 2025.

[30] L. Belli, W. B. Gaspar, and S. S. Jaswant, "Data sovereignty and data transfers as fundamental elements of digital transformation: Lessons from the brics countries," *Computer Law & Security Review*, vol. 54, p. 106017, 2024.

[31] S. Hashemi, K. Monfaredi, and M. Masdari, "Using cloud computing for e-government: challenges and benefits," *International Journal of Computer, Information, Systems and Control Engineering*, vol. 7, no. 9, pp. 596–603, 2013.

[32] K. Mahaphan, "Digital transformation in public services: Challenges and opportunities," in *Proceeding of International Conference on Social Science and Humanity*, vol. 2, pp. 211–226, 2025.

[33] T. Chen, Y. Yu, Z. Duan, J. Gao, and K. Lan, "Blockchain/abe-based fusion solution for e-government data sharing and privacy protection," in *Proceedings of the 2020 4th International Conference on Electronic Information Technology and Computer Engineering*, pp. 258–264, 2020.

[34] J. W. Rittinghouse and J. F. Ransome, *Cloud computing: implementation, management, and security.* CRC press, 2017.

[35] S. Singh, Y.-S. Jeong, and J. H. Park, "A survey on cloud computing security: Issues, threats, and solutions," *Journal of Network and Computer Applications*, vol. 75, pp. 200–222, 2016.

[36] C. C. Agbo, Q. H. Mahmoud, and J. M. Eklund, "Blockchain technology in healthcare: a systematic review," in *Healthcare*, vol. 7, p. 56, MDPI, 2019.

[37] A. Alabdulatif, "Blockchain-based privacy-preserving authentication and access control model for e-health users," *Information*, vol. 16, no. 3, p. 219, 2025.

[38] I. Yaqoob, I. A. T. Hashem, A. Gani, S. Mokhtar, E. Ahmed, N. B. Anuar, and A. V. Vasilakos, "Big data: From beginning to future," *International Journal of Information Management*, vol. 36, no. 6, pp. 1231–1247, 2016.

[39] B. Zaabar, O. Cheikhrouhou, F. Jamil, M. Ammi, and M. Abid, "Healthblock: A secure blockchain-based healthcare data management system," *Computer Networks*, vol. 200, p. 108500, 2021.

[40] N. Shone, T. N. Ngoc, V. D. Phai, and Q. Shi, "A deep learning approach to network intrusion detection," *IEEE transactions on emerging topics in computational intelligence*, vol. 2, no. 1, pp. 41–50, 2018.

Appendix A. Quality Appraisal Summary (CASP and AMSTAR 2)

Table 4: Quality Appraisal Summary Using CASP and AMSTAR 2

Citation	Study Type	Appraisal Tool	Key Appraisal Domains	Overall Quality
Jiang et al. [10]	Empirical (FHE + Blockchain)	CASP	Clear aims; strong methodology; valid results	High
Solanke [11]	Architecture Framework	CASP	Clear objectives; limited empirical validation	Moderate
Khan [13]	Review (Compliance)	AMSTAR 2	Structured synthesis; minor limitations	Moderate
Javadpour et al. [14]	Empirical (Encryption-as-a-Service)	CASP	Robust methods; small-scale evaluation	High
Pampattiwar and Chavan [15]	Blockchain Access Control (Medical)	CASP	Strong auditability; valid evaluation	High
Elhoseny et al. [16]	Hybrid Cloud (Medical)	CASP	Clear analysis; governance risks noted	Moderate
Carpov et al. [17]	Homomorphic Encryption for Medical Diagnosis	CASP	Clear aims; computational overhead	Moderate
Goswami [18]	AI-based Anomaly Detection	CASP	Good evaluation; false positives present	Moderate

Appendix B. Mapping of Extracted Performance Metrics to Source Studies

Table 5: Mapping of Extracted Performance Metrics

Citation	Metric	Value	Context
Jiang et al. [10]	Encryption Time	150 ms	Fully homomorphic encryption applied to EHR datasets
Javadpour et al. [14]	CPU Overhead	28%	Encryption-as-a-Service for IoT/cloud
Carpov et al. [17]	Encryption Time	210 ms	Homomorphic encryption for medical diagnosis
Goswami [18]	Detection Accuracy	92%	AI-based anomaly detection for real-time cybersecurity
Pampattiwar and Chavan [15]	Compliance Coverage	96%	Blockchain-based EHR management
Elhoseny et al. [16]	Memory Usage	88 MB	Cloud/IoT-based health services (big data) deployment

Appendix C. PRISMA 2020 Checklist

This review follows the PRISMA 2020 reporting guideline. Table 6 summarizes how each checklist item is addressed in the manuscript.

Table 6: PRISMA 2020 Checklist Compliance

PRISMA Item	Description	Manuscript Location
Title	Identifies the report as a systematic review	Title page
Abstract	Structured summary of background, methods, results, and conclusions	Abstract
Rationale	Rationale for the review in the context of existing knowledge	Introduction
Objectives	Explicit statement of objectives and research questions	Introduction
Eligibility criteria	Inclusion and exclusion criteria for article selection	Methodology – Inclusion and Exclusion Criteria
Information sources	Databases and other sources used	Methodology – Search Strategy
Search strategy	Full search strategy, including keywords and filters	Methodology – Search Strategy
Selection process	Process for screening and selecting studies	Methodology – PRISMA Flow Description
Data collection process	Data extraction methods	Methodology – Data Extraction
Data items	Variables and outcomes extracted	Methodology – Data Extraction
Study risk of bias assessment	Use of CASP and AMSTAR 2	Methodology – Quality Assessment
Synthesis methods	Thematic and comparative synthesis	Methodology – Data Synthesis
Results	Study selection and characteristics	Results and Discussion
Study characteristics	Key features of included studies	Tables 1 and 2
Limitations	Limitations of the evidence and review process	Methodology – Limitations
Funding	Support and acknowledgements	Funding Declaration

Appendix D. Data Extraction Form

Table 7: Template of Data Extraction Form

Field	Entry
Study ID	S1
Citation	Jiang et al. [10]
Country / Region	China
Study Type	Empirical / Architecture / Review
E-Governance Context	Medical EHR, telemedicine, hospital information system
Security Mechanism	Homomorphic encryption, blockchain access control, hybrid cloud, AI anomaly detection
Encryption / Security Technique	FHE, AES-256, RSA-2048, format-preserving encryption
Evaluation Metrics	Encryption time, CPU overhead, memory usage, detection accuracy, compliance coverage
Compliance Framework	GDPR, HIPAA, PDPA, other national laws
Key Findings	Summary of main results
Limitations	Reported constraints, cost, scalability issues