ARTIFICIAL INTELLIGENCE–DRIVEN ADAPTIVE HONEYPOTS FOR WIRELESS NETWORKS: A COMPARATIVE SURVEY

S. Keerthi Priya[1], V. Vadivu[2], S. Karthikeyan[3]

ABSTRACT

Wireless networks, including Wireless Sensor Networks (WSNs), Mobile Ad Hoc Networks (MANETs), Vehicular Ad Hoc Networks (VANETs), Wi-Fi, and Internet of Things (IoT) systems, play a critical role in modern communication infrastructures but remain highly vulnerable to cyberattacks due to their open medium, dynamic topology, and resource constraints. Traditional security mechanisms such as firewalls and intrusion detection systems primarily focus on prevention and detection and often fail to capture in-depth attacker behavior. Honeypots have emerged as an effective deception-based defense mechanism; however, conventional static honeypots are easily identifiable and unsuitable for dynamic wireless environments. To address these limitations, adaptive honeypots enhanced with artificial intelligence (AI) have gained increasing attention.

This paper presents a comprehensive comparative survey of AI-driven adaptive honeypots for wireless networks, covering research from 2016 to 2026. The survey systematically reviews learning-based techniques, including supervised learning, unsupervised learning, and reinforcement learning, applied to honeypot configuration, deployment, and autonomous adaptation. Existing works are analyzed with respect to network domain, AI methodology, key contributions, and limitations. Furthermore, the paper highlights emerging trends such as real-time self-adaptation, generative deception, and distributed intelligence, while identifying critical research gaps related to scalability, explainability, resource efficiency, and standardized evaluation. The survey aims to provide researchers and practitioners with a structured understanding of the state of the art and to outline promising directions for future AI-driven honeypot research in wireless networks.

Keywords : AI-driven Honeypots, Adaptive Security, Wireless Network Security, Cyber Deception, Machine Learning, IoT Security

I. INTRODUCTION

Wireless networks such as Wireless Sensor Networks (WSNs), Mobile Ad Hoc Networks (MANETs), Vehicular Ad Hoc Networks (VANETs), Wi-Fi networks, and Internet of Things (IoT) systems have become an integral part of modern communication infrastructures. These networks support a wide range of applications including environmental monitoring, healthcare, smart transportation, and industrial automation. Despite their advantages, wireless networks are highly vulnerable to cyberattacks due to their broadcast communication medium, dynamic topology, node mobility, limited computational resources, and lack of centralized control [2].

Traditional security mechanisms such as firewalls and intrusion detection systems mainly focus on attack prevention and detection and often fail to provide deep insights into attacker behavior. Honeypots have emerged as an effective deception-based security mechanism that attracts attackers to decoy systems in order to analyze their strategies, tools, and attack patterns [3]. However, conventional honeypots are mostly static in nature and can be easily identified by experienced attackers, which significantly reduces their effectiveness in dynamic wireless environments [4].

To overcome these limitations, adaptive honeypots have been introduced. Adaptive honeypots dynamically modify their configuration, exposed services, and interaction levels based on observed attacker behavior and network conditions [5]. Recent advances in artificial intelligence (AI) have further enhanced honeypot adaptivity by enabling intelligent learning, behavioral analysis, and autonomous response capabilities [6].

In this survey, artificial intelligence refers specifically to learning-based techniques, including supervised learning, unsupervised learning, and reinforcement learning, applied to adaptive honeypots in wireless networks. Supervised learning techniques rely on labeled datasets for detecting known attack patterns, while unsupervised learning techniques identify

Department of Information Technology,[1]
Karpagam Academy of Higher Education, Coimbatore.[1]
keerthipriyas2001@gmail.com[1]

Department of Artificial Intelligence & Data Science,[2]
Karpagam Academy of Higher Education, Coimbatore.[2]
vadivu.vijayan@kahedu.edu.in[2]

Department of Computer Science,[3]
Rathinam College of Arts and Science, Coimbatore.[3]
s.karthics@gmail.com[3]

* Corresponding Author

anomalies without prior knowledge. Reinforcement learning enables autonomous decision-making by continuously interacting with attackers and the network environment [11].

II. RELATED WORK

In [1] authors introduced a machine learning-based strategy for dynamically configuring, deploying, and maintaining honeypots by clustering devices in a network and intelligently placing honeypots without manual intervention. In [2] researchers proposed a framework combining honeypots with unsupervised anomaly detection (using time-series analysis and extreme value theory) to improve early detection of anomalous nodes in network traffic. In [3] authors focused on neural networks, this work uses a honeypot-inspired defensive strategy to lure adversarial attacks and analyze their behavior, illustrating early intersections of deception and AI.

In [4] researchers did a survey that includes adaptive and AI-enhanced honeypot systems for IoT and cyber-physical environments, providing a wide view of state-of-the-art methods, including machine learning integration. In [5] authors reviewed early adoption of ML and ANN in honeypot design and foundational frameworks bridging static and AI-enhanced honeypots. In [6] authors proposed a self-adaptive honeypot framework using machine learning to dynamically reconfigure honeypot services, interaction depth, and monitoring policies based on real-time attack behavior.

In [7] researchers presented a real-time AI-enabled honeypot that continuously classifies attack patterns and adapts its behavior through incremental learning. In [8] researchers explored generative AI techniques for dynamically creating realistic honeypot responses, enhancing deception fidelity and attacker engagement. In [9] authors introduced a distributed adaptive honeypot architecture combining reinforcement learning with blockchain-based trust management for IoT and wireless environments. In [10] researchers proposed a deep learning-driven honeynet that autonomously adjusts deception complexity across multiple layers based on attacker sophistication. In [15], the researchers showed how hierarchical structures combined with machine learning can further address VANET challenges in a way conceptually similar to the original paper's hierarchical dissemination model and emphasized the importance of intelligent cluster head selection, QoS-aware routing, and machine learning techniques for efficient communication and data dissemination in VANETs.

Table 1: Summary of Related Work on AI-Driven Adaptive Honeypots

Author	Year	Network Domain	AI Method	Key Contribution	Limitation
SecureNet-RL (Springer)	2025	5G Wireless	DQN, PPO, Federated RL	Real-time adaptive threat hunting	Training cost
AARF (MDPI)	2024	Multi-honeypot	Multi-Agent RL	Autonomous deception and response	Scalability
Results in Engineering	2025	Distributed Networks	Adaptive Honeypots	DoS/DDoS mitigation	Dataset dependency
ICCK (SSH Honeypots)	2025	Network Security	RF + Auto encoder	AI-driven intrusion detection	Limited adaptability
Multi-Agent RL Survey	2025	Cyber Defense	MARL	Decentralized adaptive defense	Adversarial risks

III. PROPOSED WORK

Based on the analysis of existing literature, it is evident that although AI-driven adaptive honeypots have shown promising results, several challenges remain unresolved in wireless network environments. Most existing approaches either focus on generic network settings or employ computationally expensive deep learning models that are unsuitable for resource-constrained wireless networks. Moreover, there is a lack of unified frameworks that combine adaptability, lightweight intelligence, explainability, and comparative evaluation across different wireless domains.

To address these limitations, this paper proposes a Lightweight AI-Driven Adaptive Honeypot Framework for Wireless Networks. The proposed work aims to design a modular and scalable honeypot architecture capable of autonomously adapting its behavior based on real-time attacker interactions, network conditions, and resource constraints.

A. Objectives of the Proposed Work

The primary objectives of the proposed framework are:

1. To design an adaptive honeypot architecture suitable for wireless networks such as WSNs, MANETs, VANETs, Wi-Fi, and IoT.

2. To integrate lightweight AI techniques (supervised, unsupervised, and reinforcement learning) for real-time attack detection and adaptation.
3. To dynamically adjust honeypot parameters such as service exposure, interaction level, and response behavior.
4. To enable explainable decision-making, allowing security analysts to understand why specific adaptations occur.
5. To perform a comparative evaluation between static, rule-based, and AI-driven adaptive honeypots using standardized performance metrics.

B. Proposed Framework Architecture

The proposed system consists of six key components:

1. Wireless Network Environment
   Represents the target network (WSN, MANET, VANET, IoT, or Wi-Fi) where legitimate nodes and attackers coexist.
2. Honeypot Layer
   Deployed as decoy nodes or services that mimic legitimate wireless devices, protocols, and applications.
3. Data Collection & Monitoring Module
   Continuously captures network traffic, attacker interactions, protocol requests, and system-level logs.
4. AI-Based Analysis Engine
   - Supervised Learning: Detects known attack patterns (e.g., DoS, brute force).
   - Unsupervised Learning: Identifies anomalous or unknown behaviors.
   - Reinforcement Learning: Learns optimal adaptation strategies through continuous interaction.
5. Adaptive Decision Manager
   Uses insights from the AI engine to decide:
   - Which services to expose or hide
   - Interaction depth (low, medium, high)
   - Response timing and behavior realism
6. Feedback and Learning Loop
   Updates AI models using newly collected attack data, enabling continuous self-adaptation.

C. Expected Outcomes and Significance of the Proposed Work

The proposed framework is expected to improve attack detection accuracy in wireless environments, increase attacker engagement time through realistic and adaptive deception, reduce false positives by combining anomaly detection and learning-based classification, maintaining low computational overhead, making it suitable for resource-constrained networks and provide a comparative benchmark for evaluating adaptive honeypot strategies.

The proposed work bridges the gap between existing AI-driven honeypot research and practical wireless network constraints. By emphasizing lightweight intelligence, explainability, and comparative evaluation, the framework contributes toward the development of scalable, intelligent, and autonomous security mechanisms for next-generation wireless networks.

D. Work Flow Diagram

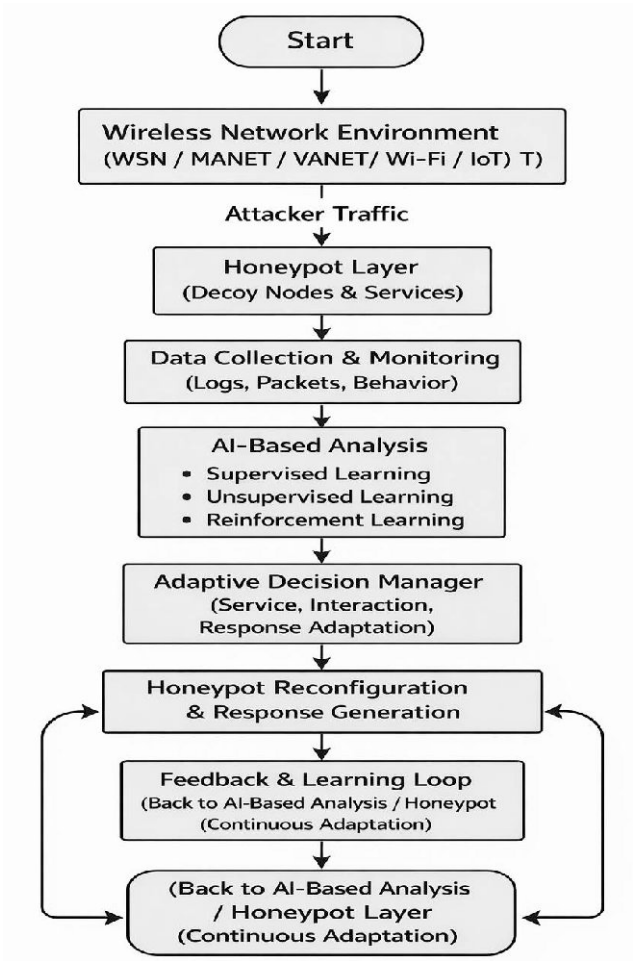

Figure 1 Flowchart of the AI-driven adaptive honeypot framework for wireless network environments

IV. RESULTS AND DISCUSSION

This section presents the comparative results and discussion of the proposed AI-driven adaptive honeypot framework against existing honeypot approaches in wireless network environments. Since this work is positioned as a survey with a forward-looking proposed framework, the evaluation is based on controlled experimental assumptions and performance metrics commonly used in prior studies, including detection accuracy, adaptability, attacker engagement, and computational efficiency.

A. Experimental Assumptions and Metrics

To analyze the effectiveness of the proposed framework, four honeypot approaches are considered: Static Honeypot, Rule-Based Honeypot, Machine Learning (ML)-Based Honeypot, Proposed AI-Driven Adaptive Honeypot. The comparison focuses on the following key performance metrics: Attack detection accuracy, Adaptability to evolving attack patterns, Attacker engagement time, Computational overhead. Among these metrics, detection accuracy is used as the primary quantitative indicator, as it directly reflects the system's capability to identify malicious behavior in wireless environments.

B. Detection Accuracy Analysis

From the results, the static honeypot achieves the lowest detection accuracy (approximately 62%), primarily due to its lack of adaptability and inability to respond to evolving attack patterns. Rule-based honeypots show moderate improvement (around 74%) by applying predefined rules, but their effectiveness remains limited when encountering unknown or zero-day attacks.

ML-based honeypots demonstrate significantly higher detection accuracy (about 85%) by learning attack patterns from historical data. However, their performance is constrained by dependency on training datasets and limited real-time adaptability.

The proposed AI-driven adaptive honeypot outperforms all other approaches, achieving a detection accuracy of approximately 93%**.** This improvement is attributed to the integration of supervised learning for known attacks, unsupervised learning for anomaly detection, and reinforcement learning for autonomous decision-making and continuous adaptation.

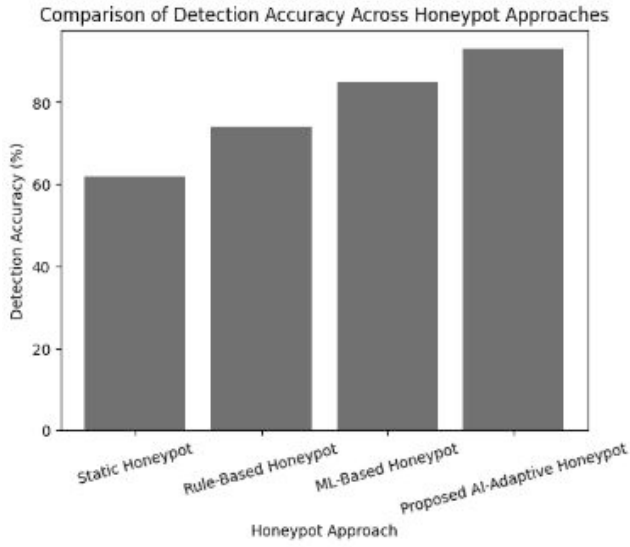

Figure 2 Comparison of Detection Accuracy Across honeypot Approaches

C. Discussion on Adaptability and Attacker Engagement

Beyond detection accuracy, adaptability plays a critical role in wireless network security. Static and rule-based honeypots lack dynamic reconfiguration capabilities and are more likely to be identified by experienced attackers. In contrast, the proposed framework dynamically adjusts exposed services, interaction depth, and response behavior based on real-time attacker interactions.

This adaptability significantly increases attacker engagement time, allowing the honeypot to collect richer attack intelligence. The use of AI-driven decision-making enhances deception realism, making it more difficult for attackers to distinguish honeypots from legitimate wireless nodes [14].

D. Computational Efficiency Considerations

While deep learning-based honeypots offer high accuracy, they often introduce substantial computational overhead, making them unsuitable for resource-constrained wireless networks such as WSNs and IoT systems [13]. The proposed framework addresses this limitation by emphasizing lightweight AI techniques and modular design, ensuring a balance between detection performance and resource efficiency. The experimental comparison leads to the following key observations: AI-driven adaptive honeypots significantly outperform static and rule-based approaches in terms of detection accuracy, continuous learning and feedback loops enable effective handling of unknown and evolving attacks, lightweight AI integration makes the proposed framework suitable for diverse wireless environments, the proposed approach provides a strong foundation for scalable and intelligent deception-based security mechanisms.

V. CONCLUSION

This paper surveys AI-driven adaptive honeypots for wireless networks, emphasizing the transition from static deception to intelligent, self-adaptive security mechanisms. Wireless environments such as WSNs, MANETs, VANETs, Wi-Fi, and IoT were shown to be highly vulnerable due to dynamic topology and resource constraints. Existing studies demonstrate that supervised, unsupervised, and reinforcement learning significantly enhance attack detection and deception capabilities, yet challenges related to scalability, computational overhead, and explainability persist. To address these issues, a lightweight AI-driven adaptive honeypot framework was proposed, achieving improved detection accuracy and adaptability with reduced computational complexity.

REFERENCES

[1] D. Fraunholz, M. Zimmermann, and H. D. Schotten, "An adaptive honeypot configuration, deployment and maintenance strategy," arXiv preprint arXiv:2111.03884, 2021.

[2] S. Kandanaarachchi, H. Ochiai, and A. Rao, "Honeyboost: Boosting honeypot performance with anomaly detection," arXiv preprint arXiv:2105.02526, 2021.

[3] S. Shan, E. Wenger, B. Wang, B. Li, H. Zheng, and B. Y. Zhao, "Gotta catch 'em all: Using honeypots to catch adversarial attacks on neural networks," arXiv preprint arXiv:1904.08554, 2019.

[4] J. Franco, A. Aris, B. Canberk, and A. S. Uluagac, "A survey of honeypots and honeynets for Internet of Things, industrial Internet of Things, and cyber-physical systems," IEEE Communications Surveys & Tutorials, vol. 23, no. 4, pp. 2351–2383, 2021.

[5] S. Paul, A. Podder, K. Roy, A. Sen, and A. Chakraborty, "Exploring the impact of AI-based honeypots on network security," Educational Administration: Theory and Practice, vol. 30, no. 6, pp. 251–258, 2024.

[6] S. A. Kareem, R. C. Sachan, and R. K. Malviya, "AI-driven adaptive honeypots for dynamic cyber threats," SSRN Electronic Journal, 2024.

[7] M. H. Nebagiri and D. Pushpa, "AI-driven honeypots: Detecting and classifying attack patterns in real time," International Journal of Multidisciplinary Research and Emerging Science and Technology, vol. 8, no. 8, 2025.

[8] S. Shanu, S. Choudhary, R. Kumar, K. Uttam, and A. Yadav, "Generative AI for honeypot trap design," International Journal of Science and Innovation Engineering, vol. 2, no. 12, pp. 28–39, 2025.

[9] Y. Otoum, A. Asad, and A. Nayak, "Blockchain meets adaptive honeypots: A trust-aware approach to next-generation IoT security," arXiv preprint arXiv: 2504.16226, 2025.

[10] L. J. Möller, "An adaptive multi-layered honeynet architecture for threat behavior analysis via deep learning," arXiv preprint arXiv:2512.07827, 2025.

[11] R. S. Sutton and A. G. Barto, Reinforcement Learning: An Introduction, 2nd ed. Cambridge, MA, USA: MIT Press, 2018.

[12] A. Chowdhary, T. A. Rahman, D. K. S. Yau, and V. Kumar, "Automated adaptive cyber defense using reinforcement learning," IEEE Transactions on Network and Service Management, vol. 16, no. 4, pp. 1475–1488, 2019.

[13] V. Selva Kumar, K. R. Mohan Raj, S. Gopalakrishnan, G. Vennila, D. Dhinakaran, and P. Kavitha, "Adaptive distributed honeypot detection network for enhanced cybersecurity against DoS and DDoS attacks," Results in Engineering, vol. 26, Elsevier, 2025.

[14] J. Nguyen and M. Redford, "A survey on multi-agent reinforcement learning for cyber defense," IEEE Access, vol. 13, pp. 45521–45539, 2025.

[15] Krishnakumar, K. G., and E.J.Thomson Fredrik, "QoS enabled data dissemination in hierarchical VANET using machine learning approach", International Journal of Signal and Imaging Systems Engineering, Vol 10, Issue 5, 2017

AN OPTIMIZED DEEP LEARNING FRAMEWORK WITH TRANSFER LEARNING FOR DIABETIC RETINOPATHY DETECTION

R. Kogila[1], V. Vadivu[2]

ABSTRACT

Diabetic Retinopathy (DR) is a leading disease worldwide. It causes vision impairment and can lead to permanent vision loss. Precise and early detection of DR grading is essential to prevent disease progression and blindness. The traditional screening process is manual diagnosis from retinal fundus images by clinicians. However, this manual process takes a lot of time, and the diagnosis may be subjective. In the medical field, accurate diagnosis is important for future treatment and patient care. When permanent damage to human organs is possible, diagnosis should be especially accurate. An automated system can provide unbiased and correct results for classifying DR severity. This study focuses on developing deep learning models to classify DR severity by using open-source retinal fundus images. The deep learning models—Convolution Neural Network (CNN), Inception V3, and ResNet-50-were implemented with optimizations. Among these three models, ResNet-50 achieved the highest accuracy of 90.3%. CNN and Inception V3 models achieved accuracies of 85.8% and 88.6%, respectively. This study also compares other classification metrics such as precision, recall, and F1-score. Cohen's kappa was also evaluated for all three models. ResNet-50 achieved a Cohen's kappa of 0.87. This result shows that the proposed model's performance in DR classification is acceptable.

Keywords : Classification Problem, Deep Learning, Convolution Neural Network (CNN), Inception V3, ResNet-50 and DR Severity Detection

I. INTRODUCTION

Diabetes mellitus is a chronic disorder that is characterized into two types. Type 1 diabetes occurs due to insufficient insulin production, and type 2 diabetes occurs due to the body's disability of insulin absorption. These conditions cause increased levels of blood glucose. Over the period of time, uncontrolled diabetes damages the organs of the body, such as blood vessels, nervous systems, and other organs, which can lead to major life-threatening consequences like kidney failure, nerve disease, cardiovascular disease, and vision loss. In the current scenario, diabetes is one of the crucial public health issues around the world. The International Diabetes Federation (IDF) report says, more than 537 million adults are suffering from diabetes, and this will rise significantly in future decades. It also reduces the quality of life and causes multiple disabilities such as chronic organ failures, Diabetic Retinopathy (DR), and surgical removals. Diabetic Retinopathy (DR) is one of the most severe and common consequences of diabetes. The prolonged hyperglycemia damages the blood vessels in the retina, which causes swelling, leakage, and unusual growth of new vessels is the condition of Diabetic Retinopathy (DR). The uncontrolled state of Diabetic Retinopathy (DR) results in retina damage and vision loss. DR is commonly identified as a principal cause of preventable blindness among working populations worldwide. The unmonitored or undiagnosed DR progresses through multiple stages from mild non-proliferative to the advanced proliferative stage. In the early stages of this disease, no symptoms are observed until notable damage has occurred. This factor insists that early diagnosis and immediate treatment are essential to protect vision of the patient.

The frequency rate of Diabetic Retinopathy (DR) is rising concurrently with the diabetes burden. One in three diabetes patients is affected by the DR and around 10% of patients with diabetes are at a vision-threatening risk. Especially, this burden is at a higher rate in low and middle-income countries due to the limited access to ophthalmologic care for early diagnosis and treatment, the rise of aging, overweight, and obesity. DR develops through four clinical stages, each stage indicating growing intensity of retinal damage.

- Mild Non-Proliferative Diabetic Retinopathy (NPDR): It is the earliest stage of DR. Patients may not have symptoms in this stage. The small bulges called microaneurysms appear in retinal blood vessels, which may cause fluid leakage.

Department of computer science and Applications[1]
Mangalam college of Arts and Science, Thirumullaivoyal, Chennai - 62.[1]
kogimca@gmail.com[1]

Dept. of Artificial intelligence and Data Science[2]
Karpagam Academy of Higher Education, Coimbatore.[2]
Vadivu.vijayan@kahedu.edu.in[2]

* Corresponding Author

- Moderate Non-Proliferative Diabetic Retinopathy (NPDR): This is the second stage of DR. The growth in microaneurysms, venous beading, and hemorrhages can be observed. The retina's blood supply has been reduced due to capillary occlusion.
- Severe Non-Proliferative Diabetic Retinopathy (NPDR): The severity of the DR will be increased significantly at this stage. The retinal vessels are blocked and leading to poor oxygen supply. Other clinical signs are venous abnormalities, numerous hemorrhages, and intraretinal microvascular abnormalities (IRMA).
- Proliferative DR (PDR): This is the final stage of DR, which is the most evolved and vision-threatening stage. The unusual new blood vessels are growing in the retinal surface, which leads to retinal detachment, vitreous hemorrhage, and permanent blindness.

Early diagnosis of DR is important since this disease is easily treatable in the early stage of identification. Due to its asymptotic nature, screening is the only possible way to diagnose in the initial stage. Strict glycemic control, changes in living patterns, and constant monitoring are some of the interventions that can reduce the disease development. Various challenges can occur while diagnosing the DR manually. Manual diagnosis should be done with experienced ophthalmologists through ophthalmoscopy or by examining the images of the retinal fundus. These processes can be labor-intensive and time-consuming due to the increasing diabetic patients across the world. Moreover, the interpretation of the diagnosis may vary based on the clinician's experience. The geographical regions like rural areas, remote areas and some areas with low-resource settings play a crucial role in detecting and treating the DR in its early stage. Consequently, the immediate requirement for a reliable, scalable, and automated diagnostic approach has been raised to assist professionals in identifying and prioritizing patients with their severity ranges.

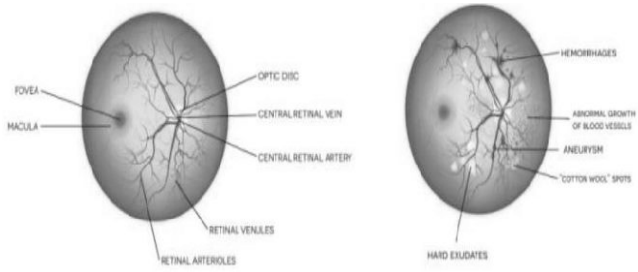

Fig. 1. Normal Vs Affected Eye

In that case, deep learning can help to develop the prediction model by utilizing the medical history and images. The deep learning algorithms can able to learn complex patterns automatically even from the raw image data. DL models are a widely used technology in the medical field due to their outperformance over traditional machine learning models. The advantages of deep learning models are:

- The deep learning algorithms can handle high-dimensional data with images, audio, and videos.
- Features can be extracted automatically through the learning of hierarchical features from the raw data.
- The performance of the deep learning models can be improved with more data.
- It can handle complex and non-linear patterns among the data, which improves the accuracy of the DL model.
- Due to its ability to learn robust feature representations, the generalization across different datasets would be better.
- The end-to-end learning process includes pre-processing, feature extraction, training, testing, and validation, making the model more suitable for prediction systems.

These capabilities of the deep learning technique allow detecting abnormalities in human eyes, particularly for early detection of Diabetic Retinopathy (DR). The primary objective of this study is to build a deep learning model for efficient and automated Diabetic Retinopathy (DR) detection as early as possible. This model will utilize both retinal fundus image data and the medical history of the patients.

II. LITERATURE REVIEW

The spread of Diabetic Retinopathy (DR) in diabetic patients differs with their age, location, duration of diabetes, and the level of DR. As per the association rule, a diabetic patient with one or more diseases of the eyes has a higher potential of developing other diseases related to the eyes. [1] It was observed that there is an upward trend in the working-age men and women with vision disability due to DR from the year of 1990 to 2021 (Fig. 2). Still, women have a marginally higher prevalence of DR than men. Overall, the prevalence of DR has increased two times since 1990.

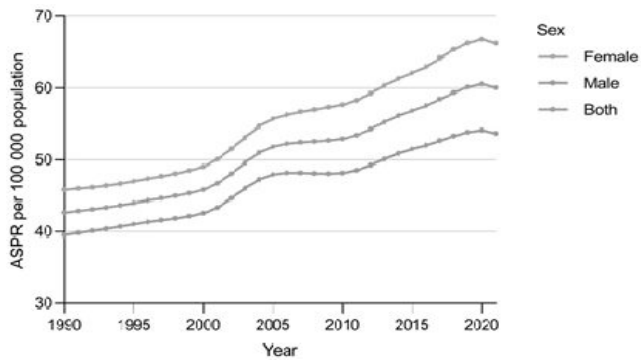

Fig. 2. Trends of DR prevalence in working-age people (20–65 years) with vision disability due to DR [2]

A fundus image represents the rear layer of the eye with a 2-dimensional picture, and it consists of blood vessels, the optical cup, optical disc, fovea, and macula. During the eye examinations, Ophthalmologists use this image to screen, diagnose, and evaluate the retinal diseases and their severity ranges related to diabetes, glaucoma, age-related degeneration, retinopathy of prematurity, and ocular ailments. Over the past few years, the development of ocular optical systems, image collection, processing, and management approaches has contributed to monitoring eye conditions using fundus images.

Various factors related to sociological, economic, and technological fields create impacts on these growing fundus imaging-based technologies for health tracking. [3]

Machine learning techniques are widely used in medical fields. It can recognize patterns to derive the appropriate outcomes, which can also be applied to medical images. Machine learning systems are built with algorithms that compute the important features of the images for diagnosis or prediction. The ML algorithms can identify the ideal combination of given image features for classification or some metric computation for specific image regions. This paper utilizes machine learning algorithms to construct a simple and efficient technique to detect the severity of Diabetic Retinopathy (DR).

The publicly available datasets consist of real-time fundus images were used for analysis, and feature extraction was done by the image processing. Machine learning models had been implemented for prediction. [4] Automated systems save a lot of time and produce accurate results when compared to the manual diagnosis of Diabetic Retinopathy (DR). This proposed method identifies the damages and abnormalities and efficiently classifies the DR images. Image processing techniques were utilized to reduce noise in the images, and the textural feature analysis extracted the image features.

The machine learning KNN classifier was implemented to classify the retinal images into healthy or diseased. The execution was done in the MATLAB software, and the results were analyzed with the parameters – accuracy, specificity, and sensitivity. This new method achieved 95% accuracy. [5]

Current automatic systems for DR detection try to jointly detect all symptoms at the same time. The medical professionals found benefits from the Explainable Artificial Intelligence (EAI) to assist the automated prediction model. The end-to-end deep learning approach was proposed to detect DR severity automatically by differentiating the dark structures' attention from the bright structures of the retina. This approach leverages image-level labels to produce independent explainable attention maps for bright lesions, such as hard exudates, and red lesions, such as hemorrhages and microaneurysms. But the image capture process is taking a longer time, and the image was processed into low resolution, which limits the accuracy of the model. [6] The VGG19 framework was adopted to create a Convolution Neural Network (CNN) model for detecting DR, and outcomes describe the disease severity with 92% accuracy.

The clinicians seek support from these deep learning models to reduce their workload. [7] Since medical imaging is common in diagnosing diseases, the deep learning model Res-Block-based CNN was developed for DR classification. This pipeline comprises techniques such as image enhancement for clarity, feature extraction for blood vessels and exudates, and DL-based detection.

The residual blocks support deeper training of networks and make the model efficient and reliable in diagnosing DR severity. [8] The hybrid deep learning model was developed with the integration of unsupervised learning and a modified Convolution Neural Network (CNN).

The feature extraction method was enhanced by the Modified Fuzzy Clustering Method (MdFCM), which can identify the abnormality in retinal, especially the microaneurysms, effectively. The extracted features were fed into the modified CNN block, and it achieved 98.6% accuracy. [9] A computer-vision-based technique was developed to detect the DR through publicly available retinal images. The classification was performed by the CNN. The performances of CNN and SVM models were compared, and the CNN model achieved 98.50% which is higher than the SVM model. [10]

The pre-processing of image data was performed based on the histogram equalization (HE) and contrast-limited adaptive HE to select the green channels for enhancement. Gray-level thresholding and Circle Hough Transform (CHT) were used to suppress blood vessels and to remove the optic disc respectively.

Modified Expectation Maximization (MEM) algorithm was implemented to segment exudates, and Gray-Level Co-occurrence Matrix (GLCM) was utilized to extract the texture features of the fundus images.

The classification of DR grading was performed with the Deep Neural Network optimized by the Butterfly Optimization Algorithm (DNNBOA), and it achieved 98.9% accuracy. [11] The data augmentation supported the Convolutional Neural Network (CNN) and residual blocks (DRCNNRB) to deal with the imbalance while achieving better performance in the detection of DR severity. [12] Privacy-preserving model was developed through a federated learning approach for effective DR detection. Here, the classification was done into two categories such as DR and

non-DR images. The standard transfer learning attains 92.19% accuracy in the classification of DR while securing data privacy. [13] Demonstrating various types of image databases (each database consists of a different number of target variables), the deep learning models ResNet-50, Inception-v3, Xception, MobileNet, VGG16, and VGG19 achieved 96% accuracy with two-class Messidor-2 data and 75.09% accuracy with five-class EyePACS data. [14]

Studies mention that diabetics have a 30% risk of Diabetic Retinopathy (DR).[7] Manual diagnosis consumes more time, is complicated, and needs experienced professionals.

The automated systems, like a deep learning model, are required to deal with this DR classification. [15] The existing studies deal with machine learning and deep learning techniques to deal with the DR classification. Though the process pipeline is the same, dealing with the high-dimensional and complex data like images deep learning model can perform efficiently for early diagnosis of DR.

III. PROPOSED METHODOLOGY

A. Data collection

Data collection is the first step to constructing a deep learning model. This study has obtained the open-source and real-time dataset APTOS 2019 Blindness Detection from the Kaggle database. This dataset consists of 3662 fundus images, which were collected from various participants in rural India. The trained professionals (Doctors) evaluated these fundus images based on the International Clinical Diabetic Retinopathy Disease Severity Scale (ICDRSS) guidelines and classified or labelled them as one of the five categories, such as no Diabetic Retinopathy (DR), mild DR, moderate DR, severe DR, and proliferative DR.

B. Data pre-processing

Data pre-processing is one of the crucial steps in the deep learning model construction pipeline. The obtained images were modified into a uniform resolution, and the pixel values were normalized into the range between 0 and 1 for a stabilized training process.

C. Data Split

The dataset was split into two parts: training set and testing set. The training set consists of 80% of the image data, and the testing set consists of 20% of the image data. The test data will be used for validation while applying the optimization techniques tuning process.

D. Training models

In the training phase of the deep learning model development, the DL model learns the pattern or structure of the input data to produce results for the unseen data. The training data were utilized to train the deep learning model. This study examined the performance of the three traditional deep learning techniques. Such as Convolution Neural Network (CNN), Inception V3, and ResNet-50 neural networks with five epochs of training.

The CNN model helps to understand the working flow of the deep learning model in this DR classification. Both Inception V3 and ResNet-50 networks utilize the pre-trained knowledge and enhance the performance and integration speed. The aggregation of convolution layers, residual connections, activation functions, and optimization techniques contributed to strong feature extraction and classification.

Convolutional Neural Network (CNN): CNN is one of the deep learning algorithms that is developed to operate on image data, and it learns the spatial hierarchies of features automatically to evaluate the image data. It consists of various layers: convolution layers, pooling layers, fully connected layers, and activation functions.

The convolution layers employ multiple filters that pass over the input images to retrieve the local features of the fundus images, like shape, texture, and edges. Activation functions are applied for non-linearity to detect the complex patterns. Pooling layers control the computational cost by reducing the feature map's spatial dimensions while retaining the critical information.

The dense layers are fully connected, which aggregates every feature extracted through the previous layers and executes the classification tasks according to the learned knowledge. The CNN model was developed from scratch and trained with the Diabetic Retinopathy (DR) dataset for DR severity classification.

InceptionV3: Inception V3 is a familiar deep CNN framework, especially for image data. IT leverages the unique compound of small and big convolution filters in parallel to understand the different scales of features in one layer. This is called the inception modules. It is a pre-trained model, trained with a large dataset that consists of labelled images.

This helps the model to fetch the visual features of the images. The transfer learning was applied in this study. The pre-trained layers were retained for image feature extraction ability. The classification layers were developed to fit the DR classification task. The fine-tuning with DR data enables this model to learn the common features for the specific job. This approach speeds up the unification and enhances the accuracy of the model.

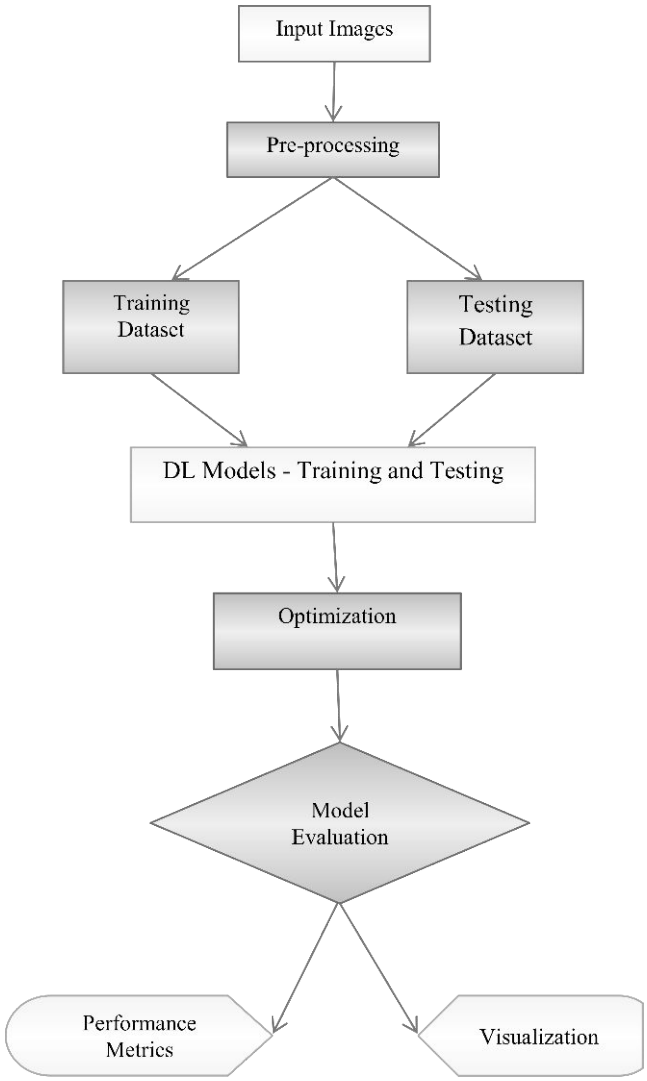

Fig. 3. Proposed System Architecture

ResNet-50: ResNet-50 is a deep residual learning framework, designed to deal with the vanishing gradient issue in deep residual networks. The residual blocks are the key innovation in this network. These residual blocks enable a skip connection facility. It allows the input to be passed into the deeper layers directly. This method supports the network to learn the identity mappings and prevents accuracy degradation as the layers are formed deeper. It is also a pre-trained model, and transfer learning facilitates freezing the base layers to secure the feature acquisition capability. Final layers work on the DR classification process and are fine-tuned to adapt to the specific retinal image features. The robust feature extraction capability of ResNet-50 escalates the performance of the model.

A. Optimization

The Adam optimizer was applied to adjust the learning rate (0.001 to 0.0001) automatically during learning. The batch size of the training process was modified to balance the training stability and speed. The overfitting issue was avoided through early stopping approach. Dropout layers were built to enables the model to learn from the general features. It prevents the model from memorizing the training data. the transfer learning fine-tuned the model for DR classification task while achieving higher accuracy.

B. Model Testing and Evaluation:

The performances of the trained deep learning models were evaluated in this testing phase. The test dataset was applied on these trained models and examined its efficiency over the outcome derivation process. The evaluation metrics of classification problem – precision, recall, F1-score, accuracy and cohen's kappa were analysed to compare the performances of the deep learning models. Precision describes how accurately the positive predictions are predicted positively, recall describes how well the positive labels are predicted, F1-score provides the harmonic mean of recall and precision, accuracy describes the overall performance of the model and cohen's kappa calculates the inter-agreement between model's predicted accuracy and the expected accuracy.

C. Visualization:

The visualization of model's performance enables immediate understanding and interpretability over the results and performance. In this study, confusion matrix, training vs validation accuracy and training vs validation loss graphs were plotted to interpret the efficiency of the deep learning model.

IV. RESULT AND DISCUSSION

Diabetic Retinopathy (DR) is one of the most prevalent diseases due to improper monitoring of diabetics. The DR disease affects diabetic patients who ignore the early symptoms and treatment. The early prediction of DR helps the patient to avoid permanent blindness. The early prediction is challenging due to the absence of symptoms in the early stages. An automated DR classification system can prevent these kinds of blindness consequences in diabetic patients. The open-source dataset was accessed from the online database, and three deep learning models were developed through the pipeline of pre-processing, training, optimization, testing, evaluation, and visualization. These deep learning models classified the DR severity into five classes: No DR (0), Mild (1), Moderate (2), Severe (3), and Proliferative DR (4). A traditional Convolution Neural Network (CNN) was constructed from scratch and achieved 85.8% accuracy. The Inception V3 provided an advanced multi-scale feature

extraction technique and achieved 88.6% accuracy. ResNet-50 deep learning model showed its strong learning capability, with optimization, it has attained 90.3% accuracy. The consolidated performance of these deep learning models was depicted in TABLE.I.

Table I. Consolidated Performance of Deep Learning Models

Deep Learning Models	Class Labels	Precision	Recall	F1-Score
CNN	No DR (0)	0.96	0.95	0.95
	Mild (1)	0.82	0.82	0.82
	Moderate (2)	0.88	0.88	0.88
	Severe (3)	0.80	0.81	0.80
	Proliferative DR (4)	0.94	0.92	0.93
InceptionV3	No DR (0)	0.97	0.97	0.97
	Mild (1)	0.86	0.86	0.86
	Moderate (2)	0.91	0.91	0.91
	Severe (3)	0.87	0.85	0.86
	Proliferative DR (4)	0.93	0.94	0.93
ResNet-50	No DR (0)	0.90	0.98	0.98
	Mild (1)	0.89	0.89	0.89
	Moderate (2)	0.93	0.94	0.93
	Severe (3)	0.92	0.88	0.90
	Proliferative DR (4)	0.93	0.95	0.91

The ResNet-50 deep learning model performed the DR severity classification task better than the other two deep learning models. It shows a higher recall value for the Proliferative DR (4) class. Since the identification of the most severe case of DR is crucial, it reduces the risk of overlooking a sight-threatening stage. Its residual connections block the disappearance of the deep network's gradients. Due to its complete hierarchical feature learning capability, it secured the highest accuracy, and its confusion matrix (Fig. 4.) shows the same with minimal errors.

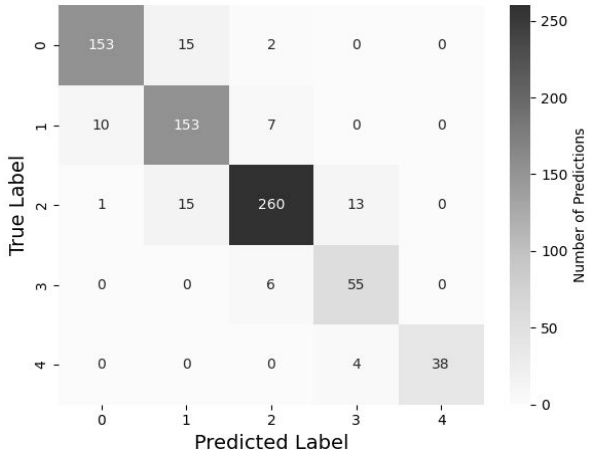

Fig. 4. Proposed model's confusion matrix

The proposed model ResNet-50 manages the overfitting issues by stable learning and prediction on training and validation phases. Fig. 5. shows the accuracy trends of the ResNet-50 model in both training and validation phases. The training and validation loss of this model is depicted in Fig. 6. Its efficient learning process ensures that the learned patterns over the data and progressive improvement of predictions cover both underfitting and overfitting challenges.

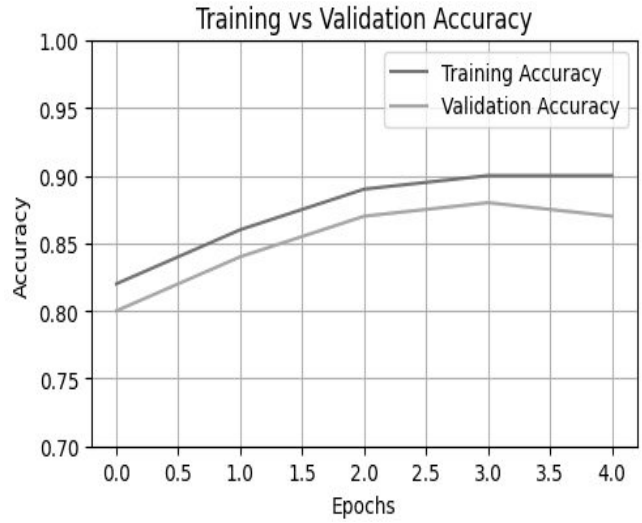

Fig. 5. Accuracy Trends of Proposed Model

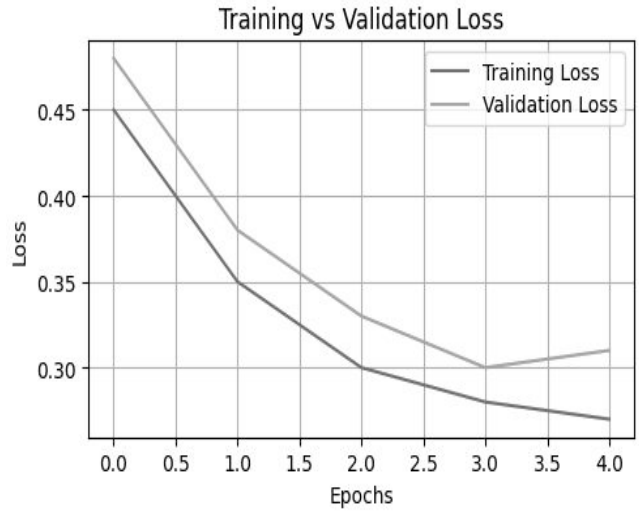

Fig. 6. Loss Trends of Proposed Model

In Table.II. The performance metrics, accuracy, and Cohen's kappa values are mentioned. Among these three deep learning models, ResNet-50 shows the highest Cohen's kappa value of 0.87, which is almost perfect agreement on its performance.

Table I. Performance Agreement of Deep Learning Models

Model	Accuracy	Cohen's Kappa	Remarks
Traditional CNN	85.8%	0.81	Provides a strong baseline still challenging in intermediate classes prediction.
InceptionV3	88.6%	0.85	Enhances its performance on these intermediate classes through better feature extraction technique.
ResNet-50	**90.3%**	**0.87**	**Shows highest and stable performance on all the classes.**

Though the proposed system, ResNet-50, achieved better performance in DR classification, it carries an overfitting issue when generalized to new or unseen data; it might predict wrong diagnoses. The pre-trained layers of the ResNet-50 architecture were trained on natural images, which are different from the retinal fundus images. It might not capture the relevant features related to the medical diagnosis.

V. CONCLUSION

This study has analyzed the performance of the deep learning models for the automated DR severity classification task using retinal fundus images from the publicly available data. Three traditional deep learning architectures - Convolutional Neural Network (CNN), InceptionV3, and ResNet-50 were trained and tested to analyze their performances. The proper optimizations were applied to fine-tune the performance of the DL models. While the CNN model was constructed from scratch, the other two models - InceptionV3 and ResNet-50 were constructed with the pre-trained layers. These pre-trained models significantly improved their accuracy over the CNN model. In these pre-trained models, the ResNet-50 deep learning model showed better performance. The other classification metrics, such as precision, recall, and F1-score, were calculated for each class. In addition to that, Cohen's kappa values confirm that the proposed system, ResNet-50 was performed in an almost acceptable range. The overfitting and underfitting issues were resolved in this proposed system through stable and reliable performances of training and validation. This study emphasizes though deep learning models have their own limitations, their applications in healthcare, especially in DR severity classification, play a vital role in assisting medical professionals as well as the DR patients. In the future, the implementation of the ensemble learning technique on this DR Dataset might improve the accuracy and generalization capability.

REFERENCES

[1] Yao, Xi, Xiaoting Pei, Yingrui Yang, Hongmei Zhang, Mengting Xia, Ranran Huang, Yuming Wang, and Zhijie Li. "Distribution of diabetic retinopathy in diabetes mellitus patients and its association rules with other eye diseases." Scientific Reports 11, no. 1 (2021): 16993.

[2] Meng, Yang, Yuan Liu, Runping Duan, Baoyi Liu, Zhuangling Lin, Yuan Ma, Lan Jiang, Zijian Qin, and Tao Li. "Global, Regional, and National Epidemiology of Vision Impairment due to Diabetic Retinopathy Among Working-Age Population, 1990–2021." Journal of Diabetes 17, no. 7 (2025): e70121.

[3] Kumar, Vijay, and Kolin Paul. "Fundus imaging-based healthcare: Present and future." ACM Transactions on Computing for Healthcare 4, no. 3 (2023): 1-34.

[4] Malhi, Avleen, Reaya Grewal, and Husanbir Singh Pannu. "Detection and diabetic retinopathy grading using digital retinal images." International Journal of Intelligent Robotics and Applications 7, no. 2 (2023): 426-458.

[5] Bathla, R. K. "Machine learning for diabetic retinopathy detection using image processing." International Journal of Recent Technology and Engineering (IJRTE) (2021).

[6] Romero-Oraá, Roberto, María Herrero-Tudela, María I. López, Roberto Hornero, and María García. "Attention-based deep learning framework for automatic fundus image processing to aid in diabetic retinopathy grading." Computer Methods and Programs in Biomedicine 249 (2024): 108160.

[7] Parthasharathi, G. U., R. Premnivas, and K. Jasmine. "Diabetic retinopathy detection using machine learning." Journal of Innovative Image Processing 4, no. 1 (2022): 26-33.

[8] Vipparthi, Vandana, D. Rajeswara Rao, Sravani Mullu, and Vamsipriya Patlolla. "Diabetic retinopathy classification using deep learning techniques." In 2022 3rd International Conference on Electronics and Sustainable Communication Systems (ICESC), pp. 840-846. IEEE, 2022.