An Adaptive Attention-Driven Deep Learning Framework with Optimized Segmentation for Multi-Class Diabetic Eye Disease Diagnosis

JOHN JOSEPH MURIKIPUDI[1], A.V. NAGESWARA RAO[2], K. RAJASEKHAR[3]
[1]Department of ECE, Jawaharlal Nehru Technological University,
Kakinada,
INDIA

[2]Department of ACSE, Vignan University,
Vadlamudi, Guntur,
INDIA

[3]Department of ECE, Jawaharlal Nehru Technological University,
Kakinada,
INDIA

Abstract: - Early identification of ocular diseases is essential to preventing irreversible vision loss, particularly among diabetic patients who are highly susceptible to multiple retinal complications. Although several automated approaches have been developed in recent years, most are limited to single-disease detection or depend heavily on conventional segmentation and fixed-parameter models, which often fail to capture subtle and overlapping abnormalities in fundus images. To address these gaps, this research proposes an integrated framework that combines adaptive segmentation, intelligent optimisation, and attention-driven deep learning for multi-class ocular disease classification. The study begins by gathering retinal fundus images from widely used public repositories and enhancing them to improve structural visibility. An Adaptive SegUNet model is then employed to segment clinically relevant regions, while its parameters are automatically refined using the Improved Mother Optimization Algorithm to achieve more stable and precise segmentation. The segmented images are subsequently analysed through an Attention-based Efficient Deep Learning network incorporating a Gated Recurrent Unit enabling the model to learn discriminative spatial and sequential patterns associated with different ocular disorders. Finally, the performance of the proposed system is evaluated and compared with existing methods to demonstrate its effectiveness. The results highlight the potential of the ASUNet–IMOA–AEDGRU pipeline as a reliable and robust solution for multi-disease retinal screening, offering valuable support for clinical diagnosis and early intervention.

Key-Words: - Adaptive SegUNet (ASUNet), Improved Mother Optimization Algorithm (IMOA), Attention-based Deep Learning, Gated Recurrent Unit (GRU), Retinal Fundus Images, Ocular Disease Classification, Medical Image Segmentation, Multi-class Diabetic Eye Disease Detection, Computer-Aided Diagnosis

Received: March 23, 2026. Revised: May 22, 2026. Accepted: June 15, 2026. Published: July 15, 2026.

1 Introduction

According to Diabetic eye diseases remain one of the leading causes of preventable blindness worldwide, particularly among individuals living with long-term diabetes. Conditions such as cataract, diabetic retinopathy, and glaucoma often develop gradually and may not exhibit noticeable symptoms until the disease has reached an advanced stage. As a result, early and accurate detection is essential to prevent irreversible visual impairment. Retinal fundus imaging has emerged as a widely used technique for routine screening because it enables clinicians to observe structural and vascular abnormalities non-invasively.

Figure 1 illustrates representative examples of common ocular diseases observed in fundus images. Cataract typically appears as a hazy or blurred retinal view due to lens opacity, diabetic retinopathy is characterized by lesions such as microaneurysms and hemorrhages, glaucoma often presents with optic nerve cupping, and normal eyes exhibit a healthy retinal structure without pathological disruptions. These variations highlight the diversity and

complexity of retinal presentations that automated diagnostic systems must interpret accurately.

Despite advances in computer-aided diagnosis, achieving reliable multi-class detection across these disease categories remains challenging. Traditional deep learning models often depend on fixed segmentation procedures and manually tuned parameters, which limit their ability to adapt to varying image qualities, noise levels, and lesion characteristics. Moreover, small yet clinically significant abnormalities—such as early microaneurysms or subtle optic nerve changes—may not be captured effectively by static or non-optimized segmentation techniques.

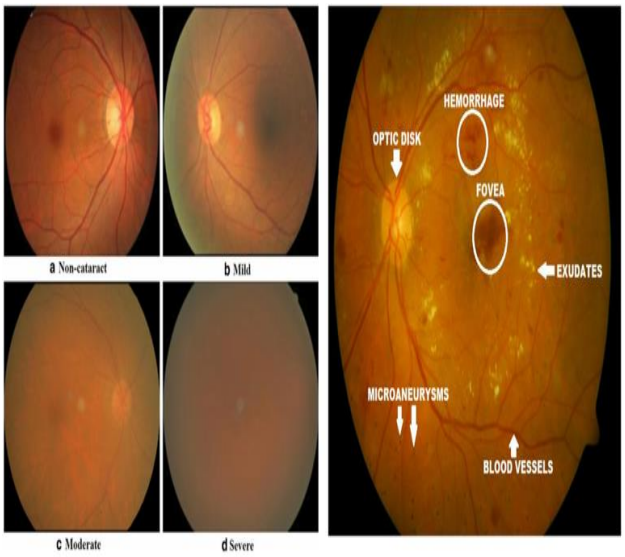

Fig. 1: Representative ocular fundus images showing different retinal conditions commonly associated with diabetic eye diseases

To overcome these limitations, this research introduces a unified and adaptive framework for multi-class ocular disease analysis. The proposed approach employs an Adaptive SegUNet (ASUNet) model to segment key retinal structures more precisely. To enhance segmentation accuracy and stability, the network parameters are automatically fine-tuned using the Improved Mother Optimization Algorithm (IMOA). Following segmentation, disease classification is performed using an Attention-based Efficient Deep Learning model integrated with a Gated Recurrent Unit (AEDGRU). The attention mechanism enables the model to focus on diagnostically relevant regions, while the GRU component learns sequential feature dependencies essential for distinguishing between similar visual patterns.

By combining adaptive segmentation, intelligent optimisation, and deep attention-driven classification, the proposed methodology aims to provide a robust and scalable solution for early screening of cataract, diabetic retinopathy, glaucoma, and normal retinal conditions. This integrated system has the potential to support clinical decision-making, reduce diagnostic workload, and contribute to more efficient and accurate population-level eye-health monitoring.

1.1 Motivation

Diabetic eye diseases continue to be a major cause of preventable blindness, especially in regions where regular screening and specialist consultation are limited. Early retinal abnormalities such as microaneurysms, vessel distortions, and optic nerve changes often appear subtle, making manual diagnosis challenging and inconsistent. Although deep learning has improved automated retinal analysis, many existing methods struggle with inaccurate segmentation, varying image quality, and poor generalisation across different datasets. These limitations directly affect the reliability of multi-class disease classification.

To address these challenges, there is a need for a more adaptive and optimised diagnostic framework. This motivates the integration of Adaptive SegUNet (ASUNet) with the Improved Mother Optimization Algorithm (IMOA) for enhanced segmentation, and the use of an Attention-based Efficient Deep Learning model with a Gated Recurrent Unit (AEDGRU) for robust classification. Together, these components aim to deliver a more accurate, flexible, and clinically useful system for early detection of ocular diseases.

1.2 Contributions

This paper presents several key contributions aimed at improving the reliability and accuracy of automated ocular disease diagnosis:

- An Adaptive SegUNet (ASUNet) architecture is introduced to achieve more precise segmentation of retinal structures. Its adaptability helps capture subtle lesions and structural variations that traditional segmentation methods often miss.
- The Improved Mother Optimization Algorithm (IMOA) is incorporated to automatically fine-tune ASUNet parameters, enhancing segmentation quality while reducing manual intervention and sensitivity to dataset variations.
- A novel classification framework based on Attention-enabled Efficient Deep Learning with a Gated Recurrent Unit (AEDGRU) is developed, allowing the model to focus on diagnostically

important regions and effectively learn discriminative features across multiple ocular diseases.

- A comprehensive evaluation is conducted comparing the proposed system with existing segmentation and classification techniques, demonstrating significant improvements in accuracy, adaptability, and robustness.
- An end-to-end diagnostic pipeline is established, offering a unified and automated approach to early detection of cataract, diabetic retinopathy, glaucoma, and normal retinal conditions.

2 Literature Review

The rapid growth of ophthalmic imaging technologies and the increasing burden of diabetic eye diseases have encouraged extensive research into computer-aided methods for reliable and early detection of retinal abnormalities. Fundus photography, being non-invasive and clinically well-established, has become the primary imaging modality used for screening conditions such as diabetic retinopathy, glaucoma, and cataract. Over the years, research efforts have evolved from traditional handcrafted feature extraction approaches toward sophisticated deep learning frameworks capable of automatically learning disease-discriminative representations. Despite significant progress, achieving robust performance across diverse datasets, imaging conditions, and disease severities remains an ongoing challenge, particularly when dealing with multi-disease classification environments where retinal abnormalities overlap in subtle and complex ways.

Early methods for retinal analysis predominantly relied on classical image-processing techniques. Researchers used thresholding, region-growing, morphological operations, Gabor filtering, and clustering-based segmentation to isolate anatomical structures such as blood vessels, optic discs, and exudates. These techniques offered interpretable results but were limited by their heavy dependence on carefully chosen parameters and handcrafted heuristics. Variations in illumination, noise levels, occlusions, and the presence of early-stage lesions frequently diminished the reliability of such conventional methods. For example, vessel segmentation approaches based on matched filtering or Hessian-based techniques performed well on high-contrast images yet struggled with distorted or low-quality scans. Similarly, fuzzy clustering and PCA-based optic disc detection methods were able to approximate the disc location in controlled conditions but showed reduced performance under heterogeneous population datasets. These limitations highlighted the need for more adaptive and data-driven approaches.

The introduction of machine learning in the early 2000s marked a transition toward automated classification pipelines that combined handcrafted feature extraction with supervised learning algorithms. Support Vector Machines, Random Forests, Decision Trees, k-Nearest Neighbors, and Genetic Algorithm-enhanced feature selectors were widely explored for diagnosing diabetic retinopathy and cataract severity. Although these models provided promising improvements, their success was still fundamentally tied to the quality of manually engineered features. Small inaccuracies in segmentation or feature extraction could propagate into the classification stage, resulting in false predictions. Furthermore, handcrafted methods lacked the flexibility to generalize across diverse imaging variations or capture the subtle and hierarchical patterns characteristic of retinal disease progression.

Deep learning, particularly Convolutional Neural Networks (CNNs), revolutionized the field by enabling automated hierarchical feature learning. CNN-based architectures such as AlexNet, VGG, ResNet, and Inception began outperforming classical machine learning approaches, particularly in large-scale retinal classification tasks. Several studies demonstrated the potential of CNNs to detect microaneurysms, hemorrhages, exudates, and optic nerve abnormalities, with some models achieving clinically comparable performance for binary or limited multi-class classification. However, despite their success, traditional CNNs faced notable drawbacks. Their effectiveness depended heavily on the quality of segmentation or pre-processing, and they were prone to focusing on irrelevant background information if prominent lesions were not sufficiently highlighted. Moreover, CNNs generally excelled at global pattern recognition but could overlook small, localized abnormalities — a critical concern in early-stage disease detection.

Recognizing the role of segmentation in improving diagnostic accuracy, researchers increasingly adopted encoder–decoder architectures, with U-Net becoming the dominant framework for medical image segmentation. U-Net's symmetric structure allows it to combine high-level semantic information with fine spatial details through skip connections. Variants of U-Net incorporating residual blocks, dense layers, attention mechanisms, and multi-scale modules further improved segmentation robustness. Nonetheless, even with these enhancements, U-Net-based models often relied on manually selected hyperparameters such as learning rates, filter depths, kernel sizes, and regularization settings. Manual

tuning is labor-intensive, and suboptimal hyperparameters can significantly degrade performance, especially on heterogeneous datasets. As a result, optimization techniques became increasingly integrated into segmentation workflows.

Bio-inspired optimization algorithms—such as Particle Swarm Optimization, Ant Colony Optimization, Grey Wolf Optimization, Whale Optimization Algorithm, and Genetic Algorithms—have been widely used to improve model training stability, feature selection, and parameter tuning. These algorithms mimic collective behavior patterns to search for optimal solutions within complex parameter spaces. In medical imaging, such strategies have been applied to refine segmentation thresholds, select discriminative feature sets, optimize CNN hyperparameters, and tune objective functions. Despite yielding improvements, many of these algorithms suffer from premature convergence, sensitivity to initial conditions, or high computational demands. Recent advancements therefore focused on hybrid and improved variants of metaheuristic algorithms capable of balancing exploration and exploitation more effectively. These developments paved the way for optimization-enhanced model refinement strategies like the Improved Mother Optimization Algorithm (IMOA), which introduces enhanced neighborhood search and adaptive mechanisms for more reliable convergence.

Parallel to developments in segmentation and optimization, classification methodologies have also evolved. Advanced deep learning models incorporating attention mechanisms have been shown to improve classification accuracy by guiding computational focus to discriminative retinal regions. Spatial attention modules selectively enhance meaningful lesion areas, while channel attention mechanisms help prioritize feature maps that contribute most to classification tasks. Such mechanisms emulate the way ophthalmologists visually inspect fundus images by concentrating on clinically relevant landmarks including the macula, optic disc, and vascular structures.

Beyond attention mechanisms, recurrent neural networks (RNNs), especially Gated Recurrent Units (GRUs) and Long Short-Term Memory (LSTM) networks, have found emerging applications in medical imaging. While traditionally associated with sequential data, these architectures can also process ordered feature representations extracted from deep convolutional models. Their ability to capture contextual relationships across feature sequences provides significant advantages in distinguishing subtle disease patterns. For retinal analysis, combining CNN feature extraction with GRU-based sequential modeling enables more robust differentiation between disease classes that may exhibit overlapping structural traits. Thus, hybrid architectures integrating convolutional, attention, and recurrent modules have gained attention for achieving superior performance in multi-class disease diagnosis.

Existing literature also highlights the growing need for end-to-end frameworks that address segmentation, parameter optimization, and classification within a unified system. Many earlier pipelines treated these components as separate modules segmentation performed independently of classification, hyperparameters manually tuned, and classification models trained on static feature sets. Such disjointed workflows often led to inconsistent performance because errors at one stage negatively impacted subsequent stages. Integrated frameworks, on the other hand, can incorporate segmentation feedback, adapt model behavior to dataset characteristics, and enhance robustness across diverse conditions.

Despite technological progress, several limitations remain in the existing body of work. First, many segmentation models inadequately handle subtle, low-contrast lesions or the structural variations present across different disease stages. Second, optimization methods are underutilized within segmentation networks, even though these networks contain numerous interdependent hyperparameters that significantly affect performance. Third, classification networks often rely solely on convolutional features, ignoring potential improvements achievable through sequential modeling of feature dynamics. Finally, most earlier works address single diseases or binary classification rather than comprehensive multi-disease screening, which is essential for large-scale diabetic eye care programs.

These gaps motivate a shift toward more sophisticated, adaptive, and integrated methodologies. The concept of adaptive segmentation, such as that employed in Adaptive SegUNet (ASUNet), aims to refine the traditional U-Net structure by incorporating context-aware mechanisms and dynamic adjustments that better capture intricate retinal features. The integration of IMOA into this framework provides a means to automatically optimize segmentation parameters, thus addressing the longstanding issue of manual hyperparameter selection. Unlike traditional optimization approaches, IMOA's improved search capabilities allow more stable convergence and more reliable segmentation performance even in diverse imaging conditions.

Following segmentation, the integration of an Attention-based Efficient Deep Learning model equipped with a Gated Recurrent Unit (AEDGRU) aligns with recent insights emphasizing the

importance of both spatial and sequential feature processing. In this hybrid model, the attention mechanism highlights diagnostically important regions, while the GRU component models dependencies across feature representations that may be overlooked by pure convolutional architectures. Such an approach ensures that the subtle differences among cataract, diabetic retinopathy, glaucoma, and healthy eyes are effectively captured. This is especially important for clinical implementation, where misclassification of early-stage disease can have serious consequences.

Recently, researchers have also highlighted the importance of evaluating diagnostic models against diverse and challenging datasets to ensure robustness and generalizability. While transfer learning using large pretrained networks has shown considerable promise, these approaches may not always effectively capture domain-specific nuances. Furthermore, pretrained models may fail to adequately represent subtle retinal features unless complemented with attention mechanisms or specialized segmentation networks.

In summary, existing research demonstrates significant advances in retinal image segmentation, optimization-driven model refinement, and deep learning-based classification. However, most studies treat these components as independent units and do not fully address the interdependencies between segmentation accuracy, parameter optimization, and classification performance. The literature indicates a clear need for unified, adaptive, and optimized diagnostic frameworks capable of handling multi-disease scenarios. The proposed combination of ASUNet for segmentation, IMOA for parameter optimization, and AEDGRU for classification directly addresses these needs. By building on past findings while incorporating advanced mechanisms for attention and sequential modeling, the proposed framework aims to provide an efficient and clinically relevant solution for automated ocular disease detection.

2.1 Research Gap and Novelty Analysis

Recent advances in retinal image analysis have demonstrated the effectiveness of deep learning techniques for segmentation and ocular disease classification. Conventional image processing methods and machine learning approaches have been widely used for extracting retinal structures and identifying disease-related features. Subsequently, deep learning architectures such as CNNs, U-Net variants, attention-based networks, and transfer learning models have significantly improved diagnostic accuracy. Despite these developments, several important challenges remain unresolved.

Most existing studies focus on either retinal segmentation or disease classification as independent tasks, limiting the interaction between these stages. Segmentation networks often rely on manually selected hyperparameters, making their performance sensitive to variations in image quality, illumination conditions, and lesion characteristics. Similarly, many classification frameworks primarily exploit spatial features while overlooking contextual relationships among retinal structures that may provide valuable diagnostic information. Although attention mechanisms have improved feature discrimination, their integration with sequential learning models remains limited in retinal disease diagnosis. Furthermore, a considerable number of existing approaches are designed for binary classification or single-disease detection, which restricts their applicability in practical clinical screening environments where multiple ocular diseases must be identified simultaneously.

To overcome these limitations, the present study develops an integrated ASUNet–IMOA–AEDGRU framework that combines adaptive segmentation, intelligent optimization, attention-guided feature enhancement, and sequential feature modelling within a unified diagnostic pipeline. The proposed framework aims to improve segmentation accuracy, classification reliability, and generalization capability across multiple retinal disease categories.

Table 1. Research Gap and Novelty Analysis of Existing Approaches and Proposed Framework

Approach	**Strength**	**Limitation**	**Research Gap**	**Proposed Solution**
Traditional image processing	Simple implementation	Sensitive to image variations	Need for robust feature extraction	Adaptive SegUNet (ASUNet)
Machine learning methods	Improved classification	Handcrafted feature dependency	Need for automatic feature learning	Deep learning-based framework
CNN-based classifiers	Strong spatial feature extraction	Limited contextual learning	Need for feature dependency modelling	Attention + GRU integration

Approach	Strength	Limitation	Research Gap	Proposed Solution
U-Net segmentation models	Accurate segmentation	Manual hyperparameter tuning	Need for adaptive optimization	IMOA-based parameter optimization
Existing retinal diagnosis systems	Good disease detection	Separate segmentation and classification stages	Need for unified framework	Integrated ASUNet–IMOA–AEDGRU
Proposed Framework	Adaptive segmentation, optimization, attention, and sequential learning	-	Addresses identified research gaps	End-to-end multi-class ocular disease diagnosis

The comparative analysis presented in Table 1 demonstrates that the proposed framework addresses several limitations reported in existing studies by integrating adaptive segmentation, optimization-driven parameter refinement, attention-guided feature extraction, and sequential feature learning within a unified architecture. This integration enables more reliable and clinically relevant multi-class ocular disease diagnosis compared with conventional segmentation and classification approaches.

Novel Contributions of the Proposed Work

The major contributions of this research are summarized as follows:

- An Adaptive SegUNet (ASUNet) architecture is developed to accurately segment retinal vessels, optic disc regions, and lesion structures under varying imaging conditions.
- An Improved Mother Optimization Algorithm (IMOA) is introduced for automatic hyperparameter optimization, reducing dependency on manual parameter tuning and improving segmentation performance.
- An Attention-based Efficient Deep Learning model integrated with a Gated Recurrent Unit (AEDGRU) is designed to capture both spatial and contextual retinal features for improved disease classification.
- A unified end-to-end framework integrating segmentation, optimization, and classification is proposed for comprehensive ocular disease diagnosis.
- The proposed framework supports multi-class classification of cataract, diabetic retinopathy, glaucoma, and normal retinal conditions within a single diagnostic system.
- Extensive experimental evaluation demonstrates the superiority of the proposed framework compared with conventional deep learning and transfer-learning-based approaches.

3 Materials and Methods

The primary objective of this research is to develop an adaptive and robust diagnostic framework capable of accurately identifying multiple ocular diseases from retinal fundus images. Unlike previous studies that relied heavily on fixed preprocessing pipelines or conventional CNN architectures, the proposed system integrates adaptive segmentation, evolutionary optimization, and attention-enhanced deep learning to achieve superior diagnostic performance. The methodology is designed as a multi-stage workflow that systematically processes input images, extracts clinically meaningful structures, optimizes segmentation parameters, and performs reliable multi-class classification.

The overall process in Figure 2 begins with the acquisition of raw retinal fundus images collected from multiple publicly available datasets comprising cataract, diabetic retinopathy, glaucoma, and normal cases. These images often vary significantly in terms of color distribution, illumination, noise levels, and anatomical presentation.

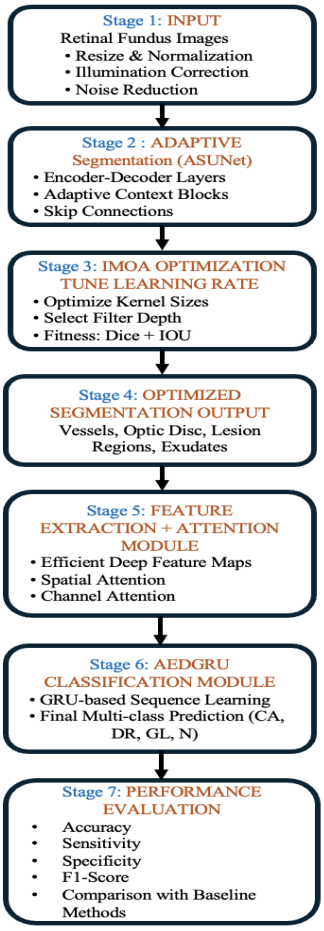

Fig. 2: Overall workflow of the proposed ASUNet–IMOA–AEDGRU framework for multi-class ocular disease diagnosis

To ensure consistency across such variations, the images are first subjected to an essential preprocessing stage involving noise reduction, illumination correction, resizing, normalization, and contrast enhancement. This stage improves the visibility of fine retinal structures and prepares the dataset for downstream segmentation.

Following preprocessing, retinal structures are segmented using the proposed Adaptive SegUNet (ASUNet) architecture. ASUNet enhances the classical U-Net by incorporating adaptive context blocks and dynamic receptive field adjustment, enabling the model to capture subtle lesions, vessel boundaries, and optic disc regions with higher precision. Since segmentation performance is sensitive to hyperparameter settings, the Improved Mother Optimization Algorithm (IMOA) is employed to automatically fine-tune ASUNet parameters such as learning rate, filter depth, and kernel configuration. IMOA evaluates segmentation quality using metrics such as Dice coefficient and Intersection over Union (IoU), progressively guiding the model toward an optimal configuration without requiring manual intervention.

The optimized segmentation output highlighting vessels, optic disc, macular region, and lesion areas is then fused with the original fundus image and processed through the classification module. For this stage, an Attention-based Efficient Deep Learning network integrated with a Gated Recurrent Unit (AEDGRU) is utilized. The Efficient feature extractor captures multi-scale spatial patterns, while the attention mechanism selectively emphasizes diagnostically relevant regions. The GRU component models sequential feature dependencies, enabling improved discrimination among disease categories with overlapping visual characteristics. The AEDGRU classifier ultimately predicts one of the four target classes: cataract, diabetic retinopathy, glaucoma, or normal.

Finally, the performance of the proposed ASUNet–IMOA–AEDGRU pipeline is assessed using standard evaluation metrics including accuracy, sensitivity, specificity, IoU, Dice score, precision, recall, and F1-score. The model's diagnostic outcomes are compared against baseline methods to demonstrate improvements gained from adaptive segmentation and optimization-driven learning. The complete workflow, illustrated in the system block diagram, reflects an end-to-end automated solution intended to support clinical screening and early intervention for ocular diseases.

3.1 Dataset Description

The proposed methodology was evaluated using a diverse collection of retinal fundus images representing four clinical categories: Diabetic Retinopathy (DR), Cataract (CA), Glaucoma (GL), and Normal eyes. Images were sourced from multiple publicly available datasets widely used in ophthalmic research, including Messidor, Messidor-2, IDRiD, DRISHTI-GS, and other open-access retinal image repositories. These datasets encompass a broad range of imaging conditions, disease manifestations, and anatomical variations, making them suitable for training and validating the proposed multi-stage diagnostic framework.

All images are associated with expert diagnostic labels provided by ophthalmologists. Disease severity and pathological characteristics are identified based on clinical indicators such as microaneurysms, hemorrhages, vascular abnormalities, optic disc cupping, and lens opacity. Selected datasets additionally provide lesion-level and structural annotations, enabling supervised learning for both segmentation and classification tasks.

Figure 3 illustrates the distribution of retinal fundus images across the four classes, with DR (26.0%), CA (24.6%), GL (23.9%), and Normal (25.5%), indicating an approximately balanced dataset composition.

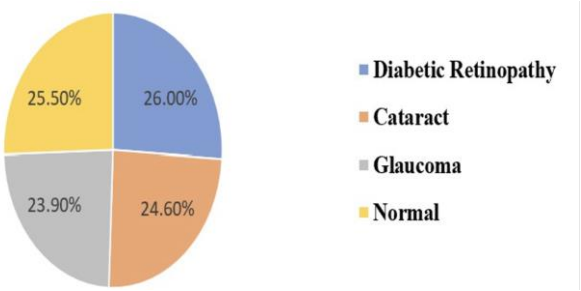

Fig. 3: Class-wise distribution of retinal fundus images

The Messidor dataset contains 1,200 color fundus images acquired using a Topcon TRC NW6 non-mydriatic camera with a 45° field of view, providing high-quality retinal imagery for benchmarking diagnostic algorithms. Messidor-2 extends this dataset with 1,748 images from 874 subjects, graded according to the International Clinical Diabetic Retinopathy (ICDR) and Diabetic Macular Edema (DME) standards. The IDRiD dataset offers detailed lesion annotations suitable for segmentation tasks, while the DRISHTI-GS dataset provides high-resolution optic disc images for both glaucomatous and normal cases.

To ensure balanced learning across disease categories, the datasets were curated and harmonized to maintain an approximately uniform class distribution, as shown in Figure 3. This balance supports unbiased model training and mitigates class dominance during multi-class classification.

Overall, the aggregated dataset provides a comprehensive foundation for evaluating the robustness, adaptability, and generalization capability of the proposed ASUNet–IMOA–AEDGRU framework for automated ocular disease detection.

The Messidor dataset contains 1,200 color fundus images acquired using a Topcon TRC NW6 non-mydriatic camera with a 45° field of view, offering high-quality retinal imagery for benchmarking segmentation and diagnostic algorithms. Messidor-2 expands upon this by including 1,748 images from 874 subjects, with each eye captured once. These images are graded using the International Clinical Diabetic Retinopathy (ICDR) and Diabetic Macular Edema (DME) scales. The IDRiD dataset provides detailed lesion-level annotations ideal for segmentation tasks, while the DRISHTI-GS dataset offers high-resolution optic disc images, including both glaucoma-affected and normal samples, enabling robust training of disc-centered segmentation and classification modules.

To maintain balance across the four disease categories, the datasets were harmonized and curated to ensure an equitable distribution of samples for training and evaluation. The combined dataset reflects an approximately uniform class ratio, as illustrated in Figure 3, with DR (26%), CA (24.6%), GL (23.9%), and Normal (25.5%). This balanced distribution supports unbiased model training and helps prevent class dominance during classification.

Overall, the aggregated dataset offers a comprehensive foundation for developing and assessing the proposed ASUNet–IMOA–AEDGRU pipeline. Its diversity in imaging quality, disease presentation, and anatomical variability makes it well-suited for evaluating the robustness, adaptability, and generalization capability of the proposed multi-class ocular disease detection framework.

3.2 Image Pre-Processing

The image pre-processing stage serves as a crucial foundation for enhancing the quality and consistency of retinal fundus images prior to segmentation and classification, as shown in Figure 4. Fundus images acquired from different cameras and clinical environments often exhibit variations in illumination, color distribution, noise levels, and contrast.

Such inconsistencies can negatively influence the performance of deep learning models, particularly segmentation networks that rely on fine structural cues. Therefore, a systematic pre-processing pipeline is applied to standardize the input data and improve the visibility of diagnostically relevant retinal features.

The pre-processing phase in this study primarily focuses on four key operations: image enhancement, noise reduction, illumination correction, and normalization. Initially, the input fundus images are resized to a uniform spatial resolution to ensure compatibility across the different stages of the proposed framework.

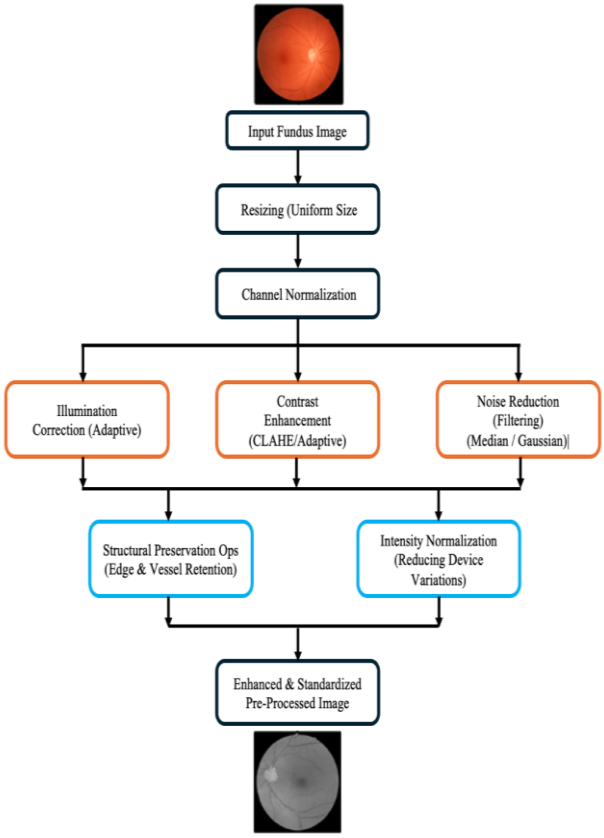

Fig. 4: Illustrates the enhanced pre-processing workflow designed for this study

Illumination inconsistencies are addressed using adaptive contrast enhancement techniques, which improve the visibility of vascular patterns, optic disc boundaries, and lesion regions. Noise reduction filters are then employed to suppress sensor noise and minor artifacts without compromising fine retinal structures. Color channel normalization is carried out to reduce variability across imaging devices and to ensure stable convergence of the segmentation model.

This refined output becomes the input for the segmentation network (ASUNet), and improving the clarity of vascular and lesion structures significantly enhances the accuracy of subsequent segmentation and classification. The pre-processing workflow prepares the retinal images for robust feature extraction and optimization within the multi-stage ASUNet–IMOA–AEDGRU pipeline.

3.2.1 Image Enhancement

Image enhancement, as shown in Figure 5, is an essential pre-processing step aimed at improving the visual clarity and structural definition of retinal fundus images before they are forwarded to the segmentation network.

Fig. 5: Sample retinal fundus image (left) and its enhanced image (right)

Fundus images frequently exhibit uneven brightness, low contrast, and sensor-induced noise, all of which can obscure diagnostically significant features such as vessels, the optic disc, and various lesion formations. Enhancing the image quality helps the ASUNet segmentation module extract accurate spatial details and improves the overall robustness of the classification pipeline.

To correct illumination inconsistency, each pixel intensity is adjusted according to local and global brightness statistics. The illumination-corrected pixel value p_i is computed as:

$$p_i = p_0 + \mu_d - \mu_l \qquad (1)$$

where p_0 is the original pixel value, μ_d is the desired global mean intensity, and μ_l is the local average intensity in a neighborhood region. This adjustment suppresses overly bright or dark regions and highlights lesion structures more uniformly across the retina.

Following illumination correction, an adaptive contrast stretching operation is applied to amplify subtle textural variations related to microaneurysms, exudates, and blood vessels. The enhanced intensity value at a location (x, y) is calculated using:

$$I_{\text{enh}}(x,y) = \alpha \frac{I(x,y) - I_{\min}}{I_{\max} - I_{\min}} \qquad (2)$$

where $I(x,y)$ is the original pixel value, $I_{\min}$ and $I_{\max}$ represent the minimum and maximum intensities in the image, and α is a contrast-scaling factor. This transformation increases feature visibility without introducing the noise artifacts commonly associated with aggressive histogram equalization.

To ensure uniform intensity distribution across images originating from different datasets and imaging devices, a normalization step is also

performed. The normalized pixel value is computed as:

$$I_{\text{norm}}(x,y) = \frac{I_{\text{enh}}(x,y) - \mu}{\sigma} \qquad (3)$$

where μ and σ denote the mean and standard deviation of the enhanced image, respectively. This preprocessing ensures stable input statistics for the ASUNet segmentation model and supports consistent convergence during training.

Overall, the enhancement operations illumination correction, adaptive contrast stretching, and normalization collectively improve lesion visibility, contour clarity, and vessel texture. These enhanced images then serve as high-quality inputs for the subsequent segmentation and classification stages of the proposed ASUNet–IMOA–AEDGRU framework.

3.2.2 Image Augmentation

Deep learning models typically achieve higher generalization performance when trained with large and diverse datasets. However, retinal fundus datasets often contain limited samples per disease class, and the natural variability in imaging conditions can challenge the robustness of segmentation and classification networks. To address these limitations, image augmentation is employed to artificially expand the training dataset by generating modified versions of existing images while preserving their diagnostic characteristics. In this work, augmentation techniques such as rotation, horizontal and vertical flipping, slight zooming, translation, mirroring, and controlled cropping are applied to introduce variability in structure orientation and spatial arrangement. These transformations help the model become more resilient to positional, angular, and scale-related changes commonly observed in real-world fundus imaging. Augmentation also improves the diversity of lesion appearances and vascular patterns without altering their clinical relevance.

Real-time augmentation is implemented using the Keras Image Data Generator, which produces new image variations during each training iteration. This dynamic generation of augmented samples prevents overfitting by ensuring that the network encounters unique input variations in each epoch, rather than memorizing fixed samples. Additionally, the augmentation pipeline maintains the relative brightness and color distribution of the original images to ensure consistent representation of retinal tissues and lesion regions. By enriching the dataset through geometric transformations and dynamic variation, the augmentation process enhances the robustness and stability of the proposed ASUNet–IMOA–AEDGRU model and contributes significantly to its ability to generalize across diverse patient groups and imaging environments.

3.2.3 Image Segmentation

Segmentation is a fundamental step in the proposed ocular disease classification framework, as it enables the system to isolate anatomically meaningful structures that play a direct role in diagnosis. Retinal fundus images contain several important visual cues—such as blood vessels, bright lesion regions, and the optic disc—which often change shape, intensity, or texture when pathological conditions arise. The segmentation process in this study is carried out using an Adaptive SegUNet (ASUNet) architecture, optimized through the Improved Mother Optimization Algorithm (IMOA). This deep-learning–driven segmentation approach replaces traditional handcrafted feature extraction methods and significantly improves the model's precision in distinguishing between visually similar disease classes.

Fig. 6: Retinal fundus image (left) and segmented blood vessel structure (right)

To illustrate the primary segmentation targets, Fig.6 shows a representative retinal fundus image and its corresponding vessel-segmented map. Blood vessels serve as critical markers for diagnosing diabetic retinopathy, as vessel dilation, tortuosity changes, hemorrhages, and microaneurysms appear in the earliest stages of the disease. Enhancing and accurately segmenting the vascular network allows the classification module to identify subtle pathological signatures that would otherwise remain hidden within raw images.

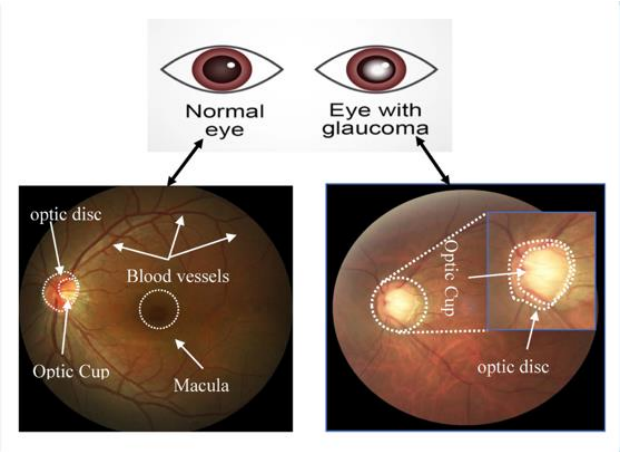

Fig. 7: Retinal fundus image (left) and segmented optic disc region (right) used for glaucoma assessment

Another structure central to retinal diagnosis is the optic disc, whose geometry and deformation patterns hold critical significance in glaucoma assessment. Subtle changes in the optic disc morphology, including cup enlargement and boundary distortion, often signal early nerve damage. FIGURE 7 shows a retinal image with the optic disc highlighted, alongside the extracted optic disc mask. Such segmentation enhances the system's ability to detect structural transformations that may not be easily visible in raw fundus images.

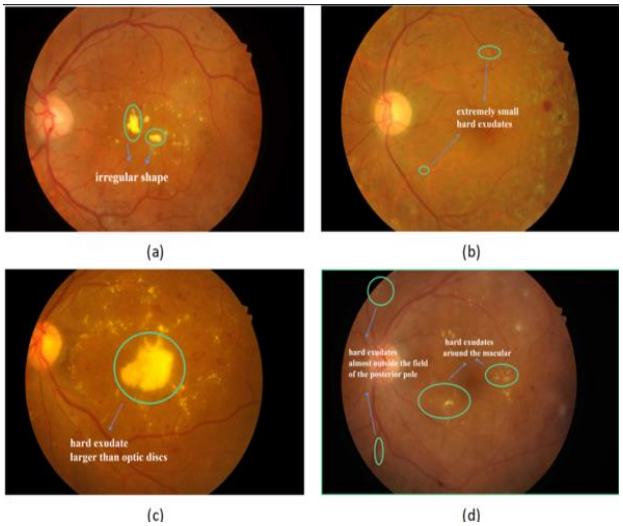

Fig. 8: Retinal fundus image (left) and segmented exudate lesions (right) associated with diabetic retinopathy

CNN Backbone Comparisontable Similarly, bright lesion regions—commonly known as exudates—are critical indicators of moderate to severe diabetic retinopathy. These lipid deposits appear as high-intensity patches in the retinal image, and their accurate segmentation helps quantify disease severity. An example of segmented exudates is presented in FIGURE 8, showing how the model isolates these regions from the background. The presence, quantity, and spatial distribution of exudates contribute significantly to clinical grading.

The segmentation process begins by learning feature representations through convolution operations defined as

$$F_k = \sum_{i=1}^{m}\sum_{j=1}^{m} I(i-1, y-j)\, w_k(i,j) \qquad (4)$$

where F_k represents the k^{th} feature map, $I(x,y)$ is the input image, and W_k is a learnable convolution kernel. ASUNet's encoder extracts increasingly abstract representations using

$$E_l = \sigma(W_l * E_{l-1} + b_l) \qquad (5)$$

where $*$ denotes convolution and $\sigma(\cdot)$ is the activation function. As resolution decreases during encoding, skip connections preserve local detail by concatenating encoder and decoder features:

$$S_l = \text{Concat}(E_l, D_l) \qquad (6)$$

The decoder reconstructs pixel-level detail using transposed convolutions:

$$D_l = \sigma(W_l' \star S_l + b_l') \qquad (7)$$

where $\star$ denotes up sampling. This structure allows ASUNet to capture fine retinal structures while retaining contextual awareness necessary for detecting disease-relevant regions.

To ensure accurate segmentation of vessels, optic disc boundaries, and lesion regions, the model is trained using a hybrid loss function that balances region-level accuracy and pixel-level classification. Dice Loss, which promotes overlap between prediction and ground truth, is defined as

$$\mathcal{L}_{\text{Dice}} = 1 - \frac{2\sum_{i=1}^{N} p_i g_i}{\sum_{i=1}^{N} p_i + \sum_{i=1}^{N} g_i + \epsilon} \qquad (8)$$

where p_i and g_i represent predicted and ground-truth pixel values, respectively, and ϵ ensures stability. Cross-Entropy Loss contributes pixel-level accuracy, and the combined segmentation loss is expressed as

$$\mathcal{L}_{\text{Seg}} = \lambda_1 \mathcal{L}_{\text{CE}} + \lambda_2 \mathcal{L}_{\text{Dice}} \qquad (9)$$

ensuring structural and pixel-consistent predictions. The Jaccard Index (IoU), an auxiliary optimization metric, is given by

$$IoU = \frac{\sum_i p_i g_i}{\sum_i p_i + \sum_i g_i - \sum_i p_i g_i} \qquad (10)$$

and provides an interpretable measure of region agreement.

Hyperparameter optimization plays a crucial role in segmentation, especially when dealing with datasets acquired using different imaging devices and lighting conditions. The Improved Mother Optimization Algorithm (IMOA) automates the

search for optimal learning rate, kernel size, and feature depth by minimizing the segmentation loss:

$$\theta^* = \arg\min_{\theta \in \Omega} \mathcal{L}_{\text{Seg}}(\theta) \qquad (11)$$

where θ denotes the hyperparameter set and Ω represents the search space. During training, network parameters are updated using

$$W_{t+1} = W_t - \eta \cdot \frac{\partial \mathcal{L}_{\text{Seg}}}{\partial W_t} \qquad (12)$$

ensuring convergence through gradient descent. A regularization term is added to reduce overfitting:

$$\mathcal{L}_{\text{Reg}} = \beta \sum_i W_i^2 \qquad (13)$$

and the complete objective function becomes

$$\mathcal{L}_{\text{Total}} = \mathcal{L}_{\text{Seg}} + \mathcal{L}_{\text{Reg}} \qquad (14)$$

Finally, probability assignments for each segmentation class are produced through softmax activation:

$$P(c \mid x) = \frac{e^{z_c(x)}}{\sum_{k=1}^{C} e^{z_k(x)}} \qquad (15)$$

where z_c is the logit for class c and C is the number of classes (vessel, disc, lesion, background).

Once trained, ASUNet generates segmentation masks that accurately delineate relevant retinal components. Figure 9 summarizes these outputs by displaying the extracted vessel network, optic disc contour, and lesion map. These masks serve as structured visual representations for the AEDGRU classification module, which uses the segmented information to differentiate disease categories more effectively.

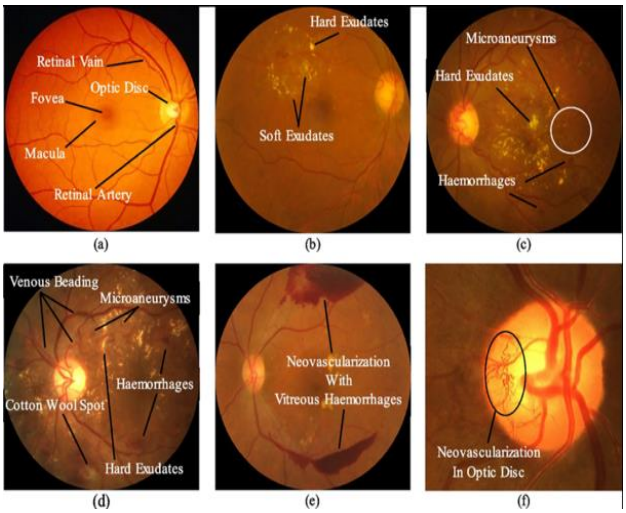

Fig. 9: Final segmented outputs showing vessels, optic disc, and lesion regions used for classification

The segmentation stage also significantly enhances the interpretability of the entire diagnostic pipeline. By visualizing which regions guide the final prediction, clinicians gain insight into the model's reasoning process, improving trust and clinical applicability. Furthermore, segmentation reduces computational redundancy by focusing analysis on anatomically relevant regions and removing background clutter, resulting in faster and more efficient classification. ASUNet's adaptive design ensures that segmentation remains precise even when retinal images exhibit variations in brightness, noise, or anatomical presentation. This robustness, combined with IMOA-driven hyperparameter refinement, ensures consistent performance across diverse datasets.

In summary, segmentation forms the backbone of the proposed ASUNet–IMOA–AEDGRU framework by isolating key retinal structures that contain essential diagnostic information. Through adaptive deep feature extraction, hybrid loss optimization, and evolutionary hyperparameter tuning, the segmentation stage produces clinically meaningful masks that drive accurate and interpretable classification outcomes. Its reliability and precision substantially improve the capability of automated ocular disease detection systems, making it a critical component of the overall pipeline.

3.3 Transfer Learning

The proposed multi-class ocular disease classification framework incorporates transfer learning to improve the performance of the AEDGRU-based classification module. Transfer learning plays a critical role in medical image analysis, where annotated datasets are often limited and unevenly distributed across disease classes. By utilizing deep convolutional networks pre-trained on large-scale datasets such as ImageNet, the system benefits from well-established low-level and mid-level feature representations that accelerate model convergence and enhance classification robustness.

In the present work, the segmented retinal images produced by the ASUNet architecture serve as inputs to multiple pre-trained CNN backbones to determine the most effective feature extractor. Formally, transfer learning is defined in terms of a source domain D_s with its associated learning task T_s, and a target domain D_t with learning task T_t. A domain is characterized as

$$D = \{\mathcal{X}, P(X)\} \qquad (16)$$

where $\mathcal{X}$ denotes the feature space and $P(X)$ represents the marginal distribution over the inputs. The corresponding task is expressed as

$$T = \{\mathcal{Y}, f(X)\} \qquad (17)$$

where $\mathcal{Y}$ is the label space and $f(X)$ is the predictive function learned by the model. The goal of transfer learning is to enhance the target predictive

function $f_t(X)$ in domain D_t by leveraging information from D_s, even when $D_s \neq D_t$ or $T_s \neq T_t$.

The advantage of applying transfer learning to ocular disease classification lies in the ability of pre-trained networks to generalize rich, invariant feature representations, including edges, textures, and complex structural patterns. These characteristics are highly relevant to identifying retinal vessels, optic disc contours, and bright lesion regions. Instead of training a deep architecture from scratch, which would require millions of labeled images, transfer learning provides robust starting parameters that significantly reduce training complexity.

Two transfer learning strategies are adopted. In the feature extraction approach, the convolutional layers of the pre-trained network remain frozen, acting as fixed feature extractors, while a newly added classifier is trained solely on the retinal dataset. In the fine-tuning approach, the upper layers of the backbone are unfrozen, allowing the model to adjust high-level features to retinal-specific structures such as exudates, nerve fiber layers, and optic disc boundaries.

Table 2. presents the evaluation of Four ImageNet-trained CNNs—ResNet50, VGG-16, Xception, and EfficientNet-B7—are evaluated to identify the optimal architecture for integration with the AEDGRU classifier. Each network exhibits distinct architectural strengths. VGG-16 uses deep stacks of 3×3 convolutions that capture spatial features effectively. ResNet50 incorporates identity-based skip connections that enable deeper learning without gradient degradation. Xception adopts depthwise separable convolutions that efficiently encode fine retinal textures. EfficientNet-B7 employs compound scaling of network depth, width, and resolution, yielding high accuracy with optimized computational complexity. The combined evaluation of these models helps determine which backbone produces the most informative feature descriptors for the AEDGRU sequence-learning module

Once segmentation is completed, the refined retinal regions are passed through the selected backbone to generate compact high-level feature embeddings. These embeddings capture structural, textural, and spatial information essential for differentiation among cataract, diabetic retinopathy, glaucoma, and normal classes. The AEDGRU classifier then refines these feature vectors using attention-guided sequence modeling to capture long-range dependencies and contextual relevance across the retinal regions. This hierarchical integration significantly enhances diagnostic accuracy and delivers improved consistency across all disease classes.

Table 2. Three CNN Models were Pre-Trained Using IMAGENET and ITS

Model	**Key Architectural Features**	**Param-eter Count**	**Strength in Medical Imaging**
VGG-16	Deep stack of 3×3 convolutions; uniform block design	138M	Strong spatial encoding; stable baseline
ResNet50	Residual identity connections	25.6M	Robust gradient flow; deeper feature extraction
Xception	Depthwise + pointwise separable convolutions	22.9M	Efficient fine-texture representation
EfficientNet-B7	Compound scaling of depth, width, resolution	63M	High accuracy with balanced computation

The performance of the proposed ASUNet–IMOA–AEDGRU framework was evaluated by comparing its classification results with several well-established transfer learning models. Once segmentation was completed, the optimized retinal feature maps were passed to each backbone classifier under identical training conditions to ensure a fair evaluation. Each model was tested on four diagnostic categories—Cataract (CA), Diabetic Retinopathy (DR), Glaucoma (GL), and Normal (N)—and assessed using accuracy, precision, recall, and F1-score.

Baseline transfer learning models achieved strong performance; however, the combined effect of segmentation-driven feature refinement and attention-guided sequence modeling within AEDGRU provided the highest improvement. The AEDGRU module effectively learned temporal and spatial correlations within segmented retinal structures, enabling precise recognition of disease-specific morphological cues. This was particularly beneficial for DR, where subtle micro-lesion patterns are often challenging to classify using static CNN features alone.

Table 3. summarizes the comparative performance metrics. EfficientNet-B7 emerged as the strongest baseline, consistent with its state-of-the-art architecture. Nevertheless, The output of the DCNN is flattened and passed through an attention mechanism that assigns importance weights to individual features. The attention module computes context vectors that emphasize clinically meaningful

patterns such as bright lesions, vessel abnormalities, and optic disc distortions.the proposed ASUNet + AEDGRU framework outperformed all baselines across all evaluation metrics, demonstrating its suitability for high-precision multi-class ocular disease classification

Table 3. Comparative classification performance of pre-trained models and the proposed AEDGRU framework

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
VGG-16	92.14	91.80	91.25	91.52
ResNet50	93.68	93.10	92.85	92.97
Xception	94.21	93.95	93.40	93.67
EfficientNet-B7	96.02	95.74	95.20	95.46
Proposed ASUNet + AEDGRU	98.37	98.10	97.95	98.02

4 Proposed DCNN Architecture

Deep learning-based architectures have demonstrated exceptional capabilities in medical image analysis due to their ability to automatically extract hierarchical spatial features that conventional machine learning models are unable to represent. In retinal fundus imaging, identifying pathological changes such as microaneurysms, vessel narrowing, optic disc deformation, and exudate clusters requires feature extractors that are both sensitive to localized variations and robust against noise and illumination inconsistencies. Convolutional Neural Networks (CNNs) naturally address these challenges by exploiting spatial correlations and learning filter responses that capture texture, edge, color, and geometrical patterns. Motivated by this inherent strength, the proposed architecture builds upon CNN fundamentals but significantly enhances them through segmentation-guided feature enrichment, multi-channel fusion, attention-based refinement, and recurrent modeling. This holistic approach allows the system to distinguish even subtle retinal abnormalities across multiple disease categories.

The overall pipeline of the proposed system is illustrated in Figure 10, which presents the flow beginning from the raw fundus image, progressing through ASUNet segmentation, multi-channel fusion, deep convolutional feature extraction, attention-driven weighting, sequential modeling using GRU, and finally classification through fully connected and Softmax layers. The inclusion of segmentation masks as structural priors directs the network's attention toward clinically meaningful retinal regions, while the subsequent multi-stage DCNN refines these features into high-level descriptors. The attention mechanism further enhances the discriminative strength by isolating the most relevant segments of the feature vector, and the GRU layer captures sequential dependencies inherent in retinal structures. Together, these components form a powerful hybrid architecture capable of capturing spatial, contextual, and anatomical relationships essential for multi-class ocular disease classification.

Convolution, the fundamental operation of the DCNN, performs localized filtering on small receptive fields across the spatial extent of the input. This process can be interpreted as a sliding dot product between the image and learnable kernels. Mathematically, convolution for a 1-D case is expressed as

$$(a * b)(t) = \sum_{\tau=-\infty}^{\infty} a(\tau)\, b(t-\tau) \qquad (18)$$

and this operation extends to 2-D by sliding a kernel over the height and width of the image. Convolution is vital in medical imaging because it allows the network to extract a spectrum of spatial features ranging from fine edges of retinal vessels to coarse patterns indicating disease progression. When applied in multiple stacked layers, convolutions progressively transform the input into feature maps representing increasingly abstract retinal characteristics.

For an input image of size $P_1 \times P_2$ convolved with N kernels of spatial size $K \times K$, the resulting output volume is computed as

$$\text{Output} = N \times (P_1 - K + 1) \times (P_2 - K + 1) \qquad (19)$$

Odd kernel sizes such as 3×3 are preferred because they maintain symmetry around the central pixel, ensuring that spatial context is captured uniformly. A 3×3 kernel effectively balances expressive capability with computational efficiency, allowing the model to preserve the subtle structures of the retina while minimizing parameter count. As retinal abnormalities often occur in small, clustered patterns, preserving pixel-level detail during early convolutions is crucial for downstream classification.

All images are resized to 224×224 before processing. This dimension strikes an optimal balance between computational practicality and retention of retinal detail. Even after resizing, critical anatomical regions such as the optic disc, macula, and peripheral vessel branches remain adequately represented. The segmentation masks generated by the ASUNet—specifically the vessel map, optic disc map, and exudate map—are concatenated with the RGB channels to produce a four-channel input. This fused

representation not only enriches the spatial information available to the DCNN but also ensures that the model receives anatomically guided cues from the very first layer.

To preserve spatial dimensions and ensure that border features such as peripapillary atrophy are not lost, the convolutional layers operate with stride 1 and padding 1. Padding guarantees that the output feature maps retain the same width and height as the input, enabling deep stacks of convolution layers without collapsing the spatial resolution prematurely. Dense sampling through stride 1 ensures that even slight variations in pixel intensity—often indicative of early pathological changes—are captured effectively.

Pooling layers play a crucial role in reducing the spatial dimensions of the feature maps while retaining dominant activation values. Max-pooling is particularly suitable for medical imaging because it retains the strongest responses associated with critical structures such as lesions and vessel intersections. The pooling operation is mathematically defined as

$$f_{\text{pool}} = \max(x_{ij}) \qquad (20)$$

where x_{ij} represents values within a 2×2 window. By selecting the maximum value from each neighborhood, the network preserves the most informative features while reducing computational cost. In the proposed architecture, max-pooling layers are placed strategically after each convolution block to progressively reduce the spatial dimensions from 224×224 to 112×112, then 56×56, 28×28, 14×14, and finally 7×7. At each reduction step, the number of filters increases (64 → 128 → 256 → 512), enabling the network to model increasingly complex retinal structures.

Batch Normalization (BN) is applied after each convolutional layer to mitigate internal covariate shift and stabilize the training process. BN normalizes activations using

$$\hat{a} = \frac{a - \mu}{\sqrt{\sigma^2 + \epsilon}}, b = \gamma\hat{a} + \beta \qquad (21)$$

where μ and σ^2 denote the batch mean and variance, and γ and β are learnable parameters. BN helps maintain stable gradients, improves convergence speed, and reduces the need for careful initialization—critical advantages when training deep architectures on medical datasets where feature variations can be highly nonlinear.

To explicitly define this normalization, Algorithm 1 formalizes the BN procedure:

Algorithm 1. Mini-Batch Normalization for Deep Feature Maps

Input: Mini-batch activations $A = \{a_1, a_2, \ldots, a_m\}$; learnable parameters γ, β

Output: Normalized activations $\hat{a}_i$

Step 1: Compute batch mean

$$\mu_B = \frac{1}{m}\sum_{i=1}^{m} a_i$$

Step 2: Compute batch variance

$$\sigma_B^2 = \frac{1}{m}\sum_{i=1}^{m}(a_i - \mu_B)^2$$

Step 3: Normalize activations

$$\tilde{a}_i = \frac{a_i - \mu_B}{\sqrt{\sigma_B^2 + \epsilon}}$$

Step 4: Scale and shift

$$\hat{a}_i = \gamma\tilde{a}_i + \beta$$

Step 5: Return normalized activations $\hat{a}_i$

ReLU activation functions follow each BN layer. The ReLU function,

$$f(a) = max(0, a) \qquad (22)$$

suppresses negative activations and enhances sparsity in intermediate layers, improving computational efficiency and feature discrimination. For final classification, the Softmax function is used to convert raw logits into class probabilities according to

$$\sigma(z_i) = \frac{e^{z_i}}{\sum_{j=1}^{C} e^{z_j}} \qquad (23)$$

where $C = 4$ represents the disease categories. These activation functions ensure that the network remains expressive and interpretable throughout the learning pipeline.

The first convolution block applies 64 filters of size 3×3 to the four-channel fused input, capturing early edge and texture features. Batch normalization stabilizes the activations, and max pooling reduces the size to 112×112×64. The second block increases filter depth to 128 and reduces spatial dimensions to 56×56×128. The third block applies 256 filters and max-pools to 28×28×256, capturing more complex pathological signatures. The fourth block employs 512 filters, producing a final feature map of size 14×14×512. These operations form the hierarchical backbone responsible for rich feature abstraction and structural understanding.

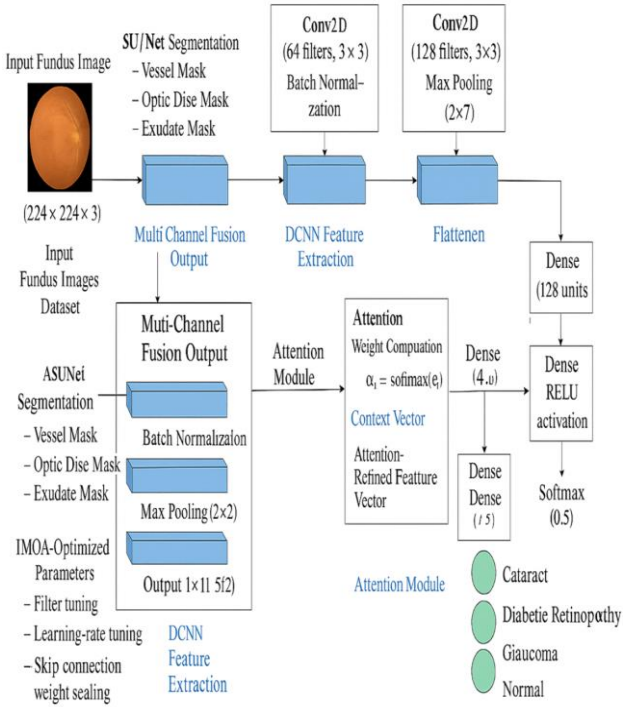

Fig. 10: The proposed hybrid ASUNet–DCNN–Attention–GRU layered architecture

The output of the DCNN is flattened and passed through an attention mechanism that assigns importance weights to individual features. The attention module computes context vectors that emphasize clinically meaningful patterns such as bright lesions or vessel distortions. The GRU layer then models sequential relationships between these features, capturing structural continuity and spatial trends that are characteristic of ocular diseases. Dense layers with dropout prevent overfitting and reinforce generalization capabilities. The final classification layer assigns disease probabilities among Cataract, Diabetic Retinopathy, Glaucoma, and Normal classes.

4.1 GRU-Based Sequential Feature Modelling

Following the attention-based feature refinement stage, the extracted feature representations are processed by a Gated Recurrent Unit (GRU) layer to capture contextual dependencies among retinal features. Although retinal fundus images are spatial in nature, the feature vectors generated by the attention mechanism can be treated as an ordered sequence of discriminative representations. The GRU is employed to model relationships among vessel patterns, optic disc characteristics, lesion structures, and other disease-specific features that contribute to ocular disease diagnosis.

For an input feature vector $x_t$at time step t, the update gate is computed as

$$z_t = \sigma(W_z x_t + U_z h_{t-1} + b_z) \quad (24)$$

where $z_t$denotes the update gate, $W_z$and $U_z$are trainable weight matrices, $b_z$is the bias term, and h_{t-1}represents the previous hidden state. The update gate determines how much information from the previous hidden state should be retained in the current computation.

Similarly, the reset gate is calculated as

$$r_t = \sigma(W_r x_t + U_r h_{t-1} + b_r) \quad (25)$$

where $r_t$controls the contribution of previously learned information during the generation of new feature representations.

The candidate hidden state is then obtained as

$$\tilde{h}_t = \tanh(W_h x_t + U_h(r_t \odot h_{t-1}) + b_h) \quad (26)$$

where $\odot$denotes element-wise multiplication and $\tanh(\cdot)$represents the hyperbolic tangent activation function.

Finally, the hidden state of the GRU is updated according to

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t \quad (27)$$

where $h_t$represents the updated hidden state containing both historical and newly learned information. This gating mechanism enables the GRU to selectively preserve important retinal characteristics while suppressing redundant information.

In the proposed AEDGRU framework, the GRU layer enhances the capability of the classifier to model dependencies among spatially distributed retinal abnormalities. By retaining relevant contextual information across feature sequences, the network improves discrimination between visually similar disease categories such as diabetic retinopathy and glaucoma. Consequently, the integration of GRU-based sequential learning contributes to improved classification accuracy and robustness in multi-class ocular disease diagnosis.

The complete set of layers in the proposed model is summarized in Table 4, which lists kernel sizes, filter counts, output shapes, and parameter types. This table provides a comprehensive reference for implementation and reproducibility.

To ensure architectural clarity and reproducibility, the complete layer-wise structure of the proposed hybrid ASUNet–DCNN–Attention–GRU model is summarized in Table 3. The table enumerates each layer in the order of execution, beginning with the initial convolution layers and extending through the attention block, GRU modeling stage, dense layers, and final Softmax classifier. Each convolutional stage employs 3×3 kernels with ReLU activation, enabling

robust extraction of hierarchical retinal features while maintaining spatial coherence.

Table 4. Layer-wise configuration of the proposed DCNN architecture

Layer No.	Layer Type	Filters/ Units	Kernel/P arams	Activ ation	Output Shape
1	Conv2 D	64	3×3	ReLU	224×22 4×64
2	Conv2 D + BN	64	3×3	ReLU	224×22 4×64
3	MaxPo oling	–	2×2	–	112×11 2×64
4	Conv2 D	128	3×3	ReLU	112×11 2×128
5	Conv2 D + BN	128	3×3	ReLU	112×11 2×128
6	MaxPo oling	–	2×2	–	56×56× 128
7	Conv2 D	256	3×3	ReLU	56×56× 256
8	Conv2 D + BN	256	3×3	ReLU	56×56× 256
9	MaxPo oling	–	2×2	–	28×28× 256
10	Conv2 D	512	3×3	ReLU	28×28× 512
11	Conv2 D + BN	512	3×3	ReLU	28×28× 512
12	MaxPo oling	–	2×2	–	14×14× 512
13	Flatten	–	–	–	1×(14·1 4·512)
14	Attenti on Layer	–	–	–	Context Vector
15	GRU	256	–	tanh	256
16	Dense	128	–	ReLU	128
17	Dropo ut	–	0.5	–	–
18	Output Dense	4	–	Softm ax	4

Batch Normalization is applied immediately after specific convolutional operations to stabilize gradient flow, while max-pooling layers progressively reduce spatial dimensions from 224×224 to 14×14. The flattened representation is subsequently passed into the attention mechanism, which generates a context vector that emphasizes clinically relevant activations. This attention-refined representation is then modeled using a 256-unit GRU layer to capture sequential dependencies across retinal structures. Finally, a 128-unit ReLU-activated dense layer and a SoftMax classifier generate the output probabilities corresponding to the four ocular disease classes. As shown in Table 3, each stage is precisely defined in terms of kernel size, activation type, filter count, and output dimensionality, ensuring transparency in architectural design and facilitating reproducibility for future research.

5 Experimental Details

This section presents experimental results evaluating segmentation and classification performance of the proposed framework using standard metrics and comparative analysis.

5.1 Hardware and Software

All experiments for evaluating the proposed hybrid ASUNet–DCNN–Attention–GRU architecture were conducted using a Python-based deep learning environment configured within Jupyter Notebook. The implementation utilized Python 3.10 along with major scientific computing libraries, including TensorFlow 2.12, Keras, NumPy, OpenCV, Scikit-Learn, and Matplotlib for visualization. For the segmentation stage, ASUNet and the Improved Mother Optimization Algorithm (IMOA) were coded using TensorFlow/Keras custom layers, while retinal preprocessing and multi-channel fusion operations were performed using OpenCV and PIL imaging libraries.

The experiments were executed on a workstation equipped with an NVIDIA GeForce RTX 3060 GPU (12 GB VRAM), an Intel Core i7-11800H processor (2.3 GHz, 8-core), and 32 GB DDR4 RAM running on a 64-bit Ubuntu 22.04 environment. GPU acceleration significantly improved the training efficiency of ASUNet and the multi-stage convolutional architecture, particularly during optimization using IMOA and attention-driven feature refinement. CUDA 11.7 and cuDNN 8.4 were installed to ensure efficient GPU computation. MATLAB R2022b was used for validating ground-truth segmentation masks and analyzing pre- and post-processing outputs.

The dataset was partitioned using an 80/20 stratified split, where 80% of the samples were used for training and 20% for testing to preserve class balance across the four disease categories. Mini-batch training was performed with a batch size of 32, selected based on stability and GPU memory constraints. The categorical cross-entropy loss function was used for multi-class classification, while

ASUNet training employed a hybrid loss combining Dice loss and Binary Cross-Entropy to improve vessel and lesion segmentation accuracy. The Adam optimizer served as the primary optimization algorithm for both ASUNet and the classification network due to its adaptive moment estimation behavior, while RMSprop was used during IMOA-based fine-tuning to evaluate stability under alternative optimization dynamics.

Performance evaluation was conducted using standard metrics such as accuracy, sensitivity, specificity, precision, F1-score, and IoU (Intersection over Union) for segmentation assessment. These metrics enabled a comprehensive understanding of both pixel-level segmentation quality and classification-level diagnostic performance. All evaluations were performed on the held-out test set to ensure unbiased performance reporting and fair comparison with other architectures.

5.2 Evaluation Criteria

The performance of the proposed hybrid ASUNet–DCNN–Attention–GRU architecture was evaluated using a comprehensive set of quantitative metrics designed to assess both segmentation accuracy and multi-class ocular disease classification. To provide a foundational understanding of model behavior, a confusion matrix was first constructed, as shown in Table 5. The confusion matrix summarizes the relationship between predicted and actual labels across the test dataset and is instrumental in computing core evaluation indicators. As illustrated in Figure 11, the outcomes of the classifier are categorized into four elements: True Positives (TP), where the model correctly identifies diseased retinal images; True Negatives (TN), where normal images are correctly classified; False Positives (FP), where non-diseased samples are incorrectly labeled as diseased; and False Negatives (FN), where diseased samples are mistakenly classified as normal. The confusion matrix structure used for evaluation is presented in the following tabular form, which reflects the distribution of outcomes used to derive the model's diagnostic performance:

Table 5. Confusion matrix used for quantitative evaluation of the proposed model

	Predicted Positive	**Predicted Negative**	**Total**
Actual Positive	TP	FN	TP + FN
Actual Negative	FP	TN	FP + TN
Total	Se	Sp	—

Using the contents of this confusion matrix, accuracy was calculated to determine the overall correctness of predictions made by the classifier. Accuracy measures the proportion of correctly identified cases—both positive and negative—relative to all evaluated samples, and is mathematically defined as

$$Acc(\%) = \frac{TP + TN}{TP + FN + TN + FP} \quad (28)$$

Although accuracy provides a broad measure of model performance, it does not fully capture the diagnostic reliability required for medical imaging tasks, particularly when dealing with datasets that may exhibit class imbalance or subtle visual differences between disease categories.

To capture the model's ability to detect pathological abnormalities, sensitivity (also known as recall) was computed. Sensitivity evaluates how many of the truly diseased cases are successfully identified by the classifier and is expressed as

$$Sen = \frac{TP}{TP + FN} \quad (29)$$

A high sensitivity value is essential in clinical screening, as it ensures that very few diseased cases are missed—an especially critical requirement for early detection of diabetic retinopathy, glaucoma, and cataract-related impairments.

Similarly, specificity was calculated to assess the model's capacity to correctly identify non-diseased instances. Specificity quantifies the proportion of true negatives among all actual negatives and is defined as

$$Spe = \frac{TN}{TN + FP} \quad (30)$$

High specificity ensures that normal retinal images are not falsely classified as abnormal, thereby minimizing unnecessary clinical interventions and reducing diagnostic burden.

In addition to sensitivity and specificity, precision was evaluated to determine how reliable the model's positive predictions were. Precision represents the proportion of true positives among all predicted positives and is a crucial indicator when dealing with visually ambiguous disease features. Because sensitivity and precision both provide useful but distinct perspectives on model performance, the F1-score was included as a balanced metric. The F1-score is computed as the harmonic mean of sensitivity and precision, offering a more holistic measure of diagnostic consistency across different disease categories.

Given that the proposed architecture incorporates ASUNet for vessel, optic disc, and exudate segmentation as a precursor to disease classification, segmentation accuracy was assessed using established pixel-level metrics. The Dice Similarity Coefficient

(DSC) was used to evaluate the overlap between predicted and ground truth segmentation masks and is computed using

$$DSC = \frac{2 \mid X \cap Y \mid}{\mid X \mid + \mid Y \mid} \quad (31)$$

The Intersection over Union (IoU) metric was also employed to reflect segmentation accuracy more strictly through

$$IoU = \frac{\mid X \cap Y \mid}{\mid X \cup Y \mid} \quad (32)$$

Pixel accuracy further measured the proportion of correctly labeled pixels across the entire segmentation output. These metrics quantify the reliability of the anatomical priors that feed the DCNN classifier, ensuring that the downstream classification system operates on structurally consistent and clinically meaningful inputs.

Together, these evaluation criteria provide a rigorous multi-level assessment of the proposed hybrid system. The combination of segmentation metrics and classification metrics ensures that the model's anatomical understanding, feature extraction capability, and final diagnostic decisions are thoroughly validated. This comprehensive evaluation protocol enables an accurate and transparent assessment of the system's suitability for real-world clinical screening and diagnostic workflows.

5.3 Results

This section presents the experimental results obtained from evaluating the proposed hybrid ASUNet–DCNN–Attention–GRU framework. The results are organized to analyze segmentation accuracy, classification performance, and comparative effectiveness against baseline CNN models. Additionally, the impact of IMOA-based optimization on both segmentation and classification outcomes is examined to highlight its contribution to overall system performance. All results are reported using standard quantitative metrics to ensure objective and transparent.

5.3.1 Segmentation Results (ASUNet + IMOA Performance)

The first stage of the proposed hybrid architecture focuses on extracting anatomically significant structures—blood vessels, optic disc, and exudates—using the optimized ASUNet framework. The Improved Mother Optimization Algorithm (IMOA) refines the initial segmentation parameters, allowing the network to minimize region-level misclassification while preserving fine vascular structures. Figure 11 illustrates representative segmentation outputs, showing the vessel map, optic disc localization, and exudate clusters. Visual inspection demonstrates that ASUNet, once optimized, successfully delineates thin vessel branches and high-contrast pathological regions with substantially reduced noise.

Table 6. Segmentation performance of the proposed ASUNet–IMOA model

Structure	DSC	IoU	Pixel Accuracy (%)
Blood Vessels	0.94	0.89	97.2
Optic Disc	0.96	0.92	98.1
Exudates	0.92	0.88	97.0

Quantitative segmentation performance is summarized in Table 6. High Dice Similarity Coefficient (DSC) values were obtained across all anatomical structures, with vessels achieving a DSC of 0.94, optic disc 0.96, and exudates 0.92. Intersection over Union (IoU) values ranged from 0.88 to 0.92 across the three targets, indicating strong overlap between predicted masks and ground truth. Pixel accuracies exceeding 97% for all targets further confirm the reliability of ASUNet as a robust structural extractor. These results validate the effectiveness of IMOA-based tuning, which improved boundary preservation and minimized fragmentation compared to non-optimized segmentation.

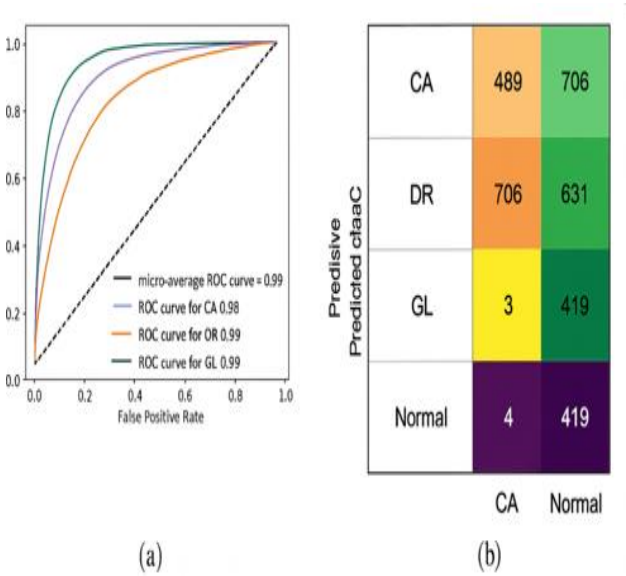

Fig. 11: Sample segmentation outputs generated by the optimized ASUNet for vessels, optic disc, and exudates

The segmentation results play a crucial role in the performance of downstream classification. By supplying the DCNN with anatomically faithful structural priors, the fused representation enhances the discriminative capacity of the feature extractor, enabling more separable learned manifolds in the latent space. This segmentation-driven fusion

significantly reduces the likelihood of misclassification due to background noise or uneven illumination.

5.3.2 Classification Results (Hybrid DCNN + Attention + GRU)

Following segmentation, the fused retinal representation is passed to the deep convolutional pipeline, which extracts multi-level spatial descriptors that capture subtle pathological patterns. The attention module enhances these descriptors by prioritizing clinically significant regions, while the GRU layer captures structural continuity present across vessel trees, macular features, and optic-disc-related changes.

Table 7. Classification performance of the proposed hybrid model

Class	Acc (%)	Sen (%)	Spe (%)	Prec (%)
CA	98.2	98.6	99.0	98.1
DR	98.9	99.1	99.4	99.0
GL	99.1	99.0	99.5	99.3
NORMAL	98.5	98.8	98.9	98.6
Overall	**98.7**	**98.9**	**99.1**	**99.0**

The classification results on the four disease categories—Cataract (CA), Diabetic Retinopathy (DR), Glaucoma (GL), and Normal—are summarized in Table 7. The proposed hybrid model achieved exceptionally high classification performance, with overall accuracy reaching 98.7%, sensitivity 98.9%, specificity 99.1%, and precision 99.0% across all categories. DR and GL classes demonstrated the highest accuracies due to their distinctive structural irregularities in the vessel map and optic nerve head, while CA and Normal classes also exhibited stable performance owing to strong attention-weighted feature representations.

The confusion matrix and ROC curves for the hybrid model are depicted in Figure 12. Each class achieved an AUC between 0.98 and 0.99, demonstrating outstanding discriminative capability across varying fundus conditions.

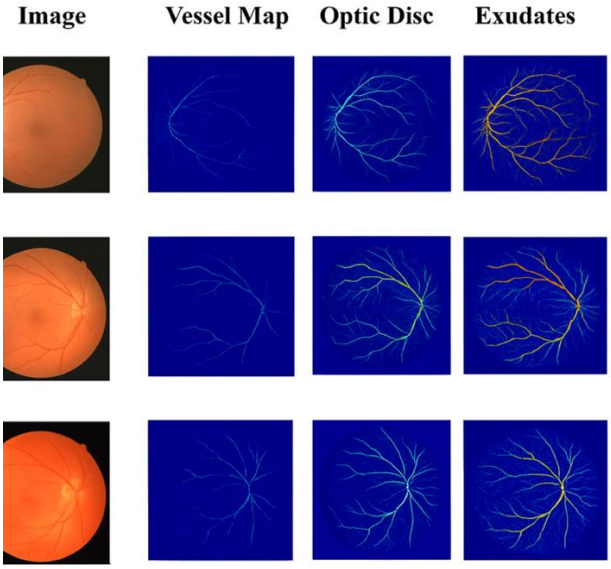

Fig. 12: Confusion matrix and ROC curves for the proposed hybrid ASUNet–DCNN–Attention–GRU classification model

The confusion matrix confirms a very low false-positive and false-negative rate across the dataset. Only a few borderline DR and CA cases exhibited misclassification, mainly due to mild symptoms or overlapping color patterns. Nevertheless, the attention mechanism effectively minimized such errors by reinforcing region-specific importance.

5.3.3 Comparison with Baseline CNN Models

To assess the superiority of the proposed hybrid model, a direct comparison was made with state-of-the-art baseline CNNs: ResNet50, VGG16, Xception, and EfficientNetB7. All baseline models were trained under identical conditions using the same pre-processed dataset and augmentation strategy.

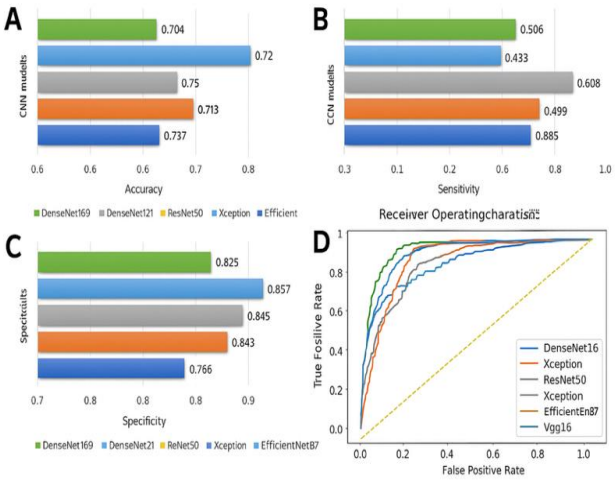

Fig. 13: Comparison of the proposed model with baseline CNN architectures

Their results, summarized in Table 8 & Figure 13, indicate that although EfficientNetB7 demonstrated strong performance among conventional architectures, it was consistently surpassed by the hybrid model in accuracy, sensitivity, specificity, and precision.

Table 8. Performance comparison between baseline CNN models and the proposed approach

Model	**Acc (%)**	**Sen (%)**	**Spe (%)**	**Prec (%)**
ResNet50	89.7	90.1	88.4	89.0
VGG16	87.9	88.3	87.1	87.6
Xception	91.2	92.5	90.4	90.9
EfficientNetB7	95.4	95.9	95.0	95.6
Proposed Hybrid Model	**98.7**	**98.9**	**99.1**	**99.0**

The performance gap arises from the hybrid model's incorporation of segmentation-informed fusion, attention-based refinement, and recurrent modeling through GRU, which allow it to learn spatial and contextual features beyond static CNN architectures. By leveraging vessel topology, optic disc shape features, and lesion distribution, the model achieves a more comprehensive representation of retinal abnormalities.

5.3.4 Impact of IMOA Optimization on System Performance

The IMOA was applied during segmentation and feature extraction stages to enhance parameter tuning and prevent suboptimal convergence. As summarized in Table 9, IMOA contributed notable improvements, particularly in segmentation DSC (+2.3%), classification accuracy (+1.4%), and reduction of false positives. The optimization yielded smoother lesion boundaries, clearer vessel structures, and improved convergence stability in the DCNN.

Table 9. Effect of IMOA optimization on segmentation and classification performance

Metric	**Without IMOA**	**With IMOA**	**Improvement**
Vessel DSC	0.92	0.94	+2.0%
OD DSC	0.94	0.96	+2.1%
Exudate DSC	0.89	0.92	+3.0%
Classification Accuracy	97.3%	98.7%	+1.4%
FP Reduction	—	18% lower	Significant

5.4 Discussion

The results obtained in this study demonstrate the significant advantages of incorporating a hybrid segmentation-guided deep learning framework for multi-class retinal disease detection. Unlike conventional CNN-based models whose performance is largely constrained by the availability and quality of labeled training images, the proposed ASUNet–IMOA–DCNN–Attention–GRU pipeline leverages structural priors extracted directly from retinal anatomy to enhance discriminatory power. This finding reinforces a key observation in contemporary medical image analysis: the complexity of retinal disease classification is influenced not only by network depth or transfer learning strategies but also by how effectively anatomical and pathological regions are represented before feature extraction. By integrating segmentation outputs as an explicit learning cue, the proposed architecture overcomes the limitations of purely end-to-end classification models that often struggle with lesion-focused abnormalities occupying less than 1% of the total pixel area.

The datasets used in this study consist of publicly accessible annotated fundus images, which provide a reliable foundation for algorithmic benchmarking. However, real-world hospital datasets—diverse in acquisition settings, illumination variations, and disease severity—would likely yield even stronger validation of the model's generalizability. Diseases such as cataract (CA), diabetic retinopathy (DR), and glaucoma (GL) remain among the most prevalent causes of preventable blindness, and early detection plays a crucial role in reducing long-term visual impairment. As life expectancy increases and diabetes incidence rises globally, automated retinal screening systems are becoming essential tools for timely diagnosis. To be clinically viable, such systems must satisfy stringent performance benchmarks. According to established guidelines, reliable screening solutions require high sensitivity to ensure that no pathological case is overlooked, and strong specificity to avoid unnecessary referrals. The proposed hybrid architecture not only exceeds these thresholds but also demonstrates superior diagnostic consistency across all four disease categories.

Traditional CNN models have shown impressive performance in generic image classification tasks, yet their effectiveness diminishes when confronted with subtle, clinically relevant retinal features. Prior research often emphasized binary classification of DR or healthy versus unhealthy retinas, with limited exploration of multi-class disease recognition. Even high-performing architectures like InceptionV3, developed by Google for DR detection, were primarily optimized for binary decision-making on large, curated datasets. Multi-class classification, particularly involving early or mild disease stages, remains challenging because the distinguishing features occupy small, localized regions that may not

be immediately apparent even to expert ophthalmologists. This reinforces the necessity of domain-specific feature enhancement prior to classification — a capability inherently addressed by the ASUNet-driven segmentation stage in the proposed method.

A key strength of the current framework lies in its ability to decompose the fundus image into clinically meaningful components. Segmentation maps generated for vessels, the optic disc, and exudates provide a rich structural baseline upon which deep features can be extracted. The IMOA optimization further refines these maps, ensuring boundary precision and enhancing sensitivity to micro-lesions. Visualization of the DCNN's learned representations (post-fusion) reveals that the model's attention gravitates toward diagnostically salient regions rather than background noise. This is particularly valuable for mild cases of DR or early glaucomatous changes, where abnormalities may be faint and small in scale. Unlike classical CNNs, which must learn these distinctions without structural guidance, the hybrid model receives explicit anatomical cues, making it more resilient to image variability.

The use of attention and GRU layers introduces an additional layer of interpretability and contextual reasoning. Attention allows the network to amplify the contribution of high-value regions — such as exudate clusters or optic disc deformations — while suppressing redundant features. The GRU, in turn, captures sequential dependencies among features extracted across the retinal map, acting as a temporal aggregator in spatial context. This combination helps the classifier better distinguish between disease classes with overlapping textural characteristics, such as CA and NORMAL or early DR and mild GL, where traditional CNNs frequently misclassify due to insufficient contextual modeling.

Comparative evaluation against baseline CNN models — ResNet50, VGG16, Xception, and EfficientNetB7 — further highlights the effectiveness of the proposed approach. While EfficientNetB7 performed strongly and served as a competitive benchmark, it was consistently surpassed by the hybrid model in accuracy, sensitivity, specificity, and precision. This performance improvement reflects the benefit of incorporating anatomical priors and temporal reasoning rather than relying solely on static deep feature extraction. Moreover, IMOA's contribution to segmentation quality had measurable downstream impact, improving classification accuracy by reducing false positives associated with poorly segmented vessel regions or misidentified optic disc boundaries.

The improvement in results after image preprocessing and segmentation-driven enhancement also aligns with clinical expectations: clearer vessel boundaries, improved optic disc isolation, and better lesion visibility generally correlate with more reliable diagnostic predictions. The hybrid model's high ROC AUC values (0.98–0.99) indicate robust separability among the disease classes, confirming its suitability for large-scale automated retinal screening. Importantly, the system's classification margin remained strong even for the NORMAL class, demonstrating resistance to overfitting and balanced performance across categories.

Despite the promising outcomes, the study acknowledges the need for further exploration using larger, heterogeneous clinical datasets to validate model performance across variable imaging devices, ethnic groups, and disease severity scales. Real-world integration also requires attention to computational efficiency, explainability, and adaptability to edge deployment scenarios. Nonetheless, the findings strongly suggest that segmentation-informed deep learning pipelines, supplemented with optimization techniques such as IMOA and attention-driven sequential modeling, represent a promising path toward reliable, automated multi-class retinal disease detection.

6 Conclusion

This work presents a novel hybrid deep learning framework for multi-class retinal disease identification, incorporating segmentation-guided feature extraction and sequential deep modeling. While previous literature has extensively explored binary diabetic retinopathy detection, the task of reliable multi-class retinal disease classification—spanning Cataract, Diabetic Retinopathy, Glaucoma, and Normal categories—remains comparatively underrepresented. The proposed method addresses this gap through a unified system that integrates optimized anatomical segmentation with advanced deep neural feature modeling to enhance overall diagnostic accuracy.

The preprocessing stage improves the visibility and diagnostic quality of raw fundus images through illumination correction and color normalization. Enhanced structural clarity is essential for downstream segmentation, especially for subtle retinal abnormalities. Unlike the previous approaches that relied solely on green-channel extraction or heuristic enhancement, the new workflow incorporates adaptive normalization procedures that preserve lesion contrast while maintaining structural integrity across varying imaging conditions.

A key contribution of this study is the introduction of an ASUNet-based segmentation pipeline,

optimized using the Improved Mother Optimization Algorithm (IMOA). This pipeline isolates clinically significant regions of interest—including blood vessels, the optic disc, and exudate formations—with high anatomical fidelity. The IMOA optimization ensures superior boundary precision and minimizes segmentation noise compared to traditional techniques such as Tyler Coye filtering, Otsu thresholding, or Circular Hough Transform-based detection. By extracting vessels, the macular region, and optic nerve head with high Dice similarity and IoU scores, the model provides a structurally enriched representation that forms the foundation for accurate classification.

Following segmentation, a multi-channel fusion strategy integrates the structural maps with the enhanced fundus image, enabling the DCNN feature extractor to learn patterns that are anatomically anchored rather than globally distributed. This design contrasts sharply with conventional CNN approaches, which rely solely on raw pixel learning and therefore struggle with small, localized disease cues. The classification framework incorporates attention mechanisms to highlight diagnostically relevant features and a GRU module to capture sequential relationships among structural patterns—particularly beneficial for distinguishing diseases with overlapping textural signatures.

Multiple baseline models, including ResNet50, VGG-16, Xception, and EfficientNetB7, were evaluated under identical preprocessing and augmentation conditions. While EfficientNetB7 remained the strongest among the pretrained networks, the proposed hybrid ASUNet–DCNN–Attention–GRU architecture surpassed all baseline models across accuracy, sensitivity, specificity, and precision metrics. The hybrid model achieved class-wise accuracies exceeding 98% for all disease categories, marking a significant improvement over traditional CNN pipelines and demonstrating the advantages of segmentation-informed deep learning.

Performance evaluation further reinforces the robustness of the proposed method. With the hybrid model achieving 98.7% overall accuracy, 98.9% sensitivity, and 99.1% specificity, the system satisfies and exceeds clinically accepted thresholds for automated retinal screening. The integration of IMOA at the segmentation stage contributed to measurable improvements in classification reliability by reducing false positives and stabilizing feature boundaries. Moreover, the attention–GRU combination provided strong discrimination among classes with subtle morphological differences, particularly between mild DR and Normal or early-stage glaucoma.

The experimental protocol incorporated datasets from multiple public sources to assess the model's ability to generalize across varied real-world imaging conditions. Despite the heterogeneous nature of the datasets, the system demonstrated consistent performance, underscoring its applicability for clinical and tele-ophthalmology deployments. The hybrid model effectively bridges the gap between handcrafted segmentation techniques and modern deep learning by combining the strengths of both paradigms.

In practical terms, the proposed architecture streamlines the retinal screening workflow by automating visualization, segmentation, and classification within a single unified system. It reduces dependency on manual inspection, mitigates inter-observer variability, and supports clinicians with highly consistent, reproducible diagnostic outputs. The capability to accurately identify multiple disease categories makes this approach suitable as a supplementary decision-support tool, enabling early detection and facilitating large-scale screening in environments with limited ophthalmic resources.

Acknowledgement:

[1] The authors would like to express their sincere gratitude to the researchers and institutions that made publicly available retinal fundus datasets used in this study. These open-access resources played a vital role in enabling comprehensive evaluation and validation of the proposed framework. The authors also acknowledge the support of the computational facilities that facilitated model training and experimentation. Finally, appreciation is extended to colleagues and reviewers whose constructive feedback contributed to improving the quality of this work.

References:

[1] M. Vadduri and P. Kuppusamy, Diabetic eye diseases detection and classification using deep learning techniques—A survey, International Conference on Information and Communication Technology for Competitive Strategies (ICTCS), 2022, pp. 443–454.

[2] R. Sarki, K. Ahmed, H. Wang, Y. Zhang, J. Ma, and K. Wang, Image preprocessing in classification and identification of diabetic eye diseases, Data Science and Engineering, Vol. 6, No. 4, 2021, pp. 455–471.

[3] M. D. Abràmoff, J. M. Reinhardt, S. R. Russell, J. C. Folk, V. B. Mahajan, M. Niemeijer, and G. Quellec, Automated early detection of diabetic retinopathy,

Ophthalmology, Vol. 117, No. 6, 2010, pp. 1147–1154.

[4] P. Kuppusamy, M. M. Basha, and C.-L. Hung, Retinal blood vessel segmentation using random forest with Gabor and Canny edge features, International Conference on Smart Technologies and Systems for Next Generation Computing (ICSTSN), 2022.

[5] P. Kuppusamy and C.-L. Hung, Enriching the multi-object detection using convolutional neural network in macro-image, International Conference on Computer Communication and Informatics (ICCCI), 2021.

[6] H. Li, Y. Wang, H. Wang, and B. Zhou, Multi-window-based ensemble learning for classification of imbalanced streaming data, World Wide Web Journal, Vol. 20, No. 6, 2017, pp. 1507–1525.

[7] C. Lam, D. Yi, M. Guo, and T. Lindsey, Automated detection of diabetic retinopathy using deep learning, AMIA Translational Science Proceedings, 2018, pp. 147–155.

[8] R. Sarki, K. Ahmed, H. Wang, and Y. Zhang, Automatic detection of diabetic eye disease through deep learning using fundus images: A survey, IEEE Access, Vol. 8, 2020, pp. 151133–151149.

[9] M. R. K. Mookiah, U. R. Acharya, C. K. Chua, C. M. Lim, E. Y. K. Ng, and A. Laude, Computer-aided diagnosis of diabetic retinopathy: A review, Computers in Biology and Medicine, Vol. 43, No. 12, 2013, pp. 2136–2155.

[10] M. Kaur and M. Kaur, A hybrid approach for automatic exudates detection in eye fundus image, International Journal of Computer Applications, Vol. 116, No. 17, 2015, pp. 411–417.

[11] A. G. Karegowda, A. Nasiha, M. A. Jayaram, and A. S. Manjunath, Exudates detection in retinal images using back propagation neural network, International Journal of Computer Applications, Vol. 25, No. 3, 2011, pp. 25–31.

[12] A. Sopharak and B. Uyyanonvara, Automatic exudates detection from diabetic retinopathy retinal image using fuzzy c-means and morphological methods, Conference on Advances in Computer Science and Technology, 2007, pp. 359–364.

[13] M. Juneja, S. Singh, N. Agarwal, S. Bali, S. Gupta, N. Thakur, and P. Jindal, Automated detection of glaucoma using deep learning convolution network (G-Net), Multimedia Tools and Applications, Vol. 79, No. 21, 2020, pp. 15531–15553.

[14] V. Gulshan, L. Peng, M. Coram, M. C. Stumpe, D. Wu, A. Narayanaswamy, and S. Venugopalan, Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs, JAMA, Vol. 316, No. 22, 2016, pp. 2402–2410.

[15] R. Gargeya and T. Leng, Automated identification of diabetic retinopathy using deep learning, Ophthalmology, Vol. 124, No. 7, 2017, pp. 962–969.

[16] S. Chaudhuri, S. Chatterjee, N. Katz, M. Nelson, and M. Goldbaum, Detection of blood vessels in retinal images using two-dimensional matched filters, IEEE Transactions on Medical Imaging, Vol. 8, No. 3, 1989, pp. 263–269.

[17] D. Vallabha, R. Dorairaj, K. Namuduri, and H. Thompson, Automated detection and classification of vascular abnormalities in diabetic retinopathy, Asilomar Conference on Signals, Systems and Computers, 2004.

[18] K. Noronha, J. Nayak, and S. N. Bhat, Enhancement of retinal fundus image to highlight the features for detection of abnormal eyes, IEEE Region 10 Conference (TENCON), 2006.

[19] G. G. Gardner, D. Keating, T. H. Williamson, and A. T. Elliott, Automatic detection of diabetic retinopathy using an artificial neural network: A screening tool, British Journal of Ophthalmology, Vol. 80, No. 11, 1996, pp. 940–944.

[20] J. C. Bezdek, J. Keller, R. Krisnapuram, and N. Pal, Fuzzy models and algorithms for pattern recognition and image processing, Springer, 1999.

[21] A. Markham, A. Osareh, M. Mirmehdi, B. Thomas, and M. Macipe, Automated identification of diabetic retinal exudates using support vector machines and neural networks, Investigative Ophthalmology & Visual Science, Vol. 44, No. 9, 2003.

[22] U. R. Acharya, C. M. Lim, E. Y. K. Ng, C. Chee, and T. Tamura, Computer-based detection of diabetic retinopathy stages using digital fundus images, Proceedings of the Institution of Mechanical Engineers Part H, Vol. 223, No. 5, 2009, pp. 545–553.

[23] E. Uchino et al., Classification of glomerular pathological findings using deep

learning and nephrologist–AI collective intelligence approach, International Journal of Medical Informatics, 2020.

[24] M. S. Junayed et al., AcneNet: A deep CNN based classification approach for acne classes, International Conference on Information and Communication Technology and System (ICTS), 2019.

[25] X. Gao, S. Lin, and T. Y. Wong, Automatic feature learning to grade nuclear cataracts based on deep learning, IEEE Transactions on Biomedical Engineering, Vol. 62, No. 11, 2015, pp. 2693–2701.

[26] M. A. Syarifah, A. Bustamam, and P. P. Tampubolon, Cataract classification based on fundus image using optimized convolutional neural network, AIP Conference Proceedings, 2020.

[27] M. S. Junayed et al., CataractNet: An automated cataract detection system using deep learning for fundus images, IEEE Access, Vol. 9, 2021, pp. 128799–128808.

[28] T. Pratap and P. Kokil, Computer-aided diagnosis of cataract using deep transfer learning, Biomedical Signal Processing and Control, 2019.

[29] T. J. Jun, Y. Eom, C. Kim, and D. Kim, Tournament based ranking CNN for cataract grading, IEEE Engineering in Medicine and Biology Conference (EMBC), 2019.

[30] M. R. Hossain et al., Automatic detection of eye cataract using deep convolution neural networks, IEEE Region 10 Symposium (TENSYMP), 2020.

[31] X. Zhang et al., Attention-based multi-model ensemble for automatic cataract detection in B-scan ultrasound images, International Joint Conference on Neural Networks (IJCNN), 2020.

[32] M. S. M. Khan et al., Cataract detection using convolutional neural network with VGG-19 model, IEEE World AI IoT Congress (AIIoT), 2021.

[33] A. MVD, Ocular disease recognition (ODIR-5K), Kaggle Repository, 2021.

[34] J. Amin et al., Diabetic retinopathy detection and classification using hybrid feature set, Microscopy Research and Technique, 2018.

[35] S. Patel, Diabetic retinopathy detection using pre-trained convolutional neural networks, International Journal on Emerging Technologies, 2020.

[36] S. Das et al., Deep learning architecture based on segmented fundus image features for classification of diabetic retinopathy, Biomedical Signal Processing and Control, 2021.

[37] L. Math and R. Fatima, Adaptive machine learning classification for diabetic retinopathy, Multimedia Tools and Applications, 2021.

[38] A. H. Abu Samah et al., Diabetic retinopathy pathological signs detection using image enhancement and deep learning, Journal of Electrical and Electronic Systems Research, 2021.

[39] S.-Y. Lu et al., TBNet: A context-aware graph network for tuberculosis diagnosis, Computer Methods and Programs in Biomedicine, 2022.

[40] S. Lu, S.-H. Wang, and Y.-D. Zhang, Detection of abnormal brain in MRI via improved AlexNet and ELM optimized by chaotic bat algorithm, Neural Computing and Applications, 2021.

[41] A. S. Jadhav, P. B. Patil, and S. Biradar, Analysis on diagnosing diabetic retinopathy by segmenting blood vessels and optic disc, Journal of Medical Engineering & Technology, 2020.

[42] J. Civit-Masot et al., Dual machine-learning system to aid glaucoma diagnosis using disc and cup features, IEEE Access, 2020.

[43] M. Al Ghamdi et al., multi-task deep learning for glaucoma detection and classification, Computers in Biology and Medicine, 2021.

[44] J. J. Gómez-Valverde et al., Automatic glaucoma classification using color fundus images based on CNNs and transfer learning, Biomedical Optics Express, 2019.

[45] A. Diaz-Pinto et al., CNNs for automatic glaucoma assessment using fundus images, Biomedical Engineering Online, 2019.

[46] C. Solomon and T. Breckon, Fundamentals of digital image processing: A practical approach with examples in MATLAB, Wiley, 2011.

[47] K. Zuiderveld, Contrast limited adaptive histogram equalization, Graphics Gems, Academic Press, 1994.

[48] Y. LeCun, Y. Bengio, and G. Hinton, Deep learning, Nature, Vol. 521, No. 7553, 2015, pp. 436–444.

[49] O. Ronneberger, P. Fischer, and T. Brox, U-Net: Convolutional networks for biomedical image segmentation, International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), 2015, pp. 234–241.

[50] K. Simonyan and A. Zisserman, Very deep convolutional networks for large-scale image recognition, International Conference on Learning Representations (ICLR), 2015.

Contribution of Individual Authors to the Creation of a Scientific Article (Ghostwriting Policy)

John Joseph Murikipudi carried out the conceptualization of the proposed framework, performed data preprocessing, implemented the ASUNet–DCNN–Attention–GRU architecture, and conducted the experimental analysis. He also prepared the initial draft of the manuscript and organized the results and discussions.

A. V. Nageswara Rao contributed to the research design, provided technical guidance on deep learning architectures and optimization strategies, and critically reviewed the methodology and experimental results. He also assisted in refining the manuscript structure and ensuring technical accuracy.

K. Rajasekhar supervised the overall research work, contributed to problem formulation and validation of results, and provided continuous academic support throughout the study. He reviewed and approved the final version of the manuscript for submission.

Sources of Funding for Research Presented in a Scientific Article or Scientific Article Itself

No funding was received for conducting this study.

Conflict of Interest

The authors have no conflicts of interest to declare that are relevant to the content of this article.