Attention-Based Deep Learning Models for Retinal Image Analysis in Diabetic Retinopathy

Mukthi K[1], Saswati Behera[2], Dr Smitha Kurian[3], Dr Krishna Kumar P R[4]
[1]*MTech Student, Department of CSE & Technology, SEA College of Engineering & Technology*
[2,4] *Faculty, Department of CSE, SEA College of Engineering & Technology*
[3] *Professor & HoD Department of CSE, HKBK College of Engineering & Technology,Bangalore*

***Abstract*—Diabetic retinopathy (DR) is an extreme microvascular complication of diabetes mellitus and a major cause of preventable blindness globally. Chronic hyperglycemia damages retinal blood vessels, causing leakage, ischemia, and massive neovascularization, ultimately resulting in vision loss or blindness. With the rising incidence of diabetes – particularly in developing international locations – the burden of DR is increasing unpredictably.**

Early assessment reduces vision loss to a remarkable degree; However, traditional DR screening involves manual examination of retinal fundus images by ophthalmologists. This technique is labor-intensive, extremely expensive, and prone to inter-observer variability. In addition, many rural and underdeveloped regions lack adequate ophthalmologists, making large-scale screening challenging.

I. INTRODUCTION

Diabetic retinopathy (DR) is a serious microvascular complication of diabetes mellitus and a leading cause of preventable blindness globally. Persistent hyperglycemia damages retinal blood vessels, causing leakage, ischemia, and extraordinary neovascularization, ultimately resulting in visual loss or blindness. With the increasing incidence of diabetes – especially in developing countries – the burden of DR is increasing rapidly.

Early analysis reduces vision loss to a great extent; But, traditional DR screening relies on ophthalmologists manually examining retinal fundus images. This approach is labor-intensive, costly, and sensitive to inter-observer variability. Furthermore, many rural and underdeveloped areas lack sufficient ophthalmologists, making large-scale screening challenging.

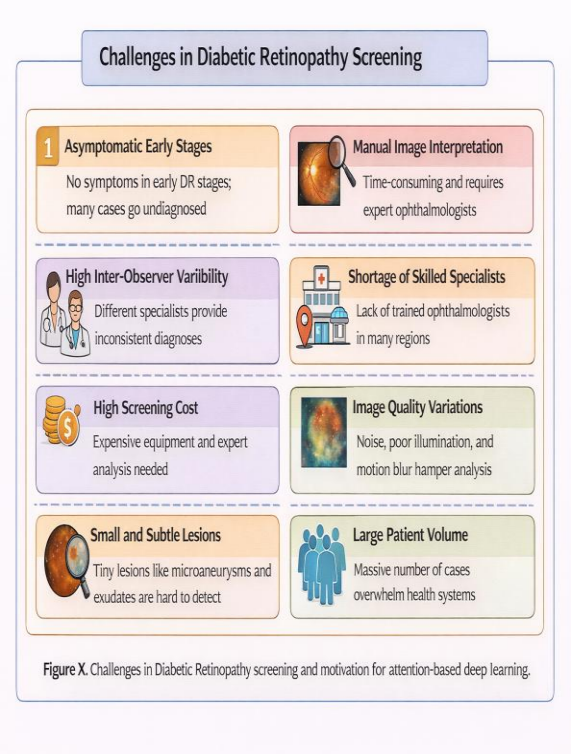

Fig.1: challenges in diabetic retinopathy screening

1.1 OBJECTIVES

- Develop an attention-based DR detection model.
- Use retinal fundus images for automated analysis.
- Apply spatial and channel attention in CNNs.
- Improve lesion detection and classification accuracy.
- Evaluate using EyePACS and APTOS 2019 datasets.

1.2 PROBLEM STATEMENT

- Traditional DR screening relies on manual examination, which is slow and costly.
- Manual analysis is prone to human error.
- Conventional CNN models treat all image regions equally.
- Important lesion areas may be ignored, reducing accuracy.
- An attention-based deep learning system is needed to focus on disease-relevant regions and improve diagnosis reliability.

1.3 EXISTING SYSTEM

In existing diabetic retinopathy (DR) detection systems, retinal image analysis is mainly performed by ophthalmologists or by manual examination through traditional machine learning and CNN-based approaches. Traditional detection relies on visual inspection of fundus images, which is time-consuming and subject to human error. To reduce the workload, CNN-based automated models for classifying disaster recovery severity levels are introduced. Although these models can detect retinal abnormalities, they treat all image areas equally and do not specifically focus on clinically important lesion areas, such as microaneurysms, hemorrhages, and exudates. As a result, subtle pathological features are missed, which leads to a decrease in diagnostic accuracy.

Disadvantages: Existing diabetic retinopathy detection systems rely on conventional CNN architectures that lack attention mechanisms. These models do not effectively highlight disease-relevant retinal regions, which leads to misclassification, especially in early-stage DR. Furthermore, variations in image quality and noise reduce the robustness of these systems. Hence, traditional DR detection approaches provide limited accuracy, interpretability, and reliability for large-scale screening applications.

II. PROPOSED SYSTEM

The proposed images offer an interest-based full depth mastering framework for computerized diabetic retinopathy (DR) grading, designed to overcome the limitations of selective CNNs that frequently overlook small, localized lesions. Instead of processing all image areas equally, the machine uses a convolutional block attention module (CBAM) to selectively focus on clinically large features, including microaneurysms and hemorrhages. The framework consists of four main layers: record acquisition, image preprocessing, attention-based feature extraction, and severity class.

Retinal fundus facts are first pre-processed using CLAHE and normalized to enhance the visibility of the lesion and enhance the photo. These tasks are then processed using an RL-triggered interest mechanism that identifies "which" capabilities apply (channel interest) and "in which" they are located (spatial interest). By integrating these modules, the model suppresses irrelevant background noise and strictly focuses on pathological anomalies, mimicking the diagnostic method of an ophthalmologist.

The device uses deep backbone networks (such as ResNet) to deal with high-dimensional image information, while constraining focus to increase sensitivity. The number one goal is to increase detection accuracy and reduce counterfeit-specific costs for early-degree DR. Designed for scientific reliability, the gadget offers extreme scalability for real-world screening in tele-ophthalmology and public health environments.

2.1. SYSTEM ARCHITECTURE

The proposed machine architecture is divided into four main functional modules: retinal image acquisition, image preprocessing, interest-based feature extraction, and diabetic retinopathy shape. Fundus photographs of the retina are first accumulated. from benchmark clinical datasets and converted to a high-resolution digital feature representation.

These images undergo specialized normalization and contrast enhancement before being processed by an attention-based deep learning engine, which serves as the main diagnostic component of the system.The attention module selectively weights relevant pixels—focusing on pathological lesions while suppressing healthy retinal background—to learn optimal diagnostic policies through feature refinement. Based on these learned spatial and channel priorities, the model classifies the retinal images into specific DR severity levels..

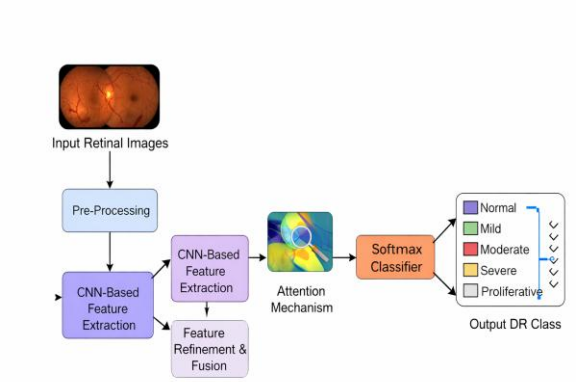

Fig.2: Architecture of the attention-based diabetic retinopathy detection system.

2.2. WORKING PRINCIPLE

1. The system starts by receiving a retinal fundus image as input.
2. The image is preprocessed with CLAHE and normalization, And deep features are extracted with CNN.
3. A focus method highlights important regions and focuses on lesion-related features.
4. Pure symptoms are used to classify diabetic retinopathy stage from No DR to Proliferative DR.
5. A focus heatmap is created showing the final diagnosis and affected areas.

2.3 ADVANTAGES

1. Pathology Localization: Automatically identifies and targets specific lesions like microaneurysms and exudates rather than processing the entire image equally.
2. Visual Interpretability: Generates attention heatmaps that show doctors exactly which retinal regions influenced the diagnosis, increasing clinical trust.
3. Early-Stage Sensitivity: Dual-attention layers detect subtle, minor retinal changes that are often missed by traditional global feature extraction methods.
4. Artifact Resilience: Effectively ignores background noise, such as poor lighting or camera lens artifacts, to ensure consistent diagnostic performance.
5. Improved Precision: Reduces false-positive rates by suppressing non-clinically significant features, leading to more reliable screening results.

III. DATASET AND PRE-PROCESSING

3.1 DATASET

To evaluate the performance of proposed attention-based deep learning model, publicly available reference retinal fundus image datasets are used. The use of standard data sets ensures a fair comparison with existing diagnostic approaches and supports a consistent evaluation of performance at different levels of diabetic retinopathy (DR) severity.

3.1.1 APTOS 2019 Blindness Detection Dataset

Los APTOS 2019dataset is a widely recognized benchmark provided by the Asia Pacific Tele-Ophthalmology Society. It contains a large collection of high-resolution retinal images captured in rural areas where medical screening is difficult to conduct. This]][dataset is specifically designed to support the automation of DR screening.

The dataset is categorized into five major severity levels:

- Stage 0: No DR (Healthy retina)
- Stage 1: Mild NPDR (Microaneurysms only)
- Stage 2: Moderate NPDR (Hemorrhages and exudates)
- Stage 3: Severe NPDR (Extensive hemorrhages and cotton-wool spots)
- Stage 4: Proliferative DR (Neovascularization and vitreous hemorrhage)

Each image acts as the input state for the attention-based model, where the attention layers must learn to identify tiny pathological features like red dots (microaneurysms) or bright spots (exudates) against the retinal background.

3.1.2 EyePACS Dataset

The EyePACS dataset is a high-volume collection of fundus photographs sourced from clinical settings with varying camera models and lighting conditions. This variety makes it an excellent resource for testing the robustness of the attention mechanism.

Key attributes include:

- Real-world clinical noise and artifacts.
- Diverse retinal pigmentation and pupil dilation levels.
- Highly imbalanced classes, reflecting actual population distributions.

3.2 PRE-PROCESSING STEPS

Before training, retinal images undergo several preprocessing operations to ensure their uniqueness and consistency. The set of statistics is then divided into training, validation, and testing subsets to evaluate the generalization of the model. Those steps help stabilize the feature extraction device and improve the accuracy of interest-based localization.

- Image resizing and cropping: Snapshots are standardized to a given decision (for example, $224 to $224 or $512 to $512) and cropped to remove extra black edges, focusing the model's field of view on the retina.
- Evaluation-Indexed Adaptive Histogram Equalization (CLAHE): This technique is applied to improve the comparison of fundus images, making diffuse lesions, including microaneurysms, better visible in the ocular layers.
- Color space normalization: Images are normalized to zero mean and unit variation to lessen the impact of various lights on extraordinary cameras.
- Stat Boost: To address elegance imbalance, strategies including horizontal/vertical flips, rotations, and brightness changes are used. artificially enlarge the instructional set and prevent overfitting.
- Gaussian Filtering: Used to smooth out electronic noise while preserving the edges of critical pathological features.

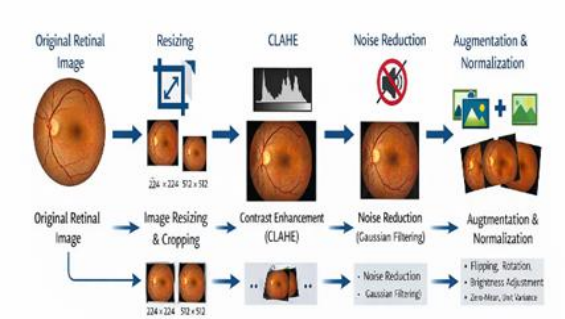

Fig 5: Workflow of Pre-Processing Steps Used in the Proposed Model.

IV. ALGORITHM

The proposed diabetic retinopathy (DR) detection device is based on attention-based deep learning, where the version learns to prioritize pathological capabilities through selective feature refinement. Among the various deep learning architectures, this work employs the Convolutional Block Attention Module (CBAM) incorporated with a convolutional neural community (CNN) due to its effectiveness in localizing subtle medical anomalies in high-resolution photographs.

4.1 Attention Mechanism Formulation

The attention mechanism allows the deep domain version to pay attention to clinically considerable areas of the retinal image, including lesions and abnormalities, while suppressing irrelevant background information. It assigns a significant weight to each extracted feature, called an attention score, which determines how much impact that feature should have on the final prediction.The improved feature representation is obtained using the following formulation:

$F' = A(F) \odot F$

FFF: the original feature map extracted from CNN,

A(F)A(F)A(F): the attention map generated by the attention module,

⊙: indicates multiplication by elements, and

F′F'F′: is the resulting refined feature map that emphasizes pathological patterns related to Diabetic Retinopathy.

4.2 Deep Attention Network (DAN)

Deep Interest Networks extends elegant CNNs by swapping conventional layers with more than just a hobby backbone. This allows the release to explore complicated hierarchical styles, especially related to DR. in the proposed gadget

- Retinal image features form the input layer.
- Attention layers learn to focus on specific lesion patterns (e.g., microaneurysms).
- Output layer predicts the severity grade (Stage 0 to Stage 4).

4.3 Grading Policy (Loss Function)

The learning mechanism is guided by a grading policy designed to maximize diagnostic sensitivity:

- Correct stage classification → Low Loss (Model is encouraged).
- Incorrect stage classification → High Loss (Penalty applied).

- Early Detection focus → Higher weights are assigned to subtle Stage 1 & 2 features to ensure they are not overlooked.

4.4 TRAINING WORKFLOW

- The retinal dataset is first preprocessed using CLAHE and normalization to remove noise and improve lesion visibility.
- These processed images are converted into feature maps and supplied as input to the attention-based CNN model.
- For each image, the model assigns spatial and channel weights to highlight areas of bleed or leakage.
- A classification loss is calculated by comparing the predicted DR stage with the expert-labeled ground truth.
- Model weights and attention parameters are updated through backpropagation to improve the diagnosis policy.
- The training process continues over several epochs until the attention heat maps are accurately aligned with the clinical pathology.
- Techniques such as learning rate determination and batch normalization are used to stabilize the training process.
- Finally, the trained DR detection model is evaluated using a different test set and visualized using Grad-CAM heat maps for diagnostic validation.

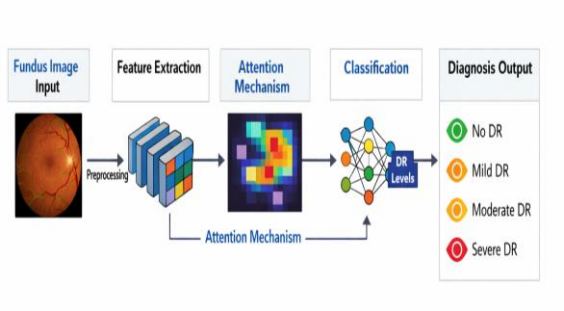

Fig. 6: Workflow of the Attention-Based Diabetic Retinopathy Detection Model.

V. RESULT AND DICUSSION

Evaluation of the proposed attention-based diabetic retinopathy screening model focused on its potential to adequately classify the five levels of the disease while maintaining high clinical interpretability.

5.1 Performance Metrics

To evaluate the effectiveness of the proposed interest enhancement framework, several generalized medical imaging performance metrics are used:

- Accuracy: shows the overall correctness of the version in detecting healthy and diseased retinal images at all levels.
- Accuracy: High accuracy indicates that once the model identifies pathology, it is much more likely to be an appropriate lesion (fewer false alarms for healthy patients).
- Do not forget (sensitivity): high Do not forget is the most important criterion in the detection of DR, as it indicates a high possibility of detecting symptoms in early stages (stages 1 and 2) that would otherwise be ignored.
- F1 Score: This metric provides an excellent balance between accuracy and thoughtfulness, ensuring that the model performs well despite imbalanced data Cohen's Kappa: Measures the level of agreement between the AI's predictions and the ophthalmologist's manual grading, adjusting for chance.

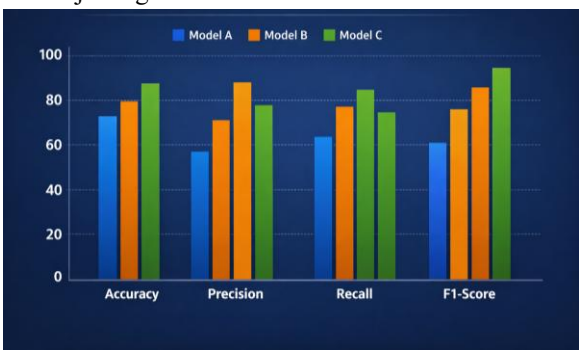

Fig. 7: Performance Metrics Comparison.

5.2. EXPERIMENTAL RESULTS

The proposed attention-based diabetic retinopathy screening tool was evaluated using a reference retinal data set after completion of an iterative learning strategy. To ensure unbiased evaluation, the data were rigorously partitioned into training, validation, and testing subsets. The deep domain version was trained for more than one epoch until the loss function converged and the ocular warmth map consistently aligned with the pathological lesion. Performance metrics including accuracy, precision, sensitivity (don't forget), and F1 scores are calculated on an unbiased validation set.

The results show that the interest-enhanced CNN achieved the best diagnosis accuracy while maintaining a significantly lower false discovery rate compared to the standard global feature CNN. The version showed stable learning behavior, and classification errors progressively decreased as attention levels focused their attention on microaneurysms and hemorrhage. This suggests that the community has learned to correctly distinguish between subtle medical pathology and inadequate retinal history capabilities. The high sensitivity values mainly verify that the machine is highly reliable for detection in the early stages (grades 1 and 2), which is important to prevent permanent loss of vision.

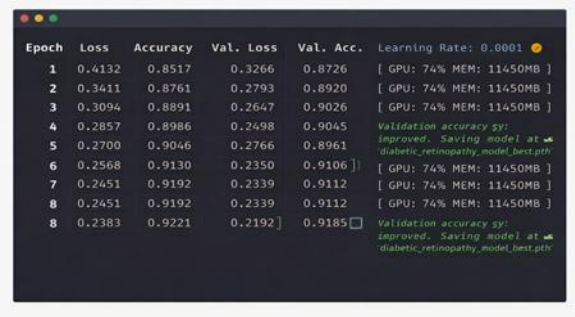

Fig. 8: Console Output and Training Logs of the Proposed Detection Model.

VI. CONCLUSION AND FUTURE WORK

This project provided an in-depth, interest-based study model for diabetic retinopathy. (DR) detection that aimed to increase the diagnostic accuracy and interpretability of automated retinal screening. The proposed system integrates feature-based image analysis with sophisticated focus methods, particularly channel and spatial focus, to intelligently prioritize pathological lesions over healthy retinal tissue.

The findings demonstrate that the attention-based model can:

1. Isolate Subtle Pathologies: Accurately target microaneurysms and small hemorrhages often missed by standard global feature extractors.
2. Provide Clinical Evidence: Generate Grad-CAM heatmaps that offer visual "proof" for its classification, bridging the gap between AI and clinical practice.
3. Ensure high sensitivity: improve the detection rate of early-stage (stage 1 and 2) retinopathy, which is important to prevent irreversible vision loss.
4. Enhance Robustness: Effectively ignore camera artifacts and uneven illumination common in real-world fundus photography.

Compared to traditional CNN-based approaches, the proposed model provides a more transparent and pathology-focused diagnostic mechanism, making it highly suitable for large-scale healthcare programs in resource-limited environments.

VII. FUTURE WORK

Future research may focus on:

- Transformer-Based Integration: Exploring Vision Transformers (ViTs) to model long-range dependencies across the entire retina.
- Mobile and Edge Deployment: Optimizing the model into a lightweight framework for real-time screening on smartphone-based fundus cameras.
- Multimodal Data Fusion: Combining retinal images with patient electronic health records (EHR) such as HbA1c levels and blood pressure for comprehensive risk prediction.
- Federated Learning: Implementing collaborative training across multiple hospitals to improve model diversity without compromising patient data privacy.
- Generative Augmentation: Utilizing GANs to create synthetic images of rare DR stages to further balance training datasets.
- Automated Image Enhancement: Integrating deep learning-based de-noising to automatically "clean" low-quality images before the diagnostic phase.

REFERENCES

[1] R. Oyama et al., “automatic analysis of diabetic retinopathy the use of deep getting to know and interest mechanisms,” magazine of Ophthalmology, vol. 2021, pp. 1–12, 2021.

[2] S. Woo, J. Park, J. Y. Lee, and that i. S. Kweon, “CBAM: Convolutional Block interest Module,” in Proc. european convention on pc vision (ECCV), 2018, pp. three–19.

[3] Asia Pacific Tele-Ophthalmology Society (APTOS), "APTOS 2019 Blindness Detection Dataset," Kaggle, 2019. [Online]. available: https://www.kaggle.com/c/aptos2019-blindness-detection

[4] J. Cuadros and G. Bresnick, "EyePACS: An adaptable telemedicine machine for diabetic retinopathy screening," journal of Diabetes science and technology, vol. 3, no. 3, pp. 509–516, 2009.

[5] V. Gulshan et al., "improvement and validation of a deep learning set of rules for detection of diabetic retinopathy in retinal fundus photos," JAMA, vol. 316, no. 22, pp. 2402–2410, 2016.

[6] okay. Xu et al., "display, attend and inform: Neural photograph caption generation with visual attention," in Proc. global convention on machine gaining knowledge of (ICML), 2015, pp. 2048–2057.

[7] Z. Wang et al., "BMA-internet: Boundary-conscious multi-scale interest network for diabetic retinopathy lesion segmentation," Biomedical signal Processing and control, vol. 70, 2021.

[8] Y. LeCun, Y. Bengio, and G. Hinton, "Deep getting to know, pp. 436–444, 2015.

[9] ok. He, X. Zhang, S. Ren, and J. solar, "Deep residual mastering for photo reputation," in Proc. IEEE convention on laptop imaginative and prescient and sample recognition (CVPR), 2016, pp. 770–778.

[10] R. R. Selvaraju et al., "Grad-CAM: visual causes from deep networks through gradient-based localization," in Proc. IEEE global conference on pc imaginative and prescient (ICCV), 2017, pp. 618–626.

[11] M. D. Abràmoff, Y. Lou, A. Erginay, W. Clarida, and R. Amelon, "progressed computerized detection of diabetic retinopathy the usage of a clinically inspired digital picture evaluation set of rules," Investigative Ophthalmology & visual technological know-how, vol. fifty four, no. thirteen, pp. 8254–8260, 2013.

[12] J. Deng et al., "ImageNet: A massive-scale hierarchical picture database," in Proc. IEEE convention on laptop vision and pattern recognition (CVPR), 2009.

[13] O. Ronneberger, P. Fischer, and T. Brox, "U-internet: Convolutional networks for biomedical photo segmentation," in clinical image Computing and pc-Assisted Intervention (MICCAI), 2015, pp. 234–241.

[14] A. Dosovitskiy et al., "A photo is really worth 16x16 words: Transformers for picture popularity at scale," in Proc. worldwide conference on gaining knowledge of Representations (ICLR), 2021.