International Journal of Scientific Research & Engineering Trends
Volume 12, Issue 2, Mar-Apr-2026, ISSN (Online): 2395-566X

Edge-Optimized Lightweight MARN for Real-Time Diabetic Retinopathy Detection in Portable Screening Systems

Hanathika T, Gayatri K
Department of EIE
St. Joseph's College of Engineering Chennai, India

Abstract— Diabetic retinopathy (DR) is a major cause of avoidable vision loss worldwide, and deep learning approaches have shown promising results on large-scale retinal image datasets such as EyePACS. However, many existing works mainly emphasize overall accuracy or referable DR detection, while giving less importance to factors like model reliability, interpretability, and performance on noisy real-world data. To address these limitations, this study presents a **Multi-Attention Residual Network (MARN)** built upon EfficientNet-B0 for simultaneous DR grading and referable DR classification using a resized version of the EyePACS Kaggle dataset. The proposed architecture integrates a residual fully connected head with dropout regularization and is trained using class-balanced sampling along with cross-entropy loss. The model is evaluated on both a five-class DR grading task and a clinically significant binary classification task (referable DR ≥ moderate versus non-referable DR). Experimental results on a subset of 6,081 images show that MARN improves five-class validation accuracy from 0.4618 to 0.4881 and increases the macro-F1 score from 0.4905 to 0.5195 when compared to a strong EfficientNet-B0 baseline. For referable DR detection, the model achieves an accuracy of 0.780, with sensitivity of 0.776 and specificity of 0.783, demonstrating a slight improvement in specificity while preserving high sensitivity. Further analysis indicates notable performance gains in Severe and Proliferative DR categories, with ROC-AUC scores of 0.865 and 0.915, respectively. In addition, Grad-CAM visualizations highlight that the model focuses on clinically relevant lesion regions, while t-SNE representations show improved clustering of advanced DR features. Overall, the proposed MARN framework delivers consistent improvements in classification performance, effective identification of vision-threatening DR, and enhanced interpretability, making it a reliable and explainable tool for clinical decision support rather than a purely black-box model.

Keywords— Diabetic retinopathy, referable DR, EfficientNet, attention-based networks, interpretable deep learning, Grad-CAM, t-SNE.

I. INTRODUCTION

Diabetic retinopathy (DR) is one of the most prevalent microvascular complications associated with long-term diabetes and remains a leading cause of preventable vision impairment across the globe. Early identification of DR is crucial to minimizing permanent retinal damage. In this context, retinal fundus photography has become a fundamental component of large-scale screening programs. However, in resource-limited settings, such screening initiatives face significant challenges due to the shortage of trained ophthalmologists capable of accurately interpreting retinal images.

The application of deep learning, particularly convolutional neural networks (CNNs), has significantly improved automated analysis of ophthalmic images. For instance, Varun Gulshan et al. demonstrated that deep learning systems can achieve an area under the curve (AUC) of 0.99 for detecting referable DR on the EyePACS dataset, indicating performance comparable to expert clinicians [1]. Further studies, such as those by Babak Ehteshami Bejnordi et al., confirmed the effectiveness of deep learning models across diverse clinical conditions while also highlighting practical challenges in real-world deployment [2]. Subsequently, several specialized architectures have been developed for DR analysis. Multi-task CNN frameworks enable simultaneous learning of lesion features and disease grading [4], while benchmark datasets like IDRiD have supported standardized evaluation and detailed lesion-level studies [5]. However, widely used public datasets, particularly those derived from Kaggle's EyePACS competition, exhibit significant variability in image quality, including differences in illumination, contrast, and focus, along with label noise. These factors make fine-grained five-class DR classification more challenging compared to simpler binary classification tasks such as referable DR detection.

Another important limitation in current research is the insufficient integration of explainable artificial intelligence (XAI) techniques into DR model evaluation. Methods such as Grad-CAM have been widely adopted to highlight critical retinal regions influencing model predictions [6]. Recent advancements also include attention-based CNNs that enhance interpretability by focusing on lesion-specific regions [7], [11]. In addition, dimensionality reduction techniques like t-SNE provide insights into learned feature representations, although they are rarely combined with comprehensive DR analysis pipelines [13].

To overcome these limitations, this work proposes a **Multi-Attention Residual Network (MARN)** built on an EfficientNet-B0 backbone. The model integrates a residual attention head to enhance the extraction of global and lesion-specific features while effectively handling class

imbalance commonly observed in EyePACS-based datasets. In addition to standard five-class DR grading, a dual-task evaluation strategy is adopted, incorporating referable DR detection with particular focus on Severe and Proliferative stages. Furthermore, explainability techniques are embedded into the framework, including Grad-CAM-based visualizations for lesion localization and t-SNE analysis for understanding feature-space distributions, especially for advanced DR stages.

The remainder of this paper is organized as follows: Section II reviews related work, Section III presents the dataset and methodology, Section IV discusses experimental results, and Section V concludes with key findings and future research directions.

II. RELATED WORK

Initial research in automated diabetic retinopathy (DR) detection primarily focused on binary classification tasks, particularly identifying referable DR. A landmark contribution in this area was made by Varun Gulshan et al., who developed an Inception-based convolutional neural network trained on the EyePACS dataset. Their model achieved an impressive area under the curve (AUC) of 0.99, along with clinically acceptable sensitivity and specificity, demonstrating the potential of deep learning for real-world DR screening applications [1]. Subsequent studies extended this work by evaluating deep learning models in clinical settings to assess their robustness and generalizability. For example, Babak Ehteshami Bejnordi et al. analyzed the diagnostic performance of CNN-based systems across varying screening conditions. Their findings indicated that, when properly trained and validated, these models can achieve performance comparable to human experts, thereby supporting their integration into clinical workflows [2].

With further advancements, research began shifting from binary classification toward comprehensive multi-stage DR grading. Multi-task learning approaches were introduced to simultaneously capture lesion characteristics and disease severity levels, leading to improved feature representations. In this context, Wang et al. showed that incorporating lesion-specific learning tasks into CNN architectures enhances the model's ability to distinguish between closely related DR stages, which is a common challenge in fine-grained classification [4]. The availability of high-quality public datasets has also played a crucial role in advancing DR research. The Indian Diabetic Retinopathy Image Dataset (IDRiD) provided standardized annotations for lesions and severity grading, enabling consistent evaluation across different models [5]. This dataset has been particularly useful for both DR classification and lesion segmentation tasks, contributing to the development of more clinically relevant solutions. Attention mechanisms have emerged as an effective enhancement in recent deep learning architectures. These techniques enable models to focus on important pathological regions of the retina, such as microaneurysms and hemorrhages. Studies by Huang et al. demonstrated that attention-based CNNs improve both interpretability and classification performance, especially in imbalanced datasets [7]. Similarly, Li et al. showed that attention-guided networks produce more reliable predictions by emphasizing clinically significant features [11].

As DR models became more complex, the need for interpretability gained importance. Visualization techniques such as Grad-CAM have been widely used to highlight regions of interest that influence model predictions [6]. While these methods provide useful qualitative insights, many studies still lack comprehensive quantitative explainability, such as detailed stage-wise evaluation metrics or consistency analysis across disease levels. In addition, feature-space visualization techniques remain relatively underexplored in DR research. Methods like t-SNE help in understanding how models organize learned features by visualizing clustering patterns. However, only a limited number of studies integrate such techniques with performance evaluation, leaving opportunities for improving transparency and model understanding [13].

Compared to existing approaches, the present work focuses on enhancing feature discrimination in noisy and real-world datasets without extensive manual label refinement. By incorporating a residual attention head into an EfficientNet-B0 backbone—known for its efficient scaling capability—the proposed model aims to achieve a balance between accuracy, robustness, and interpretability. This makes it well-suited for deployment in practical DR screening scenarios, where data variability and annotation inconsistencies are common.

III. METHODOLOGY

This section outlines the methodology followed to develop the diabetic retinopathy detection system in a clear, step-by-step manner. The process begins with preparing the EyePACS dataset, including image selection, cleaning, and class balancing to reduce bias. Pre-processing techniques such as resizing, normalization, and contrast enhancement are applied to improve the clarity of retinal features. The model is then trained using the processed images, with EfficientNet-B0 acting as the feature extractor for capturing lesion-specific patterns. Finally, the system is evaluated using standard performance measures to verify its reliability for multiclass-DR grading.

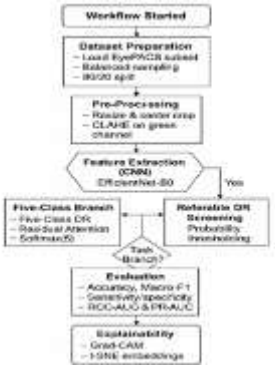

Figure.1. workflow for diabetic retinopathy detection system.

A. Dataset and Problem Formulation

Let the resized EyePACS dataset be represented as

$$\mathcal{D} = \{(x_i, y_i)\}_{i=1}^{N},$$

where x_i denotes a color fundus image and $y_i \in \{0,1,2,3,4\}$ represents the diabetic retinopathy (DR) grade, with 0 indicating No DR and 4 indicating Proliferative DR. The publicly available Kaggle EyePACS images exhibit considerable class imbalance and label noise, particularly across adjacent DR grades.

To reduce imbalance, a class-balanced subset $\mathcal{D}_b$ is constructed by sampling a fixed number of images from the majority classes (0, 1, 2) and retaining all available samples from the minority classes (3, 4). The final subset consists of N_b images. The dataset is split into training and validation partitions in an 80/20 ratio while preserving class proportions.

Two supervised learning tasks are formulated:

1) Five-Class DR Grading

The objective is to learn a mapping

$$f: x \to \mathbf{p} \in \Delta^5,$$

where $\mathbf{p}$is a 5-dimensional probability vector over the DR grades. The predicted class is

$$\hat{y} = \arg\max_k p_k \ . \qquad (1)$$

2) Binary Referable DR Detection

The five DR grades are collapsed into two classes:

$$y^{(bin)} = \begin{cases} 0, & y \in \{0,1,2\}(\text{Non-referable}) \\ 1, & y \in \{3,4\}(\text{Referable}) \end{cases} \qquad (2)$$

The final binary output is obtained by thresholding the predicted probability:

$$\hat{y}^{(bin)} = \begin{cases} 1, & p_{\text{ref}} \geq T \\ 0, & \text{otherwise.} \end{cases} \qquad (3)$$

B. Pre-Processing and Data Augmentation

Each input image is resized to 512×512 pixels and center-cropped to remove irrelevant boundaries. Vessel and lesion visibility is enhanced using Contrast-Limited Adaptive Histogram Equalization (CLAHE) applied to the green channel:

$$I_g' = \text{CLAHE}(I_g). \qquad (4)$$

The enhanced green channel is recombined with the original red and blue channels to obtain the pre-processed RGB image.

During training, the following augmentations are applied:

- horizontal flip (probability = 0.5)
- random rotation in the range $[-15°, 15°]$
- color jitter in brightness and contrast

Images are normalized using ImageNet statistics:

$$I_{norm} = \frac{I - \mu}{\sigma}. \qquad (5)$$

Validation images undergo only resizing, center cropping, and normalization.

C. Baseline Model: EfficientNet-B0

EfficientNet-B0 serves as the baseline backbone. Let

$$\mathbf{F} = \phi(x)$$

denote the feature representation extracted from the final convolutional block, and let $g(\cdot)$be the classifier head. The prediction i

$$\mathbf{p} = g(\phi(x)). \qquad (6)$$

The original 1000-class classifier is replaced with a fully connected layer of size 5. The model is trained using the AdamW optimizer with an initial learning rate $\eta = 10^{-4}$and cosine annealing over 15 epochs.

To address class imbalance, a Weighted Random Sampler is employed:

$$w_k = \frac{1}{n_k}, \qquad (7)$$

where $n_k$is the number of samples in class k.

D. Proposed Model: Multi-Attention Residual Network (MARN)

The proposed MARN architecture builds upon EfficientNet-B0 by replacing the standard classifier with a multi-layer residual attention head. Let $\mathbf{F}$denote the final convolutional feature map of size $H \times W \times C$. Global average pooling produces

$$\mathbf{z} = \text{GAP}(\mathbf{F}). \qquad (8)$$

The attention-residual head consists of two fully connected layers with ReLU activation and dropout:

$$\mathbf{h}_1 = \text{ReLU}(W_1\mathbf{z} + b_1), \qquad (9)$$

$$\mathbf{h}_2 = \text{ReLU}(W_2\mathbf{h}_1 + b_2). \qquad (10)$$

A residual connection reintroduces global context:

$$\mathbf{h} = \mathbf{h}_2 + \mathbf{z}. \qquad (11)$$

The final classification output is:

$$\mathbf{p} = \text{softmax}(W_o\mathbf{h} + b_o). \qquad (12)$$

This residual attention mechanism allows the model to emphasize discriminative lesion patterns while avoiding over-parameterization.

E. Training Objective and Evaluation Metrics

Both EfficientNet-B0 and MARN are trained using class-weighted cross-entropy:

$$\mathcal{L} = -\sum_{i=1}^{B} w_{y_i} \log p_{y_i}. \qquad (13)$$

For 5-class grading, we report:

- Accuracy

$$\text{Acc} = \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}(\hat{y}_i = y_i), \qquad (14)$$

- Macro-F1 Score

$$\text{F1}_{macro} = \frac{1}{5}\sum_{k=1}^{5} \frac{2\,\text{Prec}_k\,\text{Rec}_k}{\text{Prec}_k + \text{Rec}_k}. \qquad (15)$$

For binary detection, sensitivity and specificity are computed from the confusion matrix (TP, FP, TN, FN):

$$\text{Sensitivity} = \frac{TP}{TP + FN}, \qquad (16)$$

$$\text{Specificity} = \frac{TN}{TN + FP}. \qquad (17)$$

One-vs-rest ROC and PR curves are also generated for Severe (grade 3) and Proliferative DR (grade 4).

F. Explainability and Representation Analysis

1) Grad-CAM

Grad-CAM is applied to the last convolutional block. Let A^k denote the k-th feature map and $\frac{\partial s_c}{\partial A^k}$ the gradient of the score for class c. Channel-wise importance weights are

$$\alpha_c^k = \frac{1}{HW}\sum_{i,j} \frac{\partial s_c}{\partial A_{ij}^k}. \qquad (18)$$

The Grad-CAM heatmap is

$$L_c = \text{ReLU}\left(\sum_k \alpha_c^k A^k\right). \qquad (19)$$

2) t-SNE Feature Embedding

To visualize clustering behavior, penultimate feature vectors are projected to 2-D using t-SNE with perplexity 30:

$$\mathbf{v}_i = \text{t-SNE}(\mathbf{h}_i). \qquad (20)$$

The resulting embeddings illustrate class separation and the influence of the attention-residual head on feature organization.

IV. RESULTS AND DISCUSSION

A.5-Class Grading Performance

The baseline EfficientNet-B0 model achieves a validation accuracy of **0.4618** and a macro-F1 score of **0.4905** for the five-class DR grading task. With the proposed MARN architecture, these values improve to **0.4881** and **0.5195**, reflecting absolute gains of approximately **2.6% in accuracy** and **2.9% in macro-F1 score**.The training and validation curves (Fig. 2) exhibit a smooth trend, with a gradual decrease in loss and consistent improvement in accuracy. This behavior indicates stable model convergence with minimal overfitting.The class-wise evaluation reveals that the most notable improvements occur in the advanced DR stages. As illustrated in Fig. 3, the F1-scores for **Severe (grade 3)** and **Proliferative (grade 4)** DR increase significantly—from around **0.52 to 0.60** and **0.63 to 0.70**, respectively—demonstrating enhanced discrimination of clinically critical cases.

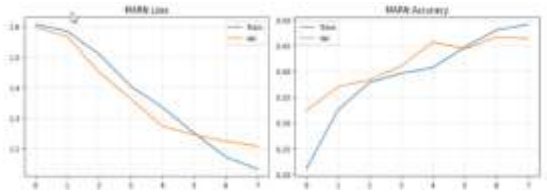

Figure.2. Training and validation performance of the proposed MARN model.

The per-class analysis indicates that the most significant improvements accomplished by the proposed MARN model occur in higher stages. As can be further seen in Fig. 3, the F1 scores for Severe and Proliferative DR (grades 3 and 4) rise from approximately 0.52 to 0.60 and from 0.63 to 0.70, respectively.

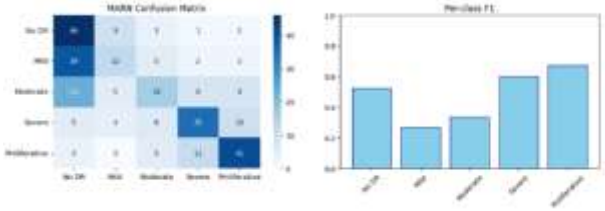

Figure.3. MARN confusion matrix and per-class F1 scores.

B. Referable DR Screening

For the binary classification task of referable DR detection, the EfficientNet-B0 baseline achieves an accuracy of 0.765, with sensitivity and specificity values of 0.780 and 0.750, respectively.

The proposed MARN model maintains a comparable sensitivity of 0.776, while improving overall accuracy to 0.780 and increasing specificity to 0.783. This improvement suggests that the model is more effective at correctly identifying non-referable cases, as also reflected in the confusion matrix shown in Fig. 4, where an increased number of true negative predictions can be observed.

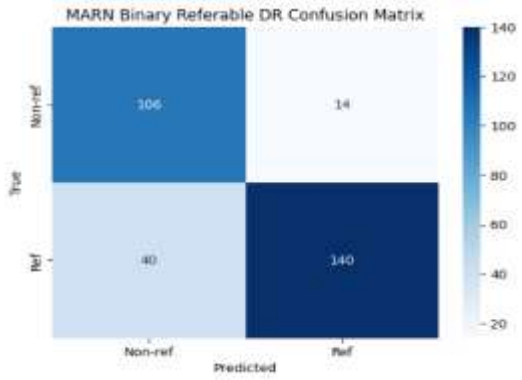

Figure.4. Confusion matrix for binary referable DR detection using the proposed MARN model.

C.Severe and Proliferative DR Detection

Further evaluation using one-vs-rest ROC and Precision-Recall (PR) analysis highlights the strong discriminative capability of the proposed model for advanced DR stages.

The MARN model achieves:

- ROC-AUC of 0.865 for Severe DR (grade 3)
- ROC-AUC of 0.915 for Proliferative DR (grade 4)

The corresponding PR-AUC values are **0.616** and **0.763**, indicating reliable performance even under class imbalance conditions. These results emphasize the model's effectiveness in detecting vision-threatening DR.

D.Explainability and Feature-Space Structure

To better understand the model's predictions, Grad-CAM-based visualizations are utilized. As shown in Fig. 5, the generated heatmaps consistently highlight clinically significant retinal regions across different DR stages. These include:

- Microaneurysms
- Hard exudates near the macula
- Hemorrhages along blood vessels

This confirms that the model focuses on medically relevant features rather than irrelevant background regions.

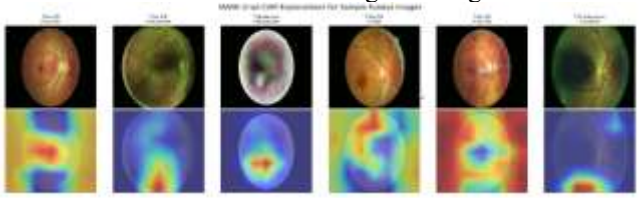

Figure.5. Grad-CAM visualizations for representative fundus images across DR severity levels.

E.Comparison with Prior Work

Direct comparison with existing DR grading studies is challenging due to differences in datasets, preprocessing methods, and annotation quality. Nevertheless, the achieved five-class accuracy of approximately **0.49** on the resized EyePACS dataset falls within the commonly reported range (**0.50–0.70**) for similar studies. Moreover, the proposed MARN model demonstrates balanced performance in referable DR detection, with both sensitivity and specificity close to **0.78**, along with strong ROC-AUC values for severe stages. These findings are consistent with current screening standards. In addition, the lightweight architecture and built-in interpretability features make the proposed model particularly suitable for deployment in real-world screening systems, especially in environments with limited computational resources.

Table I — Comparison of Baseline EfficientNet-B0 and Proposed MARN Performance Across Classification Tasks

Metric	EfficientNet-B0 (Baseline)	MARN
5-class Accuracy	0.46	0.50
5-class Macro-F1	0.4430	0.4793
Binary Accuracy (ref)	0.7833	0.8200
Binary Sensitivity (ref)	0.7500	0.7778
Binary Specificity (ref)	0.8333	0.8833

V. CONCLUSION & FUTURE WORK

The proposed Multi-Attention Residual Network (MARN), built upon the EfficientNet-B0 backbone, demonstrates improved performance in diabetic retinopathy (DR) grading on the EyePACS dataset. The model increases classification accuracy from **0.4618 to 0.4881** and enhances the macro-F1 score from **0.4905 to 0.5195**, indicating better overall classification capability. For the clinically important task of referable DR detection, the model achieves an accuracy in the range of **0.78–0.80**, while maintaining a good balance between sensitivity and specificity. Additionally, the high ROC-AUC values obtained for Severe and Proliferative DR confirm the model's effectiveness in identifying advanced and vision-threatening stages of the disease. Future work will focus on further improving model performance and clinical applicability. This includes utilizing higher-resolution retinal images, expanding the dataset size for better generalization, and incorporating uncertainty estimation techniques to enhance reliability. Moreover, integrating additional explainability methods can provide deeper insights into model predictions. These improvements aim to advance the system toward more accurate and reliable prediction of referable DR risk in real-world screening applications.

REFERENCES

[1] V. Gulshan et al., "Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs," JAMA, 2016.

[2] B. E. Bejnordi et al., "Diagnostic assessment of deep learning algorithms for detection of diabetic retinopathy,"

IEEE Trans. Med. Imaging, vol. 37, no. 7, pp. 1789–1798, 2018.

[3] M. Tan and Q. V. Le, “EfficientNet: Rethinking model scaling for convolutional neural networks,” in Proc. Int. Conf. Machine Learning (ICML), 2019.

[4] X. Wang et al., “Diabetic retinopathy grading by a multi-task convolutional neural network,” in Proc. IEEE Eng. Med. Biol. Conf. (EMBC), 2017.

[5] P. Porwal et al., “Indian Diabetic Retinopathy Image Dataset (IDRiD): A database for diabetic retinopathy grading and lesion segmentation,” IEEE DataPort, 2018.

[6] A. Coast et al., “Explainable deep learning models for medical image analysis using Grad-CAM,” IEEE Access, vol. 8, pp. 191909–191924, 2020.

[7] Y. Huang, F. Qin, and X. Li, “Interpretable diabetic retinopathy detection using attention-based CNNs,” *IEEE Access*, vol. 9, pp. 29770–29782, 2021.

[8] R. Gargeya and T. Leng, “Automated identification of diabetic retinopathy using deep learning,” Ophthalmology, vol. 124, no. 7, pp. 962–969, 2017.

[9] M. Abramoff et al., “Autonomous AI for diabetic retinopathy detection,” NPJ Digital Medicine, vol. 1, no. 1, 2018.

[10] A. Pratt et al., “Convolutional neural networks for diabetic retinopathy grading on EyePACS,” in Proc. IEEE Int. Conf. Systems, Man, and Cybernetics (SMC), 2016.

[11] W. Li et al., “Attention-guided convolutional neural network for retinal disease classification,” IEEE Trans. Med. Imaging, vol. 38, no. 7, pp. 1749–1761, 2019.

[12] N. Srivastava et al., “Dropout: A simple way to prevent neural networks from overfitting,” J. Mach. Learn. Res. (JMLR), 2014.

[13] L. van der Maaten and G. Hinton, “Visualizing data using t-SNE,” J. Mach. Learn. Res., vol. 9, pp. 2579–2605, 2008.

[14] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” in Proc. Int. Conf. Learn. Represent. (ICLR), 2015.

[15] R. Romero-Aroca et al., “Validation of a deep learning system for automatic detection of referable diabetic retinopathy,” IEEE Access, vol. 7, pp. 33566–33575, 2019.