Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.202X.XXXXXXX

RetinaSys: Generalizable Real-Time Diabetic Retinopathy Detection System

AMAN SWAR[1], K. VIJAYALAKSHMI[1], A. MAHESHWARI[1*], K. SARAVANAN[2*], G. SUMATHY[1], MAJED ALSAFYANI[3], SAEED RUBAIEE[4], AMR YOUSEF[56*], NARAYANAMOORTHI R[2]

[1]Department of Computational Intelligence, SRM Institute of Science and Technology, Kattankulathur, Chennai, 603203, India (e-mail: as2275@srmist.edu.in; vijaylak@srmist.edu.in; 78mahee@gmail.com; sumathyg@srmist.edu.in)
[2]Department of Electrical and Electronics Engineering, SRM Institute of Science and Technology, Kattankulathur, Chennai, 603203, India (e-mail: saravanank96@gmail.com; narayanamoorthi.r@gmail.com)
[3]Department of Computer Science, College of Computers and Information Technology, Taif University, P.O. Box 11099, Taif 21944, Saudi Arabia (e-mail: alsufyani@tu.edu.sa)
[4]Department of Industrial and Systems Engineering, University of Jeddah, Saudi Arabia (e-mail: salrubaiee@uj.edu.sa)
[5]Electrical Engineering Department, University of Business and Technology, Jeddah, 23435, Saudi Arabia (e-mail: a.yousef@ubt.edu.sa)
[6]Engineering Mathematics Department, Faculty of Engineering, Alexandria University, Alexandria 21544, Egypt

Corresponding authors: A. Maheshwari (e-mail: 78mahee@gmail.com), K. Saravanan (e-mail: saravanank96@gmail.com) and Amr Yousef (e-mail: a.yousef@ubt.edu.sa).

ABSTRACT Diabetic Retinopathy (DR) stands out as one of the leading causes of preventable blindness, but existing screening procedures are often inefficient, and the use of deep learning models faces significant hurdles in adaptability, usability, interpretability, and deployment in real-world clinical settings. We present RetinaSys, an innovative framework designed to overcome these critical limitations through strong generalization, real-time performance, and clinical interpretability. Leveraging self-supervised MoCo v3 pre-training on diverse datasets, RetinaSys employs a ConvNeXt backbone fine-tuned in a multi-task paradigm, integrating lesion-centric attention, ordinal grade consistency, and domain adaptation. Explainable AI (XAI) techniques, like attention maps, integrated gradients, SHAP, and Monte Carlo dropout, significantly enhance interpretability along with model optimization, which ensures efficient inference on standard CPU hardware. Comprehensive testing demonstrates competitive performance with strong ordinal agreement (Quadratic Weighted Kappa 0.80) and AUC/F1 scores comparable to state-of-the-art models, while maintaining practical efficiency post-optimization. RetinaSys represents a significant step forward, bridging the gap between sophisticated AI and its practical implementation in critical DR screening workflows.

INDEX TERMS Diabetic Retinopathy, Deep Learning, Computer Vision, Medical Imaging, Explainable AI, Real-Time Systems, Self-Supervised Learning, Multi-Task Learning, OpenVINO, ConvNeXt.

I. INTRODUCTION

Diabetic retinopathy (also referred to as diabetic eye disease) is a medical condition where the retina is damaged as a result of diabetes and is the most common cause of blindness among working-age adults [1]. The condition occurs when high blood sugar levels cause damage to the blood vessels in the retina, resulting in vision loss and possible blindness if left untreated. The prevalence of DR is serious, occurring in as many as 80 percent of those with both type 1 and type 2 diabetes for 20 years or longer [2]. Worldwide, DR is responsible for 5 percent of total blindness, with a specific severe impact on people 20 to 64 years of age. The intersection of chronic hyperglycemia, inflammation, and microvascular impairment produces a pathophysiological cascade threatening both life and vision, requiring immediate clinical and public health prioritization.

In the past, diabetic retinopathy was diagnosed using retinas that are highly refined, with an experienced ophthalmologist examining the photographs for criteria such as microaneurysms, hemorrhages, and exudates to stage the disease. While this is an accepted methodology in practice today, it is not particularly efficient. Conducting a screening of the population-scale is laborious, and indeed impossible to fully keep up with, given the lack of trained individuals and rising demand, especially in countries with minimal trained personnel. In addition, the subjective nature of interpretation

brings time-based variability in the diagnosed condition, including missing early-stage findings. The ophthalmologist is a finite and expensive resource, meaning in much of the low- and middle-income world specialized retinal screening services are either unavailable or entirely unavailable. This means that many communities and individuals are at risk of not being diagnosed with diabetic retinopathy.

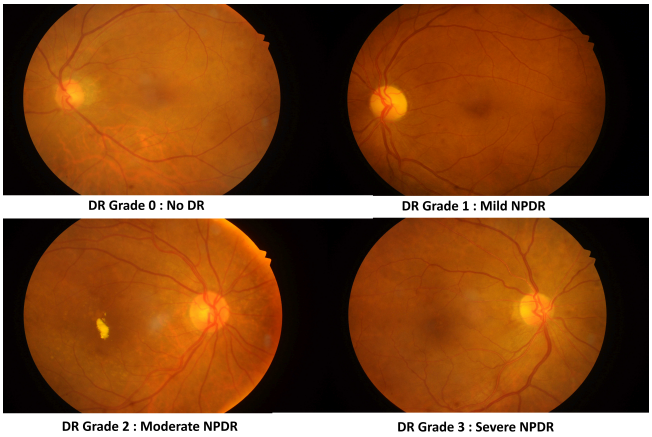


FIGURE 1: Retinal fundus images illustrating different levels of Diabetic Retinopathy (DR).

Subsequently, we have entered the era of machine learning and deep learning, which has further revolutionized the potential for autonomous diagnosis of diabetic retinopathy (DR). Considerable developments have occurred with the application of artificial intelligence (AI) systems, in particular with deep learning methods such as Convolutional Neural Networks and Vision Transformers. These approaches aim to remove reliance on human graders as well as increase screening capacity [3]. For example, the IDx-DR system [4] which has been approved by the US Food and Drug Administration (FDA), is an autonomous system in which retinal images are analyzed to provide evidence of referable DR, and demonstrated that AI provided high sensitivity and specificity for referable DR (> 85%) compared to human graders [5]. Regardless of the potential of AI, existing deep learning-based solutions for detecting DR have many limitations to their real-world deployment. As evidenced in the review by [2], many models have exhibited poor generalizability across different populations and datasets, as shown in [6], [7]. Further to this, the "black box" nature of deep learning models presents a significant barrier to clinical adoption. [8] demonstrate the growing necessity of this explainable AI aspect to build trust among clinicians, which was also pointed out in [9]. There are also many models that are computationally expensive and are a challenge to possible real-time deployment in resource-poor settings [10]. [11] pointed out the lack of efficiency focus and challenges in real-time use. Most research is directed at generating state-of-the-art models with the highest achievement metrics numbers without consideration for model size related to real-time deployments, but without considering the following computational limitations even the most accurate models may never see broad practical use.

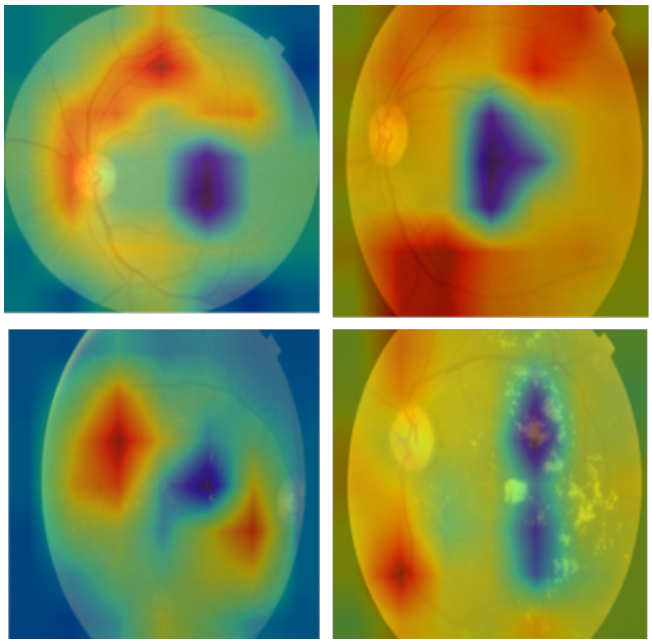


FIGURE 2: Attention maps illustrating model focus on different levels of diabetic retinopathy.

To overcome these challenges, we present RetinaSys, a novel deep learning model for diabetic retinopathy (DR) detection that enhances generalizability, explainability, and deployment efficiency. The model integrates three key components within a multi-task framework: (1) a lesion-attention mechanism, inspired by CBAM [12], that highlights clinically relevant features like microaneurysms and hemorrhages; (2) a grade consistency head that enforces the ordinal relationship of DR severity grades from 0 to 4, thus reducing critical misclassifications; and (3) a domain-adversarial module with gradient reversal to ensure robustness across diverse datasets. These components are integrated under a custom Ordinal Domain Loss function, which balances classification accuracy, ordinal consistency, and domain invariance. Taking advantage of a MoCo v3-pre-trained ConvNeXt backbone [13], RetinaSys achieves robust feature representations across datasets. To support clinical confidence, Explainable AI (XAI) techniques, such as attention maps and SHAP [14], provide transparent insights into decision-making. Optimized with OpenVINO for real-time inference on standard CPUs . RetinaSys enables efficient deployment in resource-constrained settings, offering a practical and scalable solution for DR screening.

II. RELATED WORK AND CHALLENGES

A. ADVANCES IN DEEP LEARNING FOR DIABETIC RETINOPATHY DETECTION

The rise of deep learning has changed the nature of diabetic retinopathy (DR) detection, resulting in models that possess exceptional sensitivity and specificity, frequently equaling or

surpassing the ability of trained human ophthalmologists to diagnose the disease [15], [16]. For the past 10 years, convolutional neural networks (CNNs) have been at the forefront of DR detection image processing, on account of their ability to independently perform hierarchical feature representations of data from retinal fundus images. VGGNet, Residual Networks (ResNets), and Inception Networks have been adopted into DR detection models due to their distinctions in capturing varying aspects of feature complexity. For instance, ResNet-101 is a deeper architecture that has reported an impressive accuracy rate of 98.82% on the Kaggle DR dataset [17]. Such capabilities, however, rely on unsupervised learning and high-capacity networks as the research has commonly employed large and well-curated training datasets of images. Publicly available datasets such as EyePACS and APTOS 2019 consisting of 88,000 images from the United States and 5,000 images from the Asia-Pacific region, respectively, have set standards in the field [18], [19].

While these datasets offer extensive data elements, a substantial challenge exists towards generalizability; models trained on the dataset may have deteriorated performance when applied to other populations beyond the training distribution, which threatens their dependability for utilization in clinical practice. This is exacerbated by physiologically relevant processes, such as fundus pigmentation, that can vary with ethnic grouping, which may limit visibility of DR lesions, which raises the concern of AI-based screening systems [20]–[22]. For example, a model trained on a predominantly light-pigmented fundus dataset may be ineffective when applied to a dark-pigmented retina, where lesions may be less visible [23]–[25]. These differences raise the need for AI systems trained on datasets in which participants are ethnically diverse and representative of the population to minimize bias and ensure equitable performance amongst the populations of the world.

B. THE IMPERATIVE OF EXPLAINABILITY AND TRUSTWORTHINESS

While high accuracy is still a key goal for AI models in healthcare, the needs for explainability and interpretability are equally important for eliciting clinical acceptance [26]–[28]. The "black-box" quality of deep learning models severely inhibits trust and clinical use in healthcare situations because clinicians require an open rationale for diagnostic conclusions [29]. If they do not have any explicit notion of what went into an AI prediction, healthcare providers may refrain from adding AI-driven tools into their routines or making decisions based on the predictions, especially in critically or life-critical scenarios. Explainable AI (xAI) has emerged as a way to provide extra insight into how complex models make decisions, and established methods to accomplish this include techniques like SHapley Additive exPlanations (SHAP) and Grad-CAM [30] to create explanations of which retinal image features drove the AI model predictions [31]. Therefore, there is a need for a developed cohesive xAI framework that is specific to DR that provides accurate and clinically useful explanations. For example, Figure 1 shows a retinal image with diabetic retinopathy features, and heatmaps like those generated by Grad-CAM (similar conceptually to visualizations presented later in Figure 12) can highlight key regions.

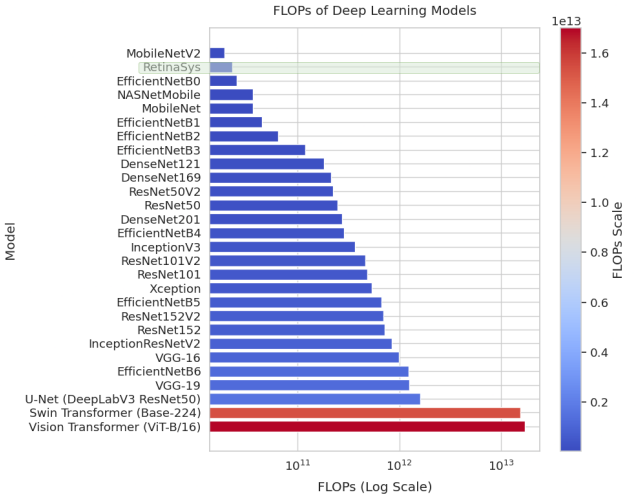


FIGURE 3: Comparison of FLOPS across past studies and our proposed model, illustrating computational efficiency.

C. CHALLENGES IN REAL-TIME CLINICAL DEPLOYMENT

Although DR detection models can achieve high accuracy and interpretability, the deployment of these models may meet additional barriers specific to clinical practice, especially in low-resource settings where computing availability is limited [32], [33]. Real-time deployment is necessary for DR screening to ensure diagnostic wait times do not reduce timely intervention, thus impacting patient outcomes. In terms of accuracy reporting, while many state-of-the-art models achieve high accuracy associated with high computational burden, complex architectures that make models impractical for deployment on edge devices available to clinicians. For example, Vision Transformers (ViT) and their variants, despite their remarkable performance in detecting diabetic retinopathy [34]–[37], face significant challenges for real-time deployment in clinical settings due to their high computational demands. These models present substantial barriers to real time implementation in resource constrained clinical environments where high-performance systems and GPUs are often unavailable. As highlighted in recent research, ViT-based architectures can reach extreme sizes, with ViT-Huge containing more than 632M parameters and extended versions scaling up to 22B parameters, making their deployment impractical without significant optimization [37]. The self-attention mechanism that gives transformers their power has a computational complexity of $O(n^2.d)$ in the input length, creating substantial bottlenecks for processing high-resolution medical images in real-time applications [38].

This quadratic complexity and inefficiencies in processing patch embeddings lead to stringent latency requirements that cannot be met in typical clinical hardware configurations [39]

Figure 3 provides a detailed comparison of the computational complexity (FLOPS) of various models, highlighting the trade-offs between accuracy and deployment feasibility. (Further details on inference time are in Table 3).

D. TOWARD A COMPREHENSIVE DR DETECTION SYSTEM

Growing evidence indicates that the effective use of AI in DR screening is more than simply getting the model right, it requires integrated systems that address both algorithmic sophistication and operational pragmatism. The majority of currently proposed frameworks tend to emphasize marginal gains in accuracy and ignore operational deployment considerations such as computational efficiency, explainability, and workflow integration. A good system for DR detection will need to effectively link advanced deep learning architectures to optimized inference pipelines, rigorous xAI approaches, and real-time deployment considerations. End-to-end solutions such as IDx-DR [40] and DeepDR [41] have shown impressive outcomes when deployed in real clinical settings. Still, they require special hardware and systems to run making it harder to reach places where such facilities are unavailable.

Recent advancements in self-supervised learning (SSL) further improve the potential for enhanced model generalization and data efficiency. Traditional supervised learning depends on using relatively large, carefully annotated datasets. However, in healthcare, this represents a considerable investment of time and resources. SSL circumvents some of this limitation by pretraining models on pretext tasks (e.g., image reconstruction, contrastive learning, predictive coding) that eliminate the need for annotation. The current state of the art—for example, SimCLR [42], BYOL [43], MoCo [44], DINO [45], and I-JEPA [46] performs excellently (e.g., up to 10% increment) against purely supervised learning baselines in chest X-ray classification without using all available labeled samples, and show in general to transfer well to the medical imaging domain. Moreover, the SSL models distill invariant anatomical information and thus combination of these characteristics helps generalization to medical images. Specific to DR detection, this flexibility of pretext task data is particularly useful since lesions may only present as slight aberrations in texture [47], [48]. To capture the advantages of SSL and continue to test its scalability, the methodology employed in this paper DINO and I-JEPA show even more promise, often surpassing supervised baselines regardless of supervised samples being limited [49]–[51].

In summary, the proposed system aims to solve the current gaps in detection of diabetic retinopathy (DR) through the integration of high-performing self-supervised learning (SSL) based pretraining, integrating of attention mechanism in CNN based network, explainable AI (xAI) techniques, and deployment via optimizations like OpenVINO [52] for real-time computation. This holistic approach will ultimately provide a model that not only is accurate, but also interpretable, efficient, and easy to implement capable of addressing the high demands of actual healthcare environments. Figure 4 illustrates the overall training architecture of the proposed *RetinaSys* framework.

III. METHODS

The approach outlined here presents an end-to-end system for the diagnosis of diabetic retinopathy (DR), with a focus on model development, real-time deployment, and explainability through explainable artificial intelligence (xAI) methods. The entire process starts with self-supervised learning (SSL) on a diverse range of retinal fundus image datasets to acquire strong and generalizable features. The model is then subjected to supervised fine-tuning followed by deployment with real-time inference abilities, and lastly utilizes xAI techniques to impart interpretability to clinicians.

A. DATASET

In order to achieve effective feature extraction and generalizability to varied populations and imaging setups, we leverage a multi-source pool of publicly available retinal fundus image datasets for training and validation purposes, as well as robust data augmentation strategies. The training dataset for self-supervised learning and supervised fine-tuning consists of EyePACS [53], DDR [54], APTOS [55], IDRiD [56], and MESSIDOR-2 [57], focusing on diversity in terms of demographics, image quality, and diabetic retinopathy grades. These data sets present a good learning representation, invariant to patient population variance and imaging equipment. In order to test and measure the generalizability of the introduced model, two independent data sets are utilized: SUSTech-SYSU [58] and DeepDR [41], ensuring the model's generalizability. These test sets introduce new data to the model, further challenging and testing its robustness and adaptability. Figure **??** illustrates the dataset composition and balancing.

To improve the strength of our model, we built a tailored augmentation pipeline that can mimic realistic variability seen during imaging while maintaining key pathological features. The pipeline has two distinct augmentation processes applied to the query and key images as mentioned in [44].

The augmentation pipeline starts with a series of transformation including random resized cropping and horizontal flipping, applied to both the query and the key. To enhance the local contrast in retinal fundus images Contrast Limited Adaptive Histogram Equalization (CLAHE) is applied. The query undergoes a more extensive set of augmentation which includes color jittering, grayscale conversion, Gaussian blur and random rotation up to 180 degrees . In contrast the key image undergoes a more conservative set of augmentations primary color jittering and gaussian blur with lower probabilities . Finally both the augmented views are converted to tensors and a random gamma correction is applied independently to

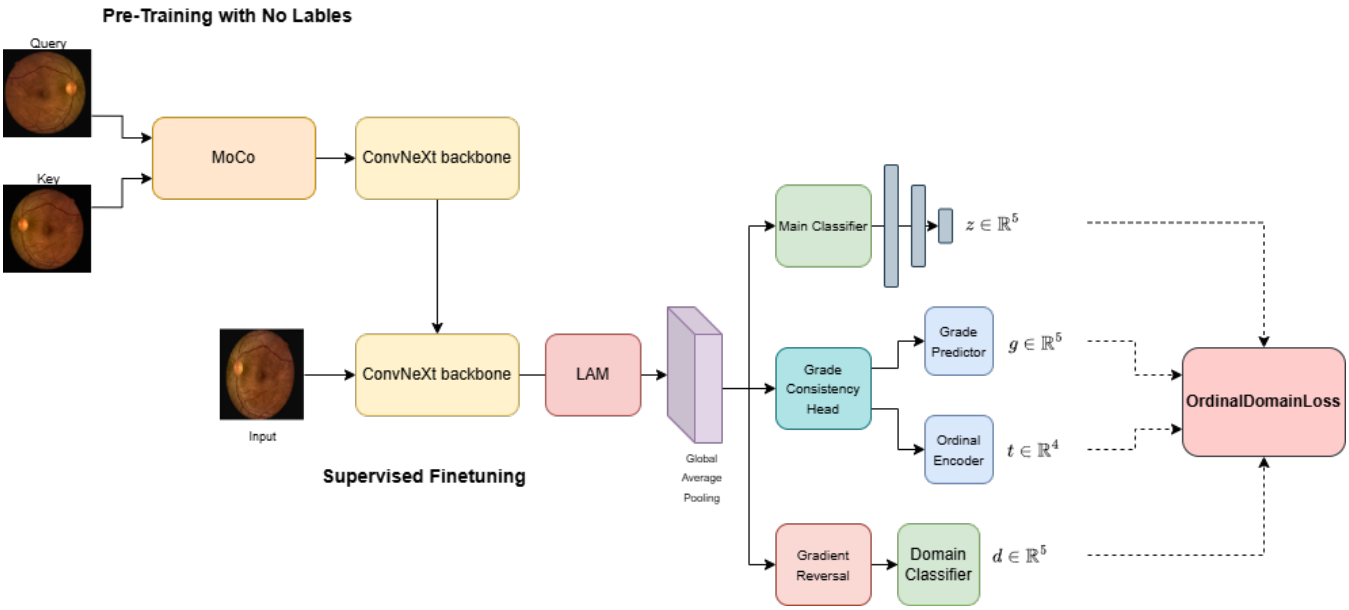


FIGURE 4: Overall architecture of the *RetinaSys* training pipeline, from pre-training to fine-tuning.

each . This asymmetric augmentation strategy with stronger augmentation applied to the query aims to facilitate learning of robust and invariant visual representations by challenging the model to recognize different augmented versions of the same image as similar while distinguishing them from others. Such a well-designed augmentation strategy protects generalizability while preserving the diagnostic validity critical to self-supervised learning in medical imaging.

B. PRE-TRAINING WITH MOCO V3

We pre-train the model using the self-supervised learning method Momentum Contrast(MoCo) (Figure 5). Our motivation for using MoCo comes from its ability as a contrastive learning framework to learn discriminative representations of retinal fundus images. By contrastive positive and negative samples, it encourages a clear separation of different DR grades within the latent space, thus yielding a better classification of retinal fundus images.

It optimizes the InfoNCE loss, a contrastive loss function designed to maximize the mutual information between two augmented views of the same retinal fundus image. Given an input image x, two augmented versions, x_q (query) and x_k (key), are generated using the augmentation pipeline (Figure 6) described in Section III.A. These augmentations are processed by the query encoder f_q and the momentum-updated key encoder f_k, respectively, to produce feature vectors $q = f_q(x_q)$ and $k^+ = f_k(x_k)$. To ensure consistent scale, these feature vectors are L_2-normalized. The InfoNCE loss is defined as

$$\mathcal{L}_{\text{MoCo}} = -\log \frac{\exp\left(\frac{q \cdot k^+}{\tau}\right)}{\exp\left(\frac{q \cdot k^+}{\tau}\right) + \sum_{i=1}^{K} \exp\left(\frac{q \cdot k_i^-}{\tau}\right)}, \quad (1)$$

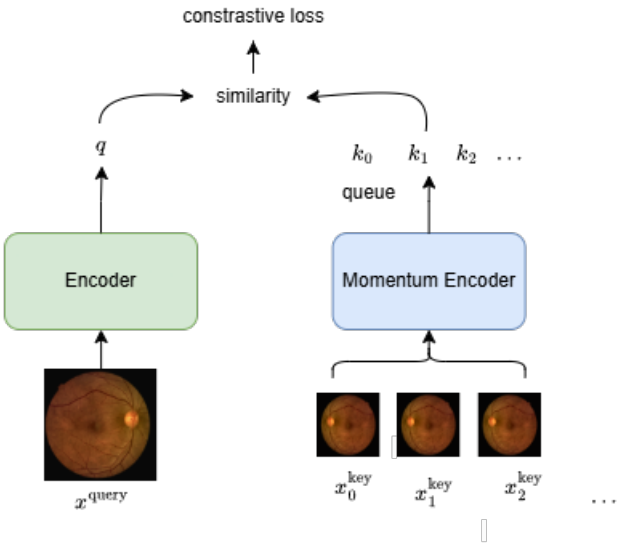


FIGURE 5: MoCo architecture utilized in the pre-training stage of *RetinaSys*.

where $q \cdot k^+$ represents the dot product between the query feature vector q and the positive key feature vector k^+, k_i^- are negative key feature vectors sampled from a dynamic queue of size K, τ is a temperature parameter controlling the concentration of the similarity distribution, and $\exp(\cdot)$ denotes the exponential function. In this implementation, $K = 65536$ and $\tau = 0.2$, as specified in Section IV.A. This loss approximates the maximization of the mutual information $I(x_q, x_k)$ between the query and key views, with a lower bound expressed as

$$I(x_q, x_k) \geq \log(K) - \mathcal{L}_{\text{MoCo}}. \quad (2)$$

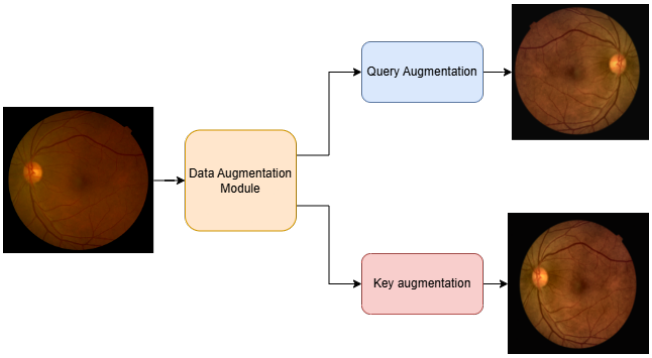


FIGURE 6: Augmentations applied to base images during the MoCo pre-training stage.

By minimizing $\mathcal{L}_{\text{MoCo}}$, the model learns to capture shared semantic content which in our case is diabetic retinopathy (DR) lesions, that remains invariant to transformations like changes in brightness or cropping. The gradient of the InfoNCE loss with respect to the query feature vector q drives the feature learning process and is given by

$$\frac{\partial \mathcal{L}_{\text{MoCo}}}{\partial q} = -\frac{1}{\tau}\left[k^+ - \frac{\sum_{i=0}^{K} k_i \exp\left(\frac{q\cdot k_i}{\tau}\right)}{\sum_{j=0}^{K} \exp\left(\frac{q\cdot k_j}{\tau}\right)}\right], \quad (3)$$

where, for notational convenience, $k_0 = k^+$ and $k_i = k_i^-$ for $i \geq 1$. This gradient encourages the query feature q to be pulled closer to the positive key k^+ while being pushed away from the negative keys k_i^-, thereby enhancing the ability of model to distinguish subtle DR features from normal retinal anatomy. To stabilize training and ensure robust representations, the key encoder f_k is updated using a momentum-based rule defined as

$$\theta_k \leftarrow m \cdot \theta_k + (1-m)\cdot \theta_q, \quad (4)$$

where θ_k and θ_q are the parameters of the key and query encoders, respectively, and $m = 0.999$ is the momentum coefficient controlling the update rate.

We use a ConvNeXt backbone to amplify MoCo's effectiveness. ConvNeXt's large kernels (e.g., 7 × 7) and deep architecture (e.g., 36 layers for ConvNeXt-Small) model both local lesion-specific details and global retinal structures. For an input image $x \in \mathbb{R}^{H\times W\times 3}$, ConvNeXt outputs a feature map $f_q(x) \in \mathbb{R}^d$ (e.g., $d = 768$ for ConvNeXt-Small). Pretraining is performed on the combined EyePACS, DDR, APTOS, IDRiD, and MESSIDOR 2 training datasets, ensuring the feature space captures fine-grained distinctions.

C. FINE-TUNING FOR DR DETECTION

Following the self-supervised learning phase, we utilize the pre-trained ConvNeXt backbone as the feature extractor. The backbone extracts high-level features from retinal images and outputs a feature map $f(x) \in \mathbb{R}^{768\times 8\times 8}$. Unlike traditional approaches, we freeze the backbone for 5 epochs and then unfreeze it, allowing it to adapt to DR-specific features during fine-tuning. Inputs after augmentations are shown in Figure 7.

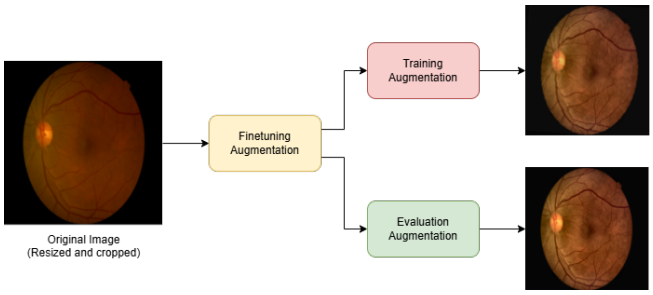


FIGURE 7: Augmentations applied during the fine-tuning stage of *RetinaSys*.

1) Lesion Attention Module (LAM)

To focus on clinically relevant regions (e.g., lesions indicative of DR), we introduce a Lesion Attention Module (LAM) inspired by the Convolutional Block Attention Module (CBAM). LAM enhances feature maps by applying channel and spatial attention sequentially, as detailed in Figure 8.

- **Channel Attention**: This sub-module enhances the importance of informative feature channels relevant to diabetic retinopathy (DR) lesions. Given an input feature map $F \in \mathbb{R}^{D\times H'\times W'}$, where $D = 768$ is the number of channels, and $H' = W' = 8$ are the spatial height and width, respectively, the sub-module first computes global average pooling (GAP) and global max pooling (GMP) across the spatial dimensions. The GAP operation accumulate spatial information by averaging over the height and width, defined as:

$$F_{\text{avg}} = \text{GAP}(F) = \frac{1}{H'W'}\sum_{i=1}^{H'}\sum_{j=1}^{W'} F_{:,i,j}, \quad (5)$$

where $F_{:,i,j} \in \mathbb{R}^D$ denotes the feature vector at spatial position (i, j). Similarly, the GMP operation captures the maximum value across spatial locations:

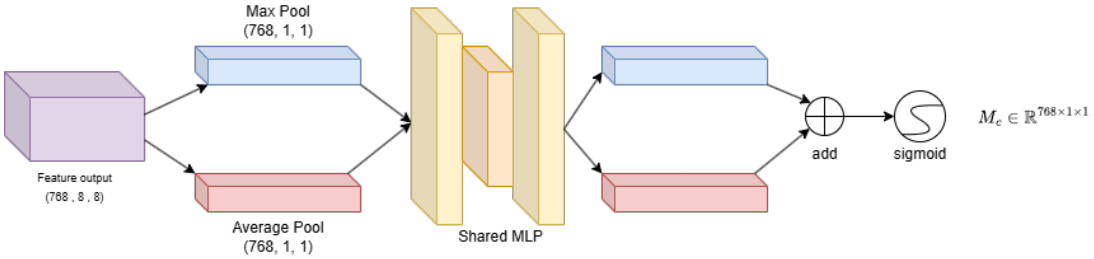
$$F_{\text{max}} = \text{GMP}(F), \quad (6)$$

Both F_{avg} and F_{max} are D-dimensional vectors ($\mathbb{R}^D$). These vectors are processed by a shared multilayer perceptron (MLP), which consists of two fully connected layers with a hidden layer of size D/r, where $r = 8$ is the reduction ratio to reduce computational complexity. The channel attention map is computed as:

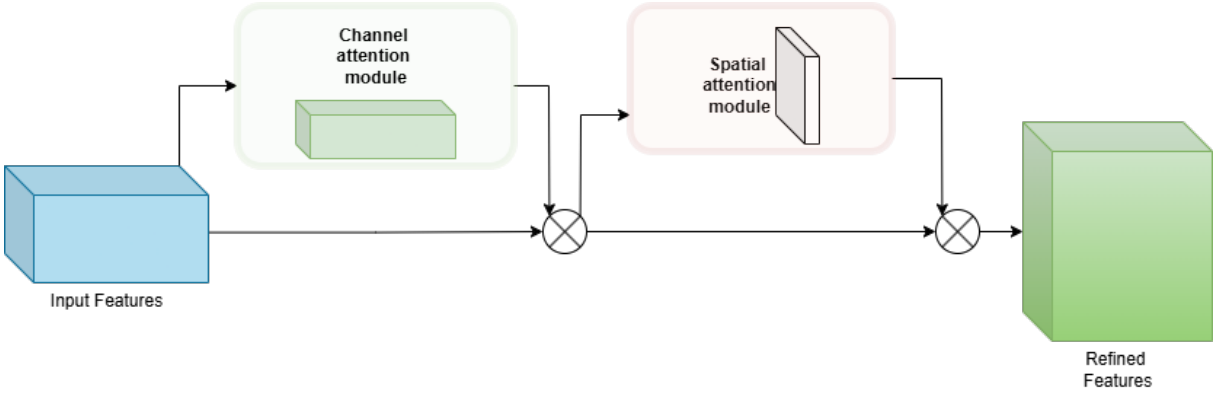
$$M_c = \sigma\left(\text{MLP}(F_{\text{avg}}) + \text{MLP}(F_{\text{max}})\right), \quad (7)$$

where σ denotes the sigmoid activation function ($\sigma(x) = 1/(1+e^{-x})$), and $M_c \in \mathbb{R}^{D\times 1\times 1}$ represents the channel-wise attention weights. The refined feature map is obtained by channel-wise multiplication, denoted by the element-wise product $\odot$:

$$F' = F \odot M_c. \quad (8)$$



(a) Channel attention mechanism in the Lesion Attention Module (LAM).



(b) Overall structure of the Lesion Attention Module (LAM).

FIGURE 8: Custom CNN attention modules inspired by CBAM, used in *RetinaSys*.

This operation amplifies channels with higher relevance to DR lesions (e.g., microaneurysms, hemorrhages) while suppressing noise from irrelevant regions.

- **Spatial Attention**: The spatial attention sub-module focuses on the spatial locations of DR lesions within the refined feature map $F' \in \mathbb{R}^{D\times H'\times W'}$. It computes channel-wise average pooling and max pooling across the channel dimension to emphasize lesion-relevant regions. The average pooling operation is defined as:

$$F'_{\text{avg}} = \text{AvgPool}_{\text{channel}}(F'), \tag{9}$$

where $\text{AvgPool}_{\text{channel}}(F')$ averages the feature values across the D channels at each spatial position, resulting in a map of size $1\times H'\times W'$. Similarly, the max pooling operation is:

$$F'_{\text{max}} = \text{MaxPool}_{\text{channel}}(F'), \tag{10}$$

where $\text{MaxPool}_{\text{channel}}(F')$ selects the maximum value across the channels at each spatial position, also yielding a map of size $1 \times H' \times W'$. These two maps are concatenated along the channel dimension, denoted by $[F'_{\text{avg}}; F'_{\text{max}}]$, producing a tensor of size $2 \times H' \times W'$. This tensor is then processed by a convolutional layer with a 7×7 kernel, followed by a sigmoid activation, to generate the spatial attention map:

$$M_s = \sigma\left(\text{Conv}_{7\times7}([F'_{\text{avg}}; F'_{\text{max}}])\right), \tag{11}$$

The final output of the Lesion Attention Module (LAM) is obtained by element-wise multiplication:

$$F'' = F' \odot M_s. \tag{12}$$

This process highlights spatial regions containing DR lesions, enhancing the model's focus on diagnostically relevant areas.

2) Grade Consistency Head

DR grading is an ordinal task where the severity levels range from 0 (no DR) to 4 (proliferative DR), and misclassifications across distant grades (e.g., predicting "Severe" instead of "Mild") are more critical. To address this, the Grade Consistency Head enforces logical ordering in predictions using two outputs derived from the attention-refined features $F'' \in \mathbb{R}^{D\times H'\times W'}$, where $D = 768$, $H' = 8$, and $W' = 8$ are the number of channels, height, and width, respectively.

- **Grade Predictor**: This component predicts the DR grade by first computing a global average pooling (GAP) over the spatial dimensions of F''. The GAP operation aggregates spatial information into a single vector:

$$h = \text{GAP}(F'') = \frac{1}{H'W'}\sum_{i=1}^{H'}\sum_{j=1}^{W'} F''_{:,i,j}, \tag{13}$$

where $h \in \mathbb{R}^D$ is the resulting D-dimensional feature vector, and $F''_{:,i,j} \in \mathbb{R}^D$ denotes the feature vector at spatial position (i, j). The vector h is then passed

through a multilayer perceptron (MLP) specific to grade prediction, denoted MLP_g, to produce logits for each DR grade:

$$z_g = \text{MLP}_g(h), \quad g \in \{0, 1, \ldots, 4\}, \tag{14}$$

where $z_g \in \mathbb{R}$ represents the logit for grade g, and the set $\{0, 1, \ldots, 4\}$ corresponds to the five DR severity levels (0: no DR, 1: mild, 2: moderate, 3: severe, 4: proliferative).

- **Ordinal Encoder**: To enforce ordinal consistency between grades, this component predicts binary thresholds between adjacent grades. Using the same feature vector h, a separate MLP, denoted MLP_t, computes threshold logits:

$$t_k = \text{MLP}_t(h), \quad k \in \{0, 1, 2, 3\}, \tag{15}$$

where $t_k \in \mathbb{R}$ represents the binary threshold between grades k and $k+1$ (e.g., $k = 0$ separates grades 0 and 1). This dual approach, combining grade prediction with ordinal thresholds, ensures that predictions respect the ordinal nature of DR severity, reducing illogical jumps in classification.

3) Domain Classifier with Gradient Reversal

To mitigate dataset-specific biases and improve generalizability across diverse datasets, RetinaSys employs domain adaptation using a Gradient Reversal Layer (GRL). The domain classifier predicts the source dataset from the feature vector $h \in \mathbb{R}^D$, which is derived from the attention-refined features as described in Section 13. The classifier applies a multilayer perceptron (MLP), denoted MLP_d, to features processed by the GRL:

$$z_d = \text{MLP}_d(\text{GRL}(h, \alpha)), \tag{16}$$

where $z_d \in \mathbb{R}^S$ represents the logits for the S source datasets (e.g., $S = 5$ for EyePACS, DDR, APTOS, IDRiD, and MESSIDOR 2), and $\text{GRL}(h, \alpha)$ is the gradient reversal operation parameterized by a scaling factor α. The GRL behaves as an identity function during the forward pass:

$$\text{GRL}(h, \alpha) = h, \tag{17}$$

but reverses the gradient during backpropagation to adversarially train the feature extractor:

$$\frac{\partial \text{GRL}}{\partial h} = -\alpha I, \tag{18}$$

where $I \in \mathbb{R}^{D \times D}$ is the identity matrix, and α is a scaling factor that increases from 0 to 2 over training epochs to balance domain adversarial learning. By maximizing the domain classifier's loss with respect to the feature extractor parameters, the backbone learns domain-invariant features, enhancing generalization across datasets.

4) Loss Function: OrdinalDomainLoss

The OrdinalDomainLoss combines three terms to jointly optimize diabetic retinopathy (DR) grade prediction, ordinal consistency, and domain invariance. The total loss is defined as:

$$L = L_{\text{main}} + \lambda_c \cdot L_{\text{consistency}} + \lambda_d \cdot L_{\text{domain}}, \tag{19}$$

where L_{main} is the main classification loss for DR grade prediction, $L_{\text{consistency}}$ enforces ordinal consistency between grades, L_{domain} encourages domain-invariant features, and λ_c and λ_d are scalar weights that balance the contribution of each term.

- **Main Loss (L_{main})**: This term uses standard cross-entropy to train the 5-class DR grade prediction. For a batch of N images, the main loss is computed as:

$$L_{\text{main}} = -\frac{1}{N} \sum_{i=1}^{N} \log p(y_i | x_i; z_g), \tag{20}$$

where x_i is the i-th input image, $y_i \in \{0, 1, \ldots, 4\}$ is the true DR grade (0: no DR, 1: mild, 2: moderate, 3: severe, 4: proliferative), and $z_g \in \mathbb{R}^5$ are the predicted logits for the five grades, as defined in Equation 14. The softmax probability for the true class is given by:

$$p(y_i | x_i; z_g) = \frac{\exp(z_{g,y_i})}{\sum_{j=0}^{4} \exp(z_{g,j})}, \tag{21}$$

where z_{g,y_i} is the logit corresponding to the true grade y_i, and the denominator sums over all grade logits.

- **Consistency Loss ($L_{\text{consistency}}$)**: This term penalizes violations of ordinality in DR grading by leveraging the ordinal encoder outputs. Using the threshold logits $t_k \in \mathbb{R}$ (Equation 15) for $k \in \{0, 1, 2, 3\}$, where k represents the threshold between grades k and $k+1$, the consistency loss is computed as a weighted binary cross-entropy (BCE):

$$L_{\text{consistency}} = 0.7 \cdot \sum_{k=0}^{3} \text{BCE}\left(\sigma(t_k), \hat{t}_k\right) + 0.3 \cdot L_{\text{grade}}, \tag{22}$$

where $\sigma(t_k) = 1/(1 + e^{-t_k})$ is the sigmoid activation mapping t_k to a probability, $\hat{t}_k = 1$ if the true grade $y_i > k$ and 0 otherwise, and $\text{BCE}(p, q) = -[q \log(p) + (1 - q) \log(1 - p)]$ is the binary cross-entropy between predicted and target probabilities. The term L_{grade} is an additional BCE loss applied to the grade predictions to ensure consistency, weighted at 0.3 to balance the contributions. This formulation ensures that the model respects the ordinal nature of DR grades, reducing significant misclassifications.

- **Domain Loss (L_{domain})**: This term encourages domain-invariant features by training the domain classifier to predict the source dataset. For a batch of N images, the domain loss is defined as a cross-entropy loss:

$$L_{\text{domain}} = -\frac{1}{N} \sum_{i=1}^{N} \log p(d_i | x_i; z_d), \tag{23}$$

where $d_i \in \{1, 2, \ldots, S\}$ is the source dataset label for the i-th image (e.g., $S = 5$ for EyePACS, DDR, APTOS, IDRiD, and MESSIDOR 2), and $z_d \in \mathbb{R}^S$ are the domain logits predicted by the domain classifier (Equation 16). The probability $p(d_i|x_i; z_d)$ is computed using a softmax over the domain logits, similar to the main loss.

The weights λ_c and λ_d are tuned hyperparameters (e.g., $\lambda_c = 0.5$, $\lambda_d = 0.1$) to balance the influence of each loss term, ensuring effective training across classification, ordinal consistency, and domain adaptation objectives.

D. EXPLAINABLE AI TECHNIQUES

To maximize transparency and interpretability of our model, we implemented a comprehensive suite of Explainable AI (XAI) methods. These included Attention maps, Grad-CAM, Integrated Gradients, SHAP, and Monte Carlo Dropout, each of which served to make clear the model's decision-making rationale, flagging areas of interest, associating feature importance, and estimating the uncertainty of predictions. Each method provides unique insights into the model's behavior and foster trust by clinicians. In the section below we will describe each of the methods, their application and relevance to ophthalmic diagnostics. The outputs are presented conceptually in Figure 12.

1) **Attention Map Visualization**: Attention maps leverage the Lesion Attention Module (LAM), described in Section III.C, which integrates channel and spatial attention mechanisms to highlight clinically relevant regions in retinal fundus images. During inference, the attended feature maps $F'' \in \mathbb{R}^{D \times H' \times W'}$, where $D = 768$, $H' = 8$, and $W' = 8$ represent the number of channels, height, and width, respectively, are extracted after the LAM output. To generate a spatial attention map, these feature maps are averaged across the channel dimension, denoted as $\text{mean}_C(F'')$, resulting in a 2D map of size $H' \times W'$. This map is then normalized to create a heatmap:

$$A = \frac{\text{mean}_C(F'') - \min(\text{mean}_C(F''))}{\max(\text{mean}_C(F'')) - \min(\text{mean}_C(F'')) + \epsilon'}, \tag{24}$$

where $A \in \mathbb{R}^{H' \times W'}$ is the normalized attention map, $\text{mean}_C(F'')$ computes the mean across the D channels for each spatial position, $\min(\cdot)$ and $\max(\cdot)$ denote the minimum and maximum values of the averaged map, and $\epsilon' = 10^{-6}$ is a small constant added to the denominator to prevent division by zero. This heatmap is overlaid on the original retinal fundus image, providing a visual representation of areas—such as lesions, hemorrhages, or exudates—that contributed to the model's prediction. In diabetic retinopathy (DR) classification, where specific retinal findings (e.g., microaneurysms, hard exudates) directly impact the assessed severity, these attention maps link clinical features to model predictions, confirming that the model prioritizes pathologically significant regions and thereby fostering clinician trust.

2) **Integrated Gradients**: Integrated Gradients is an attribution-based XAI method that measures the contribution of each pixel to the model's prediction by integrating gradients along a specified path from a baseline image to the input image. The pixel-level attributions are summed across channels, normalized, and plotted as a heatmap overlaid on the original fundus image. Integrated Gradients provide fine-grained information about the importance of features, making them the best choice to find subtle pathological changes in DR images, like early microaneurysms in mild DR. This fine-grained information allows the precise assessment of whether the model was sensitive to clinically relevant features, which is vital for early detection and accurate grading.
3) **SHAP (SHapley Additive exPlanations)**: SHAP gives the importance of the features according to Shapley values from cooperative game theory, i.e. it is a theoretically justified approach to the explaining of the predictions made by the model. Thereafter, these SHAP values, the individual pixels are then aggregated across the channels, normalized, and lastly, a heatmap of the contribution is created from an image of an input. The attribution framework of SHAP, which is rigorous, is primarily developed for the purpose of assessing the significance of features extracted from the retina in the grading of DR. It seeks to provide a crystal clear and interpretable statement regarding the decision rules made by the model. In the clinic, where the understanding of the relative importance of different features (e.g., exudates vs. hemorrhages) is crucial, SHAP offers explanations to the decision-making process of diagnosticians which are in turn consistent with their observations.
4) **Monte Carlo Dropout**: Monte Carlo Dropout provides estimation of predictive uncertainty by relaxing the dropout layers during the test phase and executing multiple forward passes to mimic the Bayesian approximation of the posterior distribution of the model. When we apply Monte Carlo Dropout to the model, we make several stochastic predictions for a particular fundus image. We take the average prediction (logits) and standard deviation over these passes as the model's confidence and uncertainty, respectively. The uncertainty is demonstrated as a column chart next to the base image, showing the standard deviation per DR grade. The assessment of uncertainty is an essential factor in medical diagnostics, especially for DR, as even small differences between grades (e.g. mild and moderate DR) can have significant clinical implications. Monte Carlo Dropout helps to expose the cases where the model is not so sure and gives the opportunity for additional clinical observation and reduces the risk of overly confident incorrect diagnoses. By demonstrating

the uncertainty, this approach informs clinicians about the dependability of the model's predictions, supporting safer decision-making in screening programs.

E. REAL-TIME DEPLOYMENT

The last part is all about putting the right DR model for its purpose into practice and doing so through immediate real-time inference, using a software that is compatible with CPU-based clinical environments. We make it very easy for the model to work on computers with a modest performance level; to this end, we take advantage of OpenVINO(Open Visual Inference and Neural Network Optimization) by Intel toolkit which among other things can adapt the deep learning models for Intel CPUs and GPUs having integrated graphics. We decide in favor of OpenVINO and against TensorRT from NVIDIA because the former seems a better option for the reason that it would be rare to find a GPU in the settings of a typical medical facility and OpenVINO was created to ensure compatibility with a number of Intel hardware options and thus it is a perfect match for CPU-only systems like regular laptops or workstations. The various methods that were applied mainly the reduction of FP16 precision and the quantization of INT8 were used in the ONNX-converted CPU-deployed model as a source of gaining computational efficiency. These methods have been developed to dramatically decrease memory and to advance the computational speed of the process, in addition, they are assisting in the problem of the large amount of the data flow that now enters the computer system. When we were in the process of improving the quality of these techniques, we were actually in the process of testing the performance of these methods in such a way that they would be of the desired effect through a validation pipeline that was paired with the optimization process.

1) FP16 Precision Reduction

FP16 precision reduction converts the weights of model and activations from 32-bit floating-point (FP32) to 16-bit floating-point (FP16) representation. OpenVINO leverages optimized CPU instructions (e.g., AVX-512) to perform half-precision arithmetic, reducing memory footprint and computational complexity. Consider the weight tensor $W \in \mathbb{R}^{m\times n}$ and activation tensor $A \in \mathbb{R}^{n\times p}$. The FP16 weight is approximately:

$$W_{\text{FP16}} \approx \text{round}(W_{\text{FP32}} \cdot 2^{e_{\max}-e_{\text{FP16}}}) \cdot 2^{e_{\text{FP16}}-e_{\max}}, \tag{25}$$

where $e_{\max}$ is the maximum exponent in FP32, and e_{FP16} adjusts the exponent to the FP16 range ($[-14, +15]$). The forward pass output $Y = W \cdot A$ is similarly scaled, with a relative error bounded by:

$$\epsilon_{\text{FP16}} \approx \frac{2^{-10}}{|W_{\text{FP32}}|}, \tag{26}$$

where 2^{-10} reflects the FP16 mantissa precision (10 bits). This ensures minimal impact on DR grade predictions, provided the weights are sufficiently large. (Note: Exact range depends on FP16 standard implementation details)

2) INT8 Quantization

INT8 quantization transforms the weights of model and activations into 8-bit integer representation, further optimizing inference performance. This method applies post-training quantization, calibrating the model on a representative dataset (a subset of the validation loader) to determine optimal scaling factors. The quantization maps a floating-point value x to an integer q using:

$$q = \text{clip}\left(\text{round}\left(\frac{x}{S} + Z\right), Q_{\min}, Q_{\max}\right), \tag{27}$$

where $S = \frac{\max(|x|)-\min(|x|)}{Q_{\max}-Q_{\min}}$ is the scale factor based on the dynamic range of tensor, Z is the zero-point (typically 0 for symmetric quantization), and $[Q_{\min}, Q_{\max}]$ is the integer range (e.g., [-127, 127] for signed INT8). The dequantized value is reconstructed as:

$$x_{\text{dequant}} = S \cdot (q - Z), \tag{28}$$

with a quantization error approximated by the Root Mean Square Error (RMSE):

$$\epsilon_{\text{INT8}} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - x_{\text{dequant},i})^2}, \tag{29}$$

minimized through calibration to preserve DR-specific features like lesions. This integer arithmetic accelerates computations and reduces memory demands.

IV. EXPERIMENTS

This section outlines the experimental setup designed to validate the proposed *RetinaSys* model for diabetic retinopathy (DR) detection. We evaluate the system's performance across multiple dimensions: feature robustness from self-supervised learning (SSL), classification accuracy and ordinal consistency after fine-tuning, interpretability via explainable AI (xAI) techniques, and real-time inference capability on modest hardware. The experiments leverage diverse datasets, rigorous training protocols, and comprehensive evaluation metrics to ensure the system's effectiveness and generalizability.

A. DATASETS AND PREPROCESSING

1) Dataset Breakdown

The experimental setup utilizes a multi-source collection of retinal fundus image datasets. For Self Supervised Learning using MoCo, we combined EyePACS (88,702 images), DDR (6835 images), APTOS 2019 (5,590 images), IDRiD (413 images), and MESSIDOR-2 (1,744 images), totaling 103,284 images. Then for finetuning we again combined EyePACS(28100 images), DDR(6260 images), APTOS (2929 images), IDRiD (330 images) and MESSIDOR-2 (1220 images). These datasets provide diversity in demographics, image quality, and DR severity (grades 0-4). To assess

generalization, we assessed both test set of trained dataset (EyePACS, DDR, APTOS, IDRiD and MESSIDOR-2) and unseen dataset (SUSTech-SYSU (1151 images) and DeepDR (400 images)).

2) Preprocessing

For MoCo Self-supervised pre-training, both the query and key images were initially subjected to random resize crop with output size of 256 pixels, a scaling factor between 0.2 and 1, followed by horizontal flip with probability of 0.5. Subsequently, CLAHE was applied with a clip limit of 2 and a tile grid size of 8×8. For the query image, we used color jitter with brightness, contrast, saturation, and hue adjustments of ±0.4, ±0.4, ±0.2, and ±0.1, respectively, with a probability of 0.8. Grayscale conversion (probability 0.2), Gaussian blur with kernel size of 23 and a sigma range of (0.1, 2.0) (probability 0.5), and random rotation with degrees in the range of ±180 (probability 0.3) were also applied to the query. The key image received less aggressive augmentations, with Color jitter (same parameters as query) used with a probability of 0.3 and Gaussian blur (same parameters as query) with a probability of 0.5. Finally, a random gamma adjustment was applied independently to both tensors with a gamma value sampled uniformly between 0.7 and 1.3, with a probability of 0.3 for each. Then, for Supervised Finetuning, it begins with resizing the image to 110% of the target image size (in this case, 256) and subsequently applying a center crop. To enhance local contrast again, CLAHE with a clip limit of 3 and a tile grid size of 8×8 was applied. Further augmentations included a random horizontal flip with a probability of 0.3 and a random rotation with a maximum rotation angle of ±10 degrees. Color variations were introduced using color jitter with brightness and contrast adjustments of ±0.2. After converting to tensor, we also normalize it using the ImageNet mean.

B. TRAINING

1) SSL Pre-Training

MoCo v3 with a ConvNeXt-Small backbone was pre-trained on the combined training dataset. Due to computational constraints, training ran for 92 epochs with a batch size 128. We used the Adam optimizer (learning rate: 5×10^{-4}) with a cosine annealing scheduler (T _max = 87 epochs, 5 warm-up epochs, min LR: 1×10^{-5}). MoCo parameters were $\tau = 0.2$, queue size $K = 4096$, momentum $m = 0.99$. Augmentations followed Figure 6. The pretraining took approximately 120 hours on a single NVIDIA A100 GPU (40GB).

2) Fine-Tuning

The pre-trained backbone was not frozen and fine-tuned with the LAM (reduction ratio 8), the Grade Consistency Head, and the domain classifier. Again, due to limited computational resources, we ran the fine-tuning process for 95 epochs with batch size 64. We used AdamW optimizer with LR: 5×10^{-5}, weight decay: 0.01 for initial 5 epochs then unfreezing backbone with LR: 5×10^{-6} for backbone and LR: 5×10^{-5} for classifier and weight decay: 0.01 . The learning rate schedule was managed by One Cycle LR policy with max LR set to 5×10^{-5} till 5 epochs and 5×10^{-5} and 5×10^{-6} for classifier head and backbone respectively after unfreezing the backbone. The OrdinalDomainLoss used weights $\lambda_c = 0.1$ (consistency) and $\lambda_d = 0.05$ (domain adversarial). The GRL scaling factor α increased linearly from 0 to 1 over training. Fine-tuning took 21 hours on the A100 GPU with mixed precision training. Validation was performed on a 40% split of the training data. Training and validation loss curves (Figure 9 and 10) show convergence without severe overfitting. Key validation metrics (F1, QWK, AUC) are plotted in Figure 11.

C. EVALUATION

To assess the performance of RetinaSys in detecting diabetic retinopathy (DR) across diverse datasets, five key metrics are employed: Sensitivity, Specificity, F1 Score, Quadratic Weighted Kappa (QWK), and Area Under the Curve (AUC).

Sensitivity (Recall) measures the proportion of true DR cases correctly identified, which is vital for minimizing missed diagnoses, particularly in early-stage DR. It is defined as:

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \tag{30}$$

where TP (true positives) is the number of correctly identified DR cases, and FN (false negatives) is the number of DR cases incorrectly classified as non-DR.

Specificity indicates the proportion of non-DR cases correctly classified, reducing unnecessary referrals and improving screening efficiency. It is given by:

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}}, \tag{31}$$

where TN (true negatives) is the number of correctly identified non-DR cases, and FP (false positives) is the number of non-DR cases incorrectly classified as DR.

F1 Score balances precision and recall, addressing class imbalance in DR datasets where severe cases are less frequent. It is computed as:

$$\text{F1} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}, \tag{32}$$

where $\text{Precision} = \frac{\text{TP}}{\text{TP+FP}}$ is the proportion of predicted DR cases that are correct, and Recall is equivalent to Sensitivity.

Quadratic Weighted Kappa (QWK) evaluates the agreement between predicted and actual DR grades, penalizing larger discrepancies more heavily, which is crucial for the ordinal multi-class DR staging (from 0: no DR to 4: proliferative DR). It is defined as:

$$\kappa = 1 - \frac{\sum_{i,j} w_{i,j} O_{i,j}}{\sum_{i,j} w_{i,j} E_{i,j}}, \quad \text{where} \quad w_{i,j} = \frac{(i-j)^2}{(N-1)^2}, \tag{33}$$

where $O_{i,j}$ is the observed agreement (number of images assigned grade i by the model and grade j by the ground truth), $E_{i,j}$ is the expected agreement by chance for the

grade pair (i, j), $w_{i,j}$ is the quadratic weight penalizing larger grade differences, $N = 5$ is the number of DR grades, and $i, j \in \{0, 1, \ldots, N-1\}$ represent the possible grades.

Area Under the Curve (AUC) quantifies the model's ability to distinguish between DR severity levels across all classification thresholds, reflecting overall discriminative power. Derived from the Receiver Operating Characteristic (ROC) curve, which plots the true positive rate (TPR = Sensitivity) against the false positive rate (FPR $= 1 -$ Specificity) over all thresholds, AUC is defined as:

$$\text{AUC} = \int_0^1 \text{TPR}(\text{FPR})\, d\text{FPR}, \tag{34}$$

where TPR(FPR) represents the true positive rate as a function of the false positive rate.

V. RESULTS

The quantitative results demonstrate the effectiveness of the *RetinaSys* framework. This section details the performance across various metrics and datasets, highlighting the accuracy, robustness, and efficiency of model post-optimization.

1) Quantitative Metrics

Model performance was assessed on the multiple test sets, including those seen during the training phase (EyePacs, DDR, APTOS, IDRiD, and Messidor) and unseen datasets (SUSTech-SYSU and DeepDR) using standard metrics for multi-class DR classification. As reported in Table 1, *RetinaSys* achieved state-of-the-art or competitive performance on specific datasets like APTOS, IDRiD, and Messidor compared to other published methods, particularly excelling in F1-score and AUC. Overall performance of the model across various datasets is shown in Table 2. The model maintains high specificity across datasets, indicating reliable identification of healthy retinas. While sensitivity varies, the strong Quadratic Weighted Kappa (QWK) scores (generally above 0.80, even reaching 0.91 on IDRiD) demonstrate excellent agreement with ground truth grades, penalizing large errors more heavily, which is clinically important. Furthermore, post-optimization performance can also be seen in Table 3, demonstrating minimal loss (or even slight gain in QWK for INT8 in this run) in key metrics like QWK and AUC, with a significant boost in deployment efficiency (up to 60% memory reduction for INT8).

2) Deployment Testing

Hardware

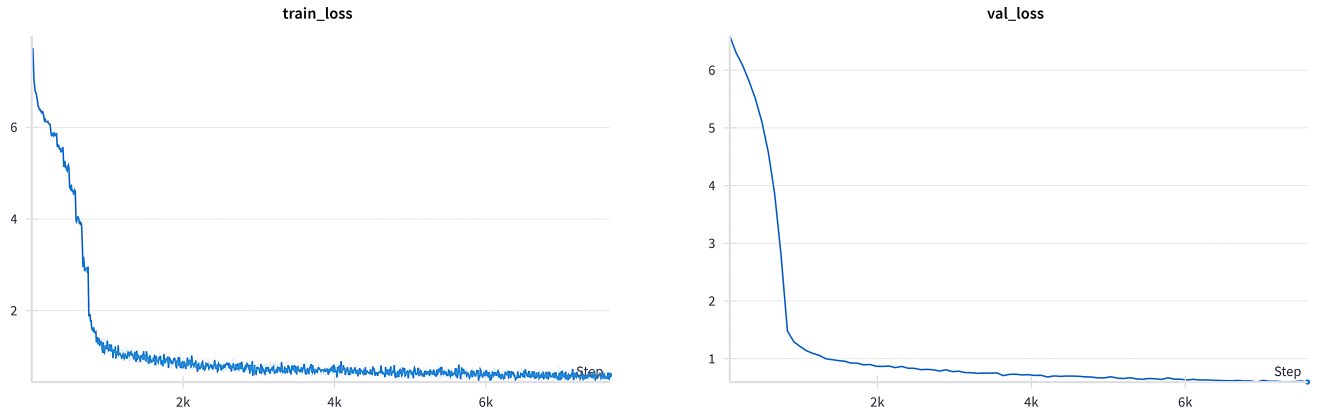
Real-time inference was evaluated on a moderate laptop: Intel Core i5-12450H CPU (8 cores/12 threads) with 8 GB RAM and integrated Intel Iris Xe Graphics (GPU not actively used for inference in this test).

Inference Speed and Optimization Impact

The performance of the optimized diabetic retinopathy (DR) classification models was rigorously evaluated using a comprehensive set of metrics, comparing the FP32 baseline model with its FP16 and INT8 variants to assess the impact of quantization on accuracy, memory usage, and inference speed. The results, derived from a validation pipeline applied to diverse datasets, are summarized in Table 3 and demonstrate the effectiveness of the optimization techniques employed. In terms of accuracy metrics following optimization, the Quadratic Weighted Kappa (QWK) for the FP32 baseline was recorded at 80.42 (on a scale of 0 to 100). The FP16 variant achieved a QWK of 80.18, reflecting a minimal decrease of 0.30%, while the INT8 variant reached 80.70, indicating a slight improvement of 0.35%, suggesting that INT8 quantization not only preserves but marginally enhances the QWK in this evaluation. The weighted F1 score for the FP32 model stood at 60.81%, with the FP16 and INT8 variants registering 61.06% and 60.98%, respectively, demonstrating negligible variation across the models. Additionally, the Area Under the Curve (AUC) for the macro-averaged one-vs-rest (OvR) classification was 87.33 for FP32, 87.51 for FP16, and 87.69 for INT8 (on a scale of 0 to 100), underscoring that classification performance remained robust and, in some instances, slightly improved post-optimization. Regarding performance metrics, memory usage exhibited significant reductions: the FP32 model required 668.11 MB, whereas the FP16 variant reduced this to 440.28 MB, achieving a 34.10% decrease, and the INT8 variant further lowered it to 266.75 MB, resulting in a substantial 60.07% reduction. Timing stability, as indicated by the standard deviation of latency, was consistent across the models, with values of 227.28 ms for FP32, 203.81 ms for FP16, and 241.67 ms for INT8, further affirming the reliability of the optimized models. These results confirm that INT8 quantization offers the most significant memory savings with negligible impact on diagnostic accuracy, making it highly suitable for deployment on resource-constrained hardware like standard laptops or edge devices , thereby supporting their practical utility in clinical settings.

VI. CONCLUSION

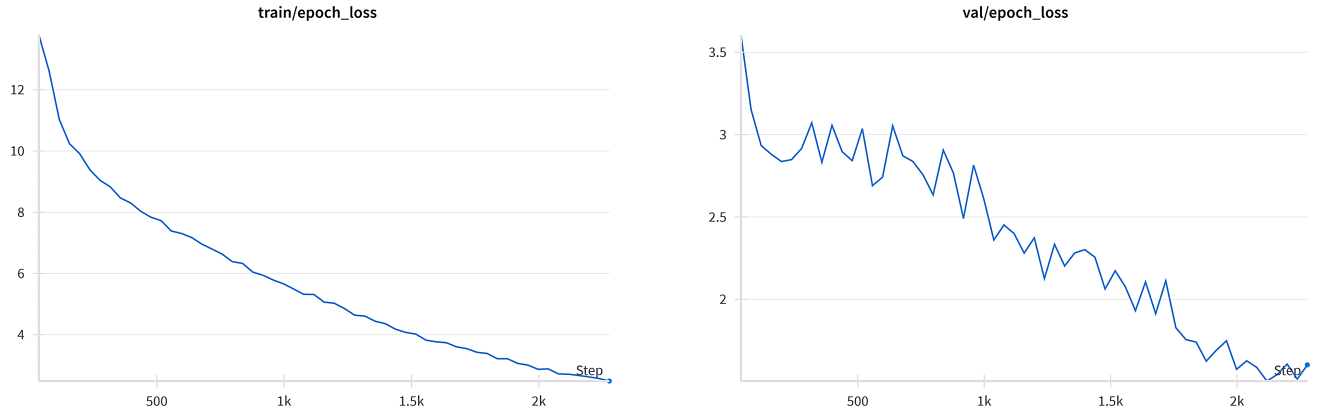
RetinaSys provides a comprehensive framework for diabetic retinopathy (DR) detection, seamlessly integrating self-supervised pre-training for robust feature learning, multi-task learning with ordinal consistency and domain adaptation for accurate grading and generalization, and optimization for real-time, interpretable deployment. The results, summarized in Section V and detailed in Tables 1, 2, and 3, demonstrate strong performance across diverse datasets. Notably, the model achieves a Quadratic Weighted Kappa (QWK) of 80.42 for the FP32 baseline (on a scale of 0 to 100), which is maintained post-optimization, alongside a high specificity as reported in Table 2. The macro-averaged Area Under the Curve (AUC) of 87.33 (baseline, on a scale of 0 to 100) further indicates robust discriminative ability across DR severity levels. Comparison with other models, as shown in Table 1, highlights the competitiveness of RetinaSys, particularly in F1 score and AUC on datasets such as IDRiD and Messidor, despite the wide variety of image



(a) MoCo Pre-Training Loss (Training Set)

(b) MoCo Pre-Training Loss (Validation Set)

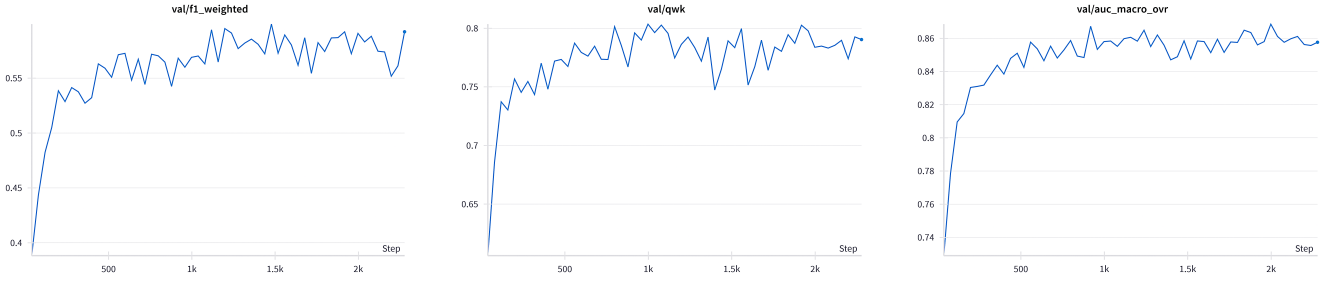
FIGURE 9: Loss curves over epochs during MoCo v3 pre-training.



(a) Finetuning Loss (Training Set)

(b) Finetuning Loss (Validation Set)

FIGURE 10: Loss curves over epochs during fine-tuning.



(a) Validation F1 score (Macro)

(b) Validation QWK

(c) Validation AUC (Macro Avg)

FIGURE 11: Validation performance metrics over epochs during fine-tuning.

qualities and population characteristics in the training and testing data. The high QWK suggests that misclassifications, when they occur, are typically between adjacent grades, which is less clinically severe than large grading errors. The integration of explainable AI (xAI) techniques, visualized comprehensively in Figure 12 and discussed in Section III-D, addresses the critical need for transparency in clinical AI by highlighting relevant features such as microaneurysms and hemorrhages. Moreover, the successful deployment optimization using OpenVINO, as evidenced in Table 3, enables efficient CPU-based inference, making RetinaSys scalable for resource-limited environments without requiring

TABLE 1: Performance Comparison on Selected Datasets (APTOS, IDRiD, Messidor). Metrics: Area Under Curve (AUC), Accuracy (ACC), F1-score. Values are percentages (%).

Target	APTOS (%)			IDRID (%)			Messidor (%)		
Metrics	AUC	ACC	F1	AUC	ACC	F1	AUC	ACC	F1
ERM	75.0	44.4	38.9	82.3	50.0	44.1	79.1	60.7	43.4
DRGen [59]	79.9	58.1	40.2	84.7	44.6	37.4	79.0	60.1	40.5
Mixup [60]	75.3	62.6	43.2	78.8	39.0	27.6	76.7	54.7	32.6
MixStyle [61]	79.0	65.8	39.9	83.0	51.4	39.2	75.2	62.2	36.5
GREEN [62]	75.1	53.8	38.9	79.9	41.3	32.2	75.8	52.0	36.8
CABNet [63]	75.8	55.5	39.4	79.2	44.8	37.3	74.2	56.1	34.1
DDAIG [64]	78.0	**67.1**	41.0	82.1	37.4	27.0	76.6	58.4	35.3
ATS [65]	77.1	56.9	38.3	83.0	41.5	34.9	77.2	64.7	35.8
Fishr [66]	79.2	66.6	43.4	82.7	40.3	27.6	76.4	**65.1**	41.1
MDLT [67]	77.3	57.2	41.5	81.5	44.2	35.4	75.4	58.9	36.9
GDRNet [68]	79.9	66.8	46.0	84.0	40.3	35.9	83.2	63.4	50.9
RetinaSys (Ours)	**85.2**	61.4	**72.9**	**90.9**	**72.3**	**82.6**	**88.7**	62.2	**74.2**

TABLE 2: Model Performance on Different Datasets. Values are percentages (%), except QWK and AUC (unitless, scale 0-100 assumed based on values).

Metric	**Test Set of Seen dataset**					**Unseen Datasets**	
	EyePacs	**DDR**	**APTOS**	**IDRiD**	**Messidor**	**SUSTech-SYSU**	**DeepDR**
Sensitivity (%)	60.80	61.79	62.91	75.36	64.55	63.24	51.76
Specificity (%)	91.69	91.89	90.35	92.97	90.57	90.29	89.67
F1 (%)	71.90	72.73	72.91	82.63	74.20	73.13	63.86
QWK	79.64	85.29	85.91	90.73	87.92	80.81	74.31
AUC	89.74	86.00	85.18	90.85	88.74	90.83	76.11

TABLE 3: Comprehensive Performance Comparison of Optimization Techniques. QWK/AUC shown as scale 0-100, Memory in MB, Inference Time in ms/img, Throughput in FPS.

Metric	FP32 (Baseline)	FP16	INT8
QWK	80.42	80.18	80.70
F1 Weighted (%)	60.81	61.06	60.98
AUC (Macro-OvR)	87.33	87.51	87.69
Memory (MB)	668.11	440.28	266.75
Inference (ms/img)	111.26	112.97	115.28
Throughput (FPS)	8.99	8.85	8.67
Timing StdDev (ms)	227.28	203.81	241.67
Speed-up vs FP32	1.00x	0.98x	0.97x
QWK Change (%)	0.00	-0.30	+0.35
Memory Reduction (%)	0.00	34.10	60.07

expensive GPU hardware—a vital attribute for addressing the global burden of DR screening. In summary, RetinaSys represents a significant advancement toward a practical, generalizable, interpretable, and efficient AI system for DR detection, demonstrating considerable promise for enhancing DR screening programs worldwide.

VII. LIMITATIONS AND FUTURE RESEARCH

Despite its strengths, this study faced notable limitations that warrant further investigation. A primary bottleneck was the computational constraint imposed by the use of a single NVIDIA A100 GPU, which limited pre-training to 92 epochs and fine-tuning to 95 epochs. In comparison, models like DeepDR utilized extended training durations, such as 800 epochs for MoCo pre-training, potentially leading to more robust feature representations and improved performance metrics. Future research should explore training with greater computational resources to assess whether extended durations can further enhance the model's accuracy and generalizability, particularly for early-stage DR detection where sensitivity remains a challenge. Another critical limitation is the lack of real-world clinical validation. While RetinaSys demonstrated strong performance on diverse public datasets and unseen dataset portions, its practical utility in clinical settings remains untested. Prospective clinical validation studies comparing RetinaSys directly against expert ophthalmologists in real-world screening scenarios are essential to confirm its impact on patient care pathways. Additionally, such studies should evaluate how clinicians perceive and utilize the xAI explanations provided by RetinaSys, as discussed in Section III-D, to determine whether these visualizations genuinely improve diagnostic

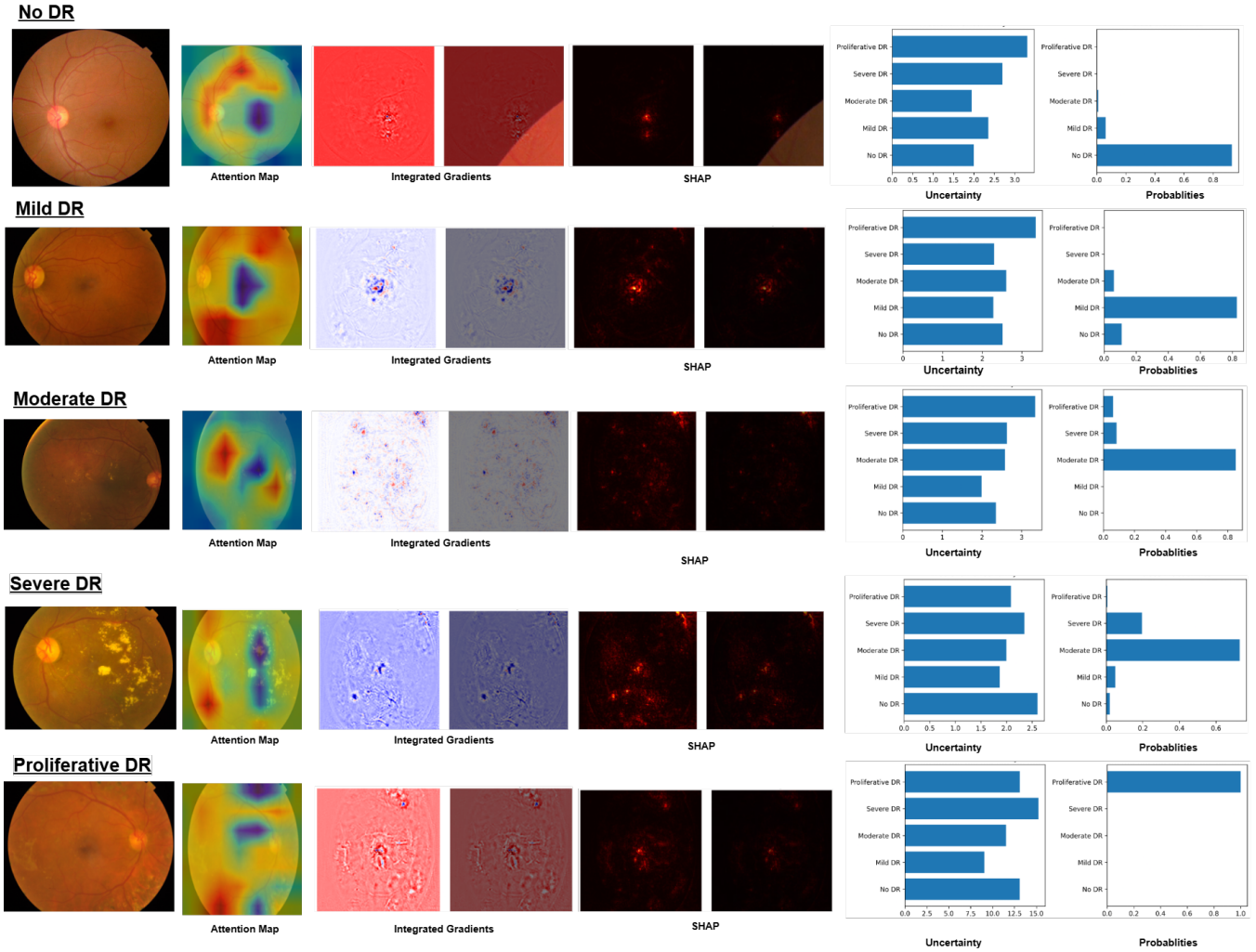


FIGURE 12: Explainable AI (xAI) analysis for representative examples across Diabetic Retinopathy grades (0-4). Methods shown include Original Image, Attention Map, Integrated Gradients, SHAP, and Monte Carlo Dropout uncertainty visualization.

confidence and workflow integration in practice. Addressing these limitations through enhanced computational resources and real-life clinical feedback will be pivotal in refining RetinaSys and ensuring its readiness for widespread clinical adoption.

.

ACKNOWLEDGMENT

The authors extend their appreciation to Taif University, Saudi Arabia, for supporting this work through project number (TU-DSPP-2024-210).

REFERENCES

[1] Z. Teo, Y. Tham, M. Yu, M. Chee, T. Rim, N. Cheung, M. Bikbov, Y. Wang, Y. Tang, Y. Lu, I. Wong, D. Ting, G. Tan, J. Jonas, C. Sabanayagam, T. Wong, and C. Cheng, "Global prevalence of diabetic retinopathy and projection of burden through 2045: Systematic review and meta-analysis," Ophthalmology, vol. 128, no. 11, pp. 1580–1591, 2021.

[2] M. Singh, A. Sharma, and S. Tailor, "A review of diabetic retinopathy disease prediction using deep learning techniques," International Journal of Advanced Research in Computer and Communication Engineering, vol. 14, no. 3, pp. 313–315, mar 2025.

[3] M. Kalavar, H. Al-Khersan, J. Sridhar, R. Gorniak, P. Lakhani, A. Flanders, and A. Kuriyan, "Applications of artificial intelligence for the detection, management, and treatment of diabetic retinopathy," Int. Ophthalmol. Clin., vol. 60, no. 4, pp. 127–145, 2020.

[4] van der Heijden AA, A. MD, V. F, van Hecke MV, L. A, and N. G, "Validation of automated screening for referable diabetic retinopathy with the idx-dr device in the hoorn diabetes care system," Acta Ophthalmol, vol. 96, no. 1, pp. 63–68, 2018.

[5] E. Anand, Rajesh, O. Davidson, C. Lee, and A. Lee, "Artificial intelligence and diabetic retinopathy: Ai framework, prospective studies, head-to-head validation, and cost-effectiveness," Diabetes Care, vol. 46, no. 10, pp. 1728–1739, 2023.

[6] A. Asia, C. Zhu, S. Althubiti, D. Al-Alimi, Y. Xiao, P. Ouyang, and M. Al-Qaness, "Detection of diabetic retinopathy in retinal fundus images using cnn classification models," Electronics, vol. 11, no. 17, p. 2740, 2022.

[7] H. Vasireddi, K. Devi, and N. Goluguri, "Dr-xai: Explainable deep learning model for accurate diabetic retinopathy severity assessment," Arabian J. Sci. Eng., vol. 49, 2024.

[8] X. Xu, M. Zhang, S. Huang, X. Li, X. Kui, and J. Liu, "The application of artificial intelligence in diabetic retinopathy: progress and prospects," Front. Cell Dev. Biol., vol. 12, p. 1473176, 2024.

[9] L. Alsadoun, H. Ali, M. Mushtaq, M. Mushtaq, M. Burhanuddin, R. Anwar, M. Liaqat, S. Bokhari, A. Hasan, and F. Ahmed, "Artificial intelligence

(ai)-enhanced detection of diabetic retinopathy from fundus images: The current landscape and future directions," Cureus, vol. 16, no. 8, p. e67844, 2024.

[10] N. Haq, T. Waheed, K. Ishaq et al., "Computationally efficient deep learning models for diabetic retinopathy detection: a systematic literature review," Artif. Intell. Rev., vol. 57, p. 309, 2024.

[11] N. Tsiknakis, D. Theodoropoulos, G. Manikis, E. Ktistakis, O. Boutsora, A. Berto, F. Scarpa, A. Scarpa, D. Fotiadis, and K. Marias, "Deep learning for diabetic retinopathy detection and classification based on fundus images: A review," Comput. Biol. Med., vol. 135, p. 104599, 2021.

[12] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "Cbam: Convolutional block attention module," in Proceedings of the European Conference on Computer Vision (ECCV), September 2018.

[13] Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, "A convnet for the 2020s," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 11 976–11 986.

[14] Q. Zheng, Z. Wang, J. Zhou, and J. Lu, "Shap-cam: Visual explanations for convolutional neural networks based on shapley value," in European Conference on Computer Vision. Springer, 2022, pp. 743–759.

[15] N. Tsiknakis et al., "Deep learning for diabetic retinopathy detection and classification based on fundus images: A review," Comput. Biol. Med., vol. 135, p. 104599, 2021.

[16] A. Jabbar, S. Naseem, J. Li et al., "Deep transfer learning-based automated diabetic retinopathy detection using retinal fundus images in remote areas," Int. J. Comput. Intell. Syst., vol. 17, p. 135, 2024.

[17] A. Asia, C. Zhu, S. Althubiti, D. Al-Alimi, Y. Xiao, P. Ouyang, and M. Al-Qaness, "Detection of diabetic retinopathy in retinal fundus images using cnn classification models," Electronics, vol. 11, no. 17, p. 2740, 2022.

[18] B. U and B. G, "Deep learning for the detection and classification of diabetic retinopathy with an improved activation function," Healthcare (Basel), vol. 11, no. 1, p. 97, 2022.

[19] L. Dai, L. Wu, H. Li et al., "A deep learning system for detecting diabetic retinopathy across the disease spectrum," Nat. Commun., vol. 12, p. 3242, 2021.

[20] J. Wang, C. Lan, C. Liu, Y. Ouyang, T. Qin, W. Lu, Y. Chen, W. Zeng, and P. S. Yu, "Generalizing to unseen domains: A survey on domain generalization," IEEE Transactions on Knowledge and Data Engineering, vol. 35, no. 8, pp. 8052–8072, 2023.

[21] S. Wang, L. Yu, K. Li, X. Yang, C.-W. Fu, and P.-A. Heng, "Dofe: Domain-oriented feature embedding for generalizable fundus image segmentation on unseen datasets," IEEE Transactions on Medical Imaging, vol. 39, no. 12, pp. 4237–4248, 2020.

[22] L. Zhang, X. Wang, D. Yang, T. Sanford, S. Harmon, B. Turkbey, B. J. Wood, H. Roth, A. Myronenko, D. Xu, and Z. Xu, "Generalizing deep learning for medical image segmentation to unseen domains via deep stacked transformation," IEEE Transactions on Medical Imaging, vol. 39, no. 7, pp. 2531–2540, 2020.

[23] Z. Shen, H. Fu, J. Shen, and L. Shao, "Modeling and enhancing low-quality retinal fundus images," IEEE Transactions on Medical Imaging, vol. 40, no. 3, pp. 996–1006, 2021.

[24] X. Wang, M. Xu, J. Zhang, L. Jiang, and L. Li, "Deep multi-task learning for diabetic retinopathy grading in fundus images," Proceedings of the AAAI Conference on Artificial Intelligence, vol. 35, no. 4, pp. 2826–2834, May 2021. [Online]. Available: https://ojs.aaai.org/index.php/AAAI/article/view/16388

[25] H. Liu, H. Li, X. Wang, H. Li, M. Ou, L. Hao, Y. Hu, and J. Liu, "Understanding how fundus image quality degradation affects cnn-based diagnosis," in 2022 44th Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC), 2022, pp. 438–442.

[26] J. Hou, S. Liu, Y. Bie, H. Wang, A. Tan, L. Luo, and H. Chen, "Self-explainable ai for medical image analysis: A survey and new outlooks," arXiv preprint arXiv:2410.02331, 2024. [Online]. Available: https://arxiv.org/abs/2410.02331

[27] B. H. M. van der Velden, H. J. Kuijf, K. G. A. Gilhuijs, and M. A. Viergever, "Explainable artificial intelligence (xai) in deep learning-based medical image analysis," Medical Image Analysis, vol. 79, p. 102470, Jul 2022. [Online]. Available: https://doi.org/10.1016/j.media.2022.102470

[28] V. Tulsani, P. Sahatiya, J. Parmar, and J. Parmar, "Xai applications in medical imaging: A survey of methods and challenges," International Journal on Recent and Innovation Trends in Computing and Communication, vol. 11, no. 9, pp. 181–186, Oct 2023. [Online]. Available: https://doi.org/10.17762/ijritcc.v11i9.8332

[29] D. Muhammad and M. Bendechache, "Unveiling the black box: A systematic review of explainable artificial intelligence in medical image analysis," Computational and Structural Biotechnology Journal, vol. 24, pp. 542–560, Aug 2024. [Online]. Available: https://doi.org/10.1016/j.csbj.2024.08.005

[30] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," in Proceedings of the IEEE international conference on computer vision, 2017, pp. 618–626.

[31] S. Matta, M. Lamard, P. Conze et al., "Towards population-independent, multi-disease detection in fundus photographs," Sci. Rep., vol. 13, p. 11493, 2023.

[32] C. K. Wong, M. Ngo, M. Lin, Z. Bashir, A. Heen, M. B. S. Svendsen, M. G. Tolsgaard, A. N. Christensen, and A. Feragen, "Deployment of deep learning model in real world clinical setting: A case study in obstetric ultrasound," arXiv preprint arXiv:2404.00032, 2024. [Online]. Available: https://arxiv.org/abs/2404.00032

[33] A. Bastos de Carvalho, S. Lee Ware, T. Belcher, F. Mehmeti, E. B. Higgins, R. Sprang, C. Williams, J. L. Studts, and C. R. Studts, "Evaluation of multi-level barriers and facilitators in a large diabetic retinopathy screening program in federally qualified health centers: a qualitative study," Implement Sci Commun, vol. 2, no. 1, p. 54, May 2021.

[34] Z. Gu, Y. Li, Z. Wang, J. Kan, J. Shu, and Q. Wang, "Classification of diabetic retinopathy severity in fundus images using the vision transformer and residual attention," Computational Intelligence and Neuroscience, vol. 2023, p. 1305583, 2023.

[35] J.-H. Wu, N. D. Koseoglu, C. Jones, and T. Y. A. Liu, "Vision transformers: The next frontier for deep learning-based ophthalmic image analysis," Saudi Journal of Ophthalmology, vol. 37, no. 3, pp. 173–178, Jul–Sep 2023.

[36] T. Karkera, C. Adak, S. Chattopadhyay, and M. Saqib, "Detecting severity of diabetic retinopathy from fundus images: A transformer network-based review," arXiv preprint arXiv:2301.00973, 2023, journal reference: Neurocomputing, Elsevier, 2024.

[37] Z. Yao, Y. Yuan, Z. Shi, W. Mao, G. Zhu, G. Zhang, and Z. Wang, "Funswin: A deep learning method to analysis diabetic retinopathy grade and macular edema risk based on fundus images," Front Physiol, vol. 13, p. 961386, Jul 2022.

[38] S. Saha and L. Xu, "Vision transformers on the edge: A comprehensive survey of model compression and acceleration strategies," arXiv preprint arXiv:2503.02891, 2025.

[39] T. C. Nauen, S. Palacio, and A. Dengel, "Which transformer to favor: A comparative analysis of efficiency in vision transformers," arXiv preprint arXiv:2308.09372, 2023.

[40] J. Huang, R. Channa, R. Wolf, and et al., "Autonomous artificial intelligence for diabetic eye disease increases access and health equity in underserved populations." npj Digital Medicine, vol. 7, p. 196, 2024. [Online]. Available: https://doi.org/10.1038/s41746-024-01197-3

[41] L. Dai, B. Sheng, T. Chen et al., "A deep learning system for predicting time to progression of diabetic retinopathy," Nat. Med., vol. 30, pp. 584–594, 2024.

[42] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple framework for contrastive learning of visual representations," in International Conference on Machine Learning, 2020, pp. 1597–1607.

[43] J. Grill, F. Strub, F. Altché, C. Tallec, P. Richemond, O. Bachem et al., "Bootstrap your own latent: A new approach to self-supervised learning," Adv. Neural Inf. Process. Syst., vol. 33, pp. 21 271–21 284, 2020.

[44] K. He, H. Fan, Y. Wu, S. Xie, and R. Girshick, "Momentum contrast for unsupervised visual representation learning," arXiv, 2019.

[45] M. Caron et al., "Emerging properties in self-supervised vision transformers," arXiv, 2021.

[46] M. Assran et al., "Self-supervised learning from images with a joint-embedding predictive architecture," arXiv, 2023.

[47] C. Zhang and Y. Gu, "Dive into self-supervised learning for medical image analysis: Data, models and tasks," arXiv, 2022.

[48] J. Anton, L. Castelli, M. Chan, M. Outters, W. Tang, V. Cheung, P. Shukla, R. Walambe, and K. Kotecha, "How well do self-supervised models transfer to medical imaging?" J. Imaging, vol. 8, no. 12, p. 320, 2022.

[49] D. Wolf, T. Payer, C. Lisson et al., "Self-supervised pre-training with contrastive and masked autoencoder methods for dealing with small datasets in deep learning for medical imaging," Sci. Rep., vol. 13, p. 20260, 2023.

[50] S. Tayebi Arasteh, L. Misera, J. Kather, D. Truhn, and S. Nebelung, "Enhancing diagnostic deep learning via self-supervised pretraining on large-scale, unlabeled non-medical images," Eur. Radiol. Exp., vol. 8, no. 1, p. 10, 2024.

[51] M. Nielsen, L. Wenderoth, T. Sentker, and R. Werner, "Self-supervision for medical image classification: State-of-the-art performance with 100 labeled training samples per class," Bioengineering, vol. 10, p. 895, 2023.
[52] V. V. Zunin, "Intel openvino toolkit for computer vision: Object detection and semantic segmentation," in 2021 International Russian Automation Conference (RusAutoCon), 2021, pp. 847–851.
[53] J. Cuadros and G. Bresnick, "EyePACS: an adaptable telemedicine system for diabetic retinopathy screening," Journal of Diabetes Science and Technology, vol. 3, no. 3, pp. 509–516, May 2009.
[54] T. Li, Y. Gao, K. Wang, S. Guo, H. Liu, and H. Kang, "Diagnostic assessment of deep learning algorithms for diabetic retinopathy screening," Information Sciences, vol. 501, pp. 511 – 522, 2019. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0020025519305377
[55] "Aptos 2019 diabetic retinopathy dataset." [Online]. Available: https://www.kaggle.com/c/aptos2019-blindness-detection/data
[56] P. Porwal, S. Pachade, R. Kamble, M. Kokare, G. Deshmukh, V. Sahasrabuddhe, and F. Meriaudeau, "Indian diabetic retinopathy image dataset (idrid)," 2018. [Online]. Available: https://dx.doi.org/10.21227/H25W98
[57] E. Decencière et al., "Feedback on a publicly distributed database: the messidor database," Image Analysis Stereology, vol. 33, no. 3, pp. 231–234, aug 2014. [Online]. Available: http://www.ias-iss.org/ojs/IAS/article/view/1155
[58] L. Deng, J. Lyu, H. Huang, J. Yuan, Z. Wang, M. Li, L. Fang, D. Zhang, L. Lin, and X. Tang, "The sustech-sysu dataset for automatically segmenting and classifying corneal ulcers," Scientific Data, vol. 7, no. 1, p. 23, 2020.
[59] M. Atwany and M. Yaqub, "Drgen: Domain generalization in diabetic retinopathy classification," in Medical Image Computing and Computer Assisted Intervention – MICCAI 2022, L. Wang, Q. Dou, P. T. Fletcher, S. Speidel, and S. Li, Eds. Cham: Springer Nature Switzerland, 2022, pp. 635–644.
[60] H. Zhang, M. Cisse, Y. N. Dauphin, and D. Lopez-Paz, "mixup: Beyond empirical risk minimization," in International Conference on Learning Representations, 2018. [Online]. Available: https://doi.org/10.48550/arXiv.1710.09412
[61] K. Zhou, Y. Yang, Y. Qiao, and T. Xiang, "Domain generalization with mixstyle," in International Conference on Learning Representations, 2021. [Online]. Available: https://doi.org/10.48550/arXiv.2104.02008
[62] S. Liu, L. Gong, K. Ma, and Y. Zheng, "Green: a graph residual re-ranking network for grading diabetic retinopathy," in Medical Image Computing and Computer Assisted Intervention – MICCAI 2020, A. L. Martel, P. Abolmaesumi, D. Stoyanov, D. Mateus, M. A. Zuluaga, S. K. Zhou, D. Racoceanu, and L. Joskowicz, Eds. Cham: Springer International Publishing, 2020, pp. 585–594.
[63] A. He, T. Li, N. Li, K. Wang, and H. Fu, "Cabnet: Category attention block for imbalanced diabetic retinopathy grading," IEEE Transactions on Medical Imaging, vol. 40, no. 1, pp. 143–153, 2021.
[64] K. Zhou, Y. Yang, T. Hospedales, and T. Xiang, "Deep domain-adversarial image generation for domain generalisation," Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, pp. 13 025–13 032, 04 2020.
[65] F.-E. Yang, Y.-C. Cheng, Z.-Y. Shiau, and Y.-C. F. Wang, "Adversarial teacher-student representation learning for domain generalization," in Proceedings of the 35th International Conference on Neural Information Processing Systems, ser. NIPS '21. Red Hook, NY, USA: Curran Associates Inc., 2021.
[66] A. Rame, C. Dancette, and M. Cord, "Fishr: Invariant gradient variances for out-of-distribution generalization," in International Conference on Machine Learning, ser. Proceedings of Machine Learning Research, vol. 162. PMLR, 2022, pp. 18 309–18 330. [Online]. Available: https://proceedings.mlr.press/v162/rame22a.html
[67] Y. Yang, H. Wang, and D. Katabi, "On multi-domain long-tailed recognition, imbalanced domain generalization and beyond," in European Conference on Computer Vision, 2022, pp. 236–253. [Online]. Available: https://doi.org/10.48550/arXiv.2203.09513
[68] H. Che, Y. Cheng, H. Jin, and H. Chen, "Towards generalizable diabetic retinopathy grading in unseen domains," in Medical Image Computing and Computer Assisted Intervention – MICCAI 2023, ser. Lecture Notes in Computer Science, vol. 14226. Springer, 2023, pp. 200–210.

AMAN SWAR (Student Member, IEEE) received the B.Tech. degree in Computer Science and Engineering from SRM Institute of Science and Technology, Kattankulathur, Chennai, India, where he is currently pursuing further studies/research. His research interests include deep learning, computer vision, medical image analysis, self-supervised learning, and efficient AI deployment.

k_vijayalakshmi.png

K. VIJAYALAKSHMI K. Vijayalakshmi is working as Assistant Professor in the Department of Computational Intelligence at S.R.M.Institute of Science and Technology, Kattankulathur, Chennai. She acquired B.E. Degree in Electronics and Communication Engineering from Valliammai Engineering College, Kattankulathur in 2005. She received M.Tech Degree in Information Technology and doing Ph.D degree in the Faculty of Information and communication engineering from Sathyabama University, Chennai. She has over 16 years of experience in teaching and guiding projects for Undergraduate and post graduate students. She has published research papers in various SCI, Scopus Indexed Journals, international journals and conferences. Her research areas include Machine Learning, Deep Learning, and Internet of Things.

a_maheshwari.png

A. MAHESHWARI Maheshwari A obtained her BE (CSE) from Madras University in 1999, ME (CSE) from Sathyabama University, Chennai, in 2008 and she received her Ph. D. in the area of Machine Learning from Sathyabama University, Chennai, in 2019. She is having 22 years of experience in teaching. She has published research papers in various SCI, Scopus Indexed Journals, international journals and conferences. Her research areas include Machine Learning, Machine Transliteration for Brahmi Letters, Deep Learning, and Internet of Things.

k_saravanan.png

K. SARAVANAN (Senior Member, IEEE) received his B.E. degree in Electrical and Electronics Engineering from Alagappa Chettiar College of Engineering and Technology, Madurai Kamaraj University, Karaikudi, Tamilnadu, India in 1998. He received his M.E. degree in Power Electronics and Drives from Anna University, Guindy Campus, Chennai, Tamilnadu, India in 2001. He completed his Ph.D. degree in JNTUK-Kakinada, AP, India in 2018. At present, he holds the post of Associate Professor in EEE at SRM IST, Kattankulathur, Chennai, India. He has published several technical papers in national and international journals. His current research interests include Renewable Energy Systems, Electric Vehicles, and multilevel inverters. He is a Fellow member in FIE, a Life Member in ISTE, and a Life member in SESI.

G. SUMATHY Dr. Sumathy G Working as Assistant Professor in the Department of Computational Intelligence, at SRM IST, Kattankulathur 603 203, chennai, Tamilnadu, India. Sumathy G obtained her BE (CSE) from Bharathidasan University in 2004, M.Tech (SW Engg) from Bharathidasan University, Trichy, in 2007 and she received her Ph. D. in the area of Machine Learning from Sathyabama University, Chennai, in 2020. She is having 15 years of experience in teaching. She has published research papers in various SCI, Scopus Indexed Journals, international journals and conferences. She has participated several Workshops, Seminars and Conferences. Her research interests include Machine Learning, Image processing, Vision computing, Embedded system and Data Mining.

MAJED ALSAFYANI Majed Alsafyani received the bachelor's degree (Hons.) in computer science from the University of Hertfordshire (UH), U.K., in 2013, and the master's degree (Merit) in Advance computer science from the University of Hertfordshire (UH), U.K., in 2014, and the Ph.D. degree in computer science from the University of Hertfordshire, U.K., in 2015 and 2020, respectively. He is currently an Assistant Professor with the College of Computers and Information Technology, Taif University, Saudi Arabia. His research interests on image processing, machine learning, applications of Internet of Things, artificial intelligence and software engineering.

SAEED RUBAIEE Saeed Rubaiee received the B.S. degree in the Chemical Engineering from Tennessee Tech University, Cookeville, USA, the M.Sc. degree in Industrial Engineering (Systems) from University of South Florida, Florida, USA, and the PhD degree from the Department of Industrial and Manufacturing Engineering, Wichita State University, Wichita. He is currently an Associate Professor in the department of the industrial and systems engineering, University of Jeddah, Jeddah, Saudi Arabia. His research interest includes engineering systems, sustainability and green manufacturing, production equipment operation, renewable energy, manufacturing engineering, material engineering, advanced materials, applied optimization.

AMR YOUSEF AMR YOUSEF is an assistant professor with the Electrical Engineering Department at the University of Business and Technology, KSA. He was a post-doctoral research associate at Old Dominion Vision Lab, USA. He obtained his Ph.D. degree in Electrical and Computer Engineering from Old Dominion University (ODU) in May 2012 and MSc and BSc. degrees from The Engineering Mathematics Department and The Electrical Engineering Department at Alexandria University in 2001 and 2006 respectively. His research is in optimization techniques, image processing/computer vision and machine learning algorithms. He is a member of SPIE, OSA and IEEE.

NARAYANAMOORTHI R NARAYANAMOORTHI R received the bachelor's degree in electrical engineering and the master's degree in control and instrumentation from Anna University, India, in 2009 and 2011, respectively, and the Ph.D. degree from the SRM Institute of Science and Technology, India, in 2019. He is currently and working as an Associate Professor at the Department of Electrical and Electronics Engineering, SRM Institute of Science and Technology. His research interests include Wireless Power Transfer, Electric Vehicle, Power Electronics, Artificial Intelligence, Machine learning in Renewable energy systems and embedded system for smart sensors.