Bulletin of Electrical Engineering and Informatics
Vol. 15, No. 4, August 2026, pp. 3095~3108
ISSN: 2302-9285, DOI: 10.11591/eei.v15i4.10407

Detection of stages in diabetic retinopathy using computer aided ensemble network

Rayudu Prasanthi[1,2], Rajyalakshmi Uppada[2], Leela Kumari Balivada[1]
[1]Department of Electronics and Communications Engineering, University College of Engineering Kakinada (UCEK), Jawaharlal Nehru Technological University Kakinada (JNTUK), Kakinada, India
[2]Department of Electronics and Communications Engineering, School of Engineering, Aditya University, Surampalem, India

Article Info

Article history:

Received Apr 6, 2025
Revised May 30, 2026
Accepted Jun 19, 2026

Keywords:

Analysis of medical images
Automated grading
Convolutional neural network
Diabetic retinopathy
Ensemble deep learning
Retinal fundus images

ABSTRACT

Diabetic retinopathy (DR) is one of the progressive micro vascular disorders of diabetes and a major cause of preventable blindness in the world. Manual ophthalmologist evaluation is costly in terms of time and more likely to have inter-observer error, whereas current automated methods tend to fail to differentiate between intermediate stages of DR; yet, proper assessment of DR severity is critical to its successful intervention. This paper suggests a framework of hybrid ensemble deep learning (DL) model that incorporates VGG16, InceptionV3 and ResNet50 based on a weighted feature fusion model and a feature-scaled parametric activation (FSPA) model to maximize inter-classes separability. The Kaggle EyePACS dataset containing 35,126 retinal fundus images was used. Image normalization, contrast enhancement, and data augmentation improved robustness to class imbalance. The overall accuracy of the proposed ensemble was 95.0% and the sensitivity and specificity were 92.0 and 94.0 respectively and the quadratic weighted Kappa (QWK) was 0.91, with up to +5% improvements in accuracy over individual convolutional neural network (CNN) baselines. Although there are still limitations to differentiating moderate versus severe DR, the findings demonstrate that the proposed framework enables consistent and reliable DR grading for large-scale clinical screening.

Corresponding Author:

Rayudu Prasanthi
Department of Electronics and Communications Engineering
University College of Engineering Kakinada (UCEK)
Jawaharlal Nehru Technological University Kakinada (JNTUK)
Kakinada, Andhra Pradesh, 533003, India
Email: santhi.476@gmail.com

1. INTRODUCTION

One of the most serious complications in diabetes mellitus both in microvascular or other aspects is diabetic retinopathy (DR) as one of the major causes of blindness that is still preventable among the working population in any nation [1]. The high rate of diabetes prevalence has largely contributed to the high risks of vision impairment due to DR. The estimates indicate that the world will have more than 640 million people with diabetes as at 2030 and almost 1/3 of patients with diabetes will have had some kind of DR at some point in their lives at least according to the global epidemiological estimation [2]. DR advances slowly into non-proliferative mild abnormalities developing progressively into proliferative stages, which include neovascularization and severe retinal damage. Notably, in many cases, the early occurrence of the disease remains asymptomatic, and proper diagnosis and the accurate severity score are critical in preventing the untreatability of irreversible vision loss with proper clinical intervention [3].

Traditionally, the diagnosis and grading of DR are conducted by hand using retinal fundus images by trained ophthalmologists, using standardized guidelines, including the early treatment DR study (ETDRS) scale. This method is based on the detection of typical lesions, such as microaneurysms, hemorrhages, exudates, and neovascularization [4]. Although manual grading is the clinical gold standard, it is labor intensive, time consuming and subject to inter and intra-observer variability especially in the identification of adjacent stages of severity like mild, moderate and severe non-proliferative DR [5]. These constraints make mass screening difficult, particularly in areas where there is a shortage of ophthalmic skills, which is why automated computer-aided diagnostic systems are under development [6].

The recent developments in deep learning (DL), especially convolutional neural networks (CNNs), have drastically enhanced automated retinal image processing through the ability to extract features in a hierarchical manner directly out of raw images. CNN-based models have proven to be very effective in medical image classification tasks, with researchers like Khatun and Hossain showing performance similar to expert ophthalmologists in DR detection [7]. DR detection, lesion segmentation, and severity grading have been extensively performed using different architectures such as VGGNet, Inception, ResNet, DenseNet, and U-Net [8]–[10]. Most methods, however, are mostly binary or coarse-grade.

Multi-class DR grading is also a difficult task because of the small inter-class differences, class imbalance in data, and generalization to a wide range of imaging conditions [11]–[13]. Such problems tend to lead to diminished model performance in clinical practice. Ensemble learning is one such promising solution to these challenges, which combines several models to improve robustness and classification results [14]. However, the current ensemble techniques are mostly based on basic feature averaging or decision-level fusion and might not be effective enough to identify discriminative features between adjacent DR stages.

This paper will address these shortcomings by suggesting a hybrid ensemble DL system to multi-stage DR grading. The model combines VGG16, InceptionV3 and ResNet50 using a weighted feature fusion approach to leverage complementary hierarchical representations. Also, a feature-scaled parametric activation (FSPA) process is proposed to boost lesion-specific responses and inhibit irrelevant activations. The suggested framework is trained and tested on the large-scale EyePACS dataset (35,126 retinal images) with five levels of DR severity. The clinically relevant measures of performance, such as accuracy, sensitivity, specificity, and quadratic weighted Kappa (QWK) to determine agreement with expert ophthalmologist grading, are used to evaluate performance.

The contributions of the major contributions of this work are summed up as:

- Multi-stage classification of DR using complementary feature representations by creation of weighted ensemble CNN architecture composed of VGG16, InceptionV3, and ResNet50.
- Introduction of parametric scale-of-feature activation mechanism that was to enhance inter-class separability, in the circumstances of visual similar DR scale.
- Extensive testing over a large public dataset with a large skewed distribution of samples between classes on clinically relevant performance measures.
- Evidence of the model as a clinical decision-support system to support ophthalmologists in larger DR screening systems.

The rest of this paper will be structured in the following way. Section 2 is the literature review related to DL-based DR detection and grades. Section 3 presents the proposed methodology, which consists of dataset preparation, preprocessing, model architecture and training strategy. Section 4 displays results and comparative analysis of the experiment and section 5 is the conclusion of the paper containing information on the future research directions.

2. LITERATURE SURVEY

DR is a significant issue that leads to poor vision among diabetics in the world and to avoid irreversible blindness, it is crucial to diagnose it at an early stage. Automated screening systems have emerged as a promising field of research in medical image analyzing especially with the rise in the number of diabetic patients. During the last 20 years, methods of DR detection have changed their antique hand-made feature-oriented methods to recent developments of DL and explainable artificial intelligence (XAI).

2.1. Diabetic retinopathy detection using deep learning

Recently, DL and specifically CNNs have contributed immensely to automated diagnosing of DR. Gulshan *et al.* [1] created a DL algorithm based on huge volumes of retinal fundus images and showed that the system was able to match the diagnostic accuracy of top skilled ophthalmologists. Pratt *et al.* [2] also explored CNN models at DR detection, demonstrated that DL models have the ability to extract discriminative retinal features automatically without the need of manual feature engineering. On the same note, Lam *et al.* [3] came with an automated detection framework of DR that applies DL methods to classify retinal images end-to-end.

Different CNN architectures have been developed in an attempt to improve the power of feature extraction. Huang *et al.* [4] suggested DenseNet architecture which enhances gradient propagation and the use of features to conserve space in the connections between the layers. One of the architectures that have been extensively used in the retinal vessel and lesion segmentation work is the U-Net architecture presented by Ronneberger *et al.* [5] to perform biomedical image segmentation. Moreover, Li *et al.* [6] also illustrated the usefulness of CNNs in the morphing of medical images and revealed considerable better performance compared to the conventional machine learning (ML) frameworks.

The DL-based DR detection progress has also been summed up through the survey studies. Shekar *et al.* [15] made a review of the CNN-based approaches to DR classification and emphasized the benefits of the reviewed methods in automated feature extraction. In detail, Khatun and Hossain [7] conducted a survey of the existing techniques of early DR detection and provided their views on the perspectives of research in the field of automated screening systems.

2.2. Transfer learning and pretrained models

Transfer learning has gained widespread popularity in DR detection to seek solutions to the problem of low-labeled medical datasets. Comparative study the authors performed a comparative analysis of multiple pretrained deep neural networks (DNN) in the classification of DR [8]. They discovered that enhanced accuracy in the transfer learning methodology is improved. Akhtar *et al.* [9] developed a DL based grading model that can identify the various stages of DR with better performance.

More studies have been conducted on computer-aided diagnosis of the severity of DR. Khanna *et al.* [10] construct the system which predicts the severity levels of DR on the basis of retinal fundus pictures using DL. Very large publicly available datasets like EyePACS dataset [11] have been significant in the research of DR detection by providing large scale annotated image of the retina to be used in training and testing of models. The study by Sebastian *et al.* [12] offered a thorough review of the lesion detection and segmentation methods of DR applied to DL methods.

2.3. Explainable artificial intelligence diabetic retinopathy diagnosis

Even though the DL models have been quite accurate, the lack of interpretability and black-box nature restricts clinical trust and interpretation. Torre *et al.* [13] suggested an interpretable DL classifier that is used in DR grading, which enhances the transparency of an automated diagnosis system. Tsao *et al.* [14] also investigated interpretable ML algorithms with the view to determining clinically relevant biomedical factors in relation to DR.

Clear artificial intelligent methods have been used to represent model choices graphically. Jiang *et al.* [16] used the Grad-CAM visualization to emphasize significant retinal areas used in making classification decisions within the CNN-based DR detection systems. To enhance the transparency of a model, Niu *et al.* [17] developed an explainable framework of DL that combines the image generation model with classification of retinal images. Another article by Jiang *et al.* [18] explored the eye-tracking DL analysis to identify DR at a very early stage by comparing the model predictions and clinician attention patterns.

2.4. Deep learning models and emerging

Recent papers have discussed state-of-the-art DL based models with multimodal reasoning and enhanced interpretability. Ito *et al.* [19] came up with an explainable vision-language model capable of diagnosing DR that combines both visual and textual features in a bid to offer clinical decision support. Hesse and Namburete [20] came up with the INSightR-Net, which can improve the interpretability through the use of prototypical examples, and the evaluation of neural networks forecasts.

XAI methods have also been examined to be integrated in DR detection systems to enhance clinical trust. Alavee *et al.* [21] created a DL model that combines both DR detection and XAI. Lalithadevi *et al.* [22] present a feasibility study that proves XAI-based systems may positively influence clinician confidence in automated DR diagnosis. One of the big problems associated with the evaluation of DR is variability in its expert grading. Krause *et al.* [23] analyzed the problem of grader discrepancies in retinal image identification and highlighted the relevance of sound reference standards in the assessment of ML models in DR identification.

Ting *et al.* [24] introduced a clinically validated AI system to identify DR on multiethnic retinal images, which suggests the possibility of retinopathy ophthalmic screening with the help of AI. Likewise, Quellec *et al.* [25] came up with a deep image mining framework to detect pathological retinal appearances linked with DR and showed that DL is effective in revealing clinically significant lesions.

The deep convolutional neural network (DCNN) has advanced to enhance performance in classifying medical images. The authors included He *et al.* [26] proposed the ResNet architecture, or slightly more specifically, led by deep residual connections, which allow solving the vanishing gradient issue, as well as, can be used to train very deep networks. Similarly, VGG16 is a DCNN that uses small convolutional filters to extract features over a hierarchy proposed by Simonyan and Zisserman [27]. Moreover, Szegedy *et al.* [28] presented InceptionV3 that learns multi-scale spatial features via parallel convolutional modules and has served extensively to accomplish image classification tasks.

2.5. Gaps and motivation of the research

Although automated DR detection systems have made great leaps forward, there are still a number of limitations:
– Weak interpretability of DL models.
– Advanced architectures have a high level of computation.
– Relying on extensive labeled medical data.
– Lack of fine detail with real-world imaging conditions.

Such obstacles encourage the creation of more powerful and understandable diagnosis paradigms. Hence, the suggested work is aimed at completing the ensemble DL models with XAI methods to enhance the accuracy of classification, enhancing reliability, and clinical usefulness of automated DR detection.

3. METHOD

The proposed automated DR classification system is made up of multiple CNN as an aggregation and weighted ensemble. The sequence of operations used in the methodology consists of five major parts namely;
– Description of dataset
– Image preprocessing
– Model architecture design
– Training and optimization
– Evaluation

Figure 1 shows the detail of workflow of the proposed system. DL methods, especially CNN-based models have performed strikingly in the task of automated DR detection [1], [15]. The input information is the retinal fundus images accessed to publicly available EyePACS dataset used previously to conduct numerous studies involving DR detection [1], [11]. These pictures go through a series of preprocessing, such as image size, contrast, and augmentation, to enhance the visual, as well as image generalization [12]. The resulting processed images are then inputted into three well-established CNN architectures including ResNet50 [26], VGG16 [27], and InceptionV3 [28] that are able to be trained to learn hierarchical feature representations of retinal images.

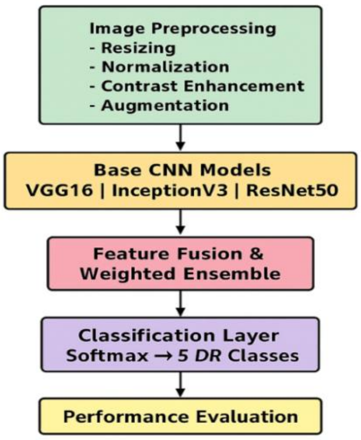

Figure 1. Hybrid ensemble framework proposal to DR classification

A weighted ensemble approach combines the features extracted out of these models such that they can combine complementary information of various architectures [25]. The combined attributes are then forwarded to a SoftMax classification layer which classifies the images into five levels of the severity of DR. Lastly, it is the performance that is used to evaluate the proposed framework based on the several performance measures including accuracy, sensitivity, specificity, and QWK which have been commonly employed in the study of DR classification [15], [21]. This pipeline is modular and therefore guarantees that there are reproducibility, scalability, and enhanced performance in automated DR grading.

3.1. Dataset description

The publicly available EyePACS DR dataset with 35,126 high-resolution retinal fundus images with five levels of annotations of DR severity were used in the experiments [1], [11].

- Class 0: No DR
- Class 1: Mild
- Class 2: Moderate
- Class 3: Severe
- Class 4: Proliferative DR

The dataset has a high level of class imbalance with most of the images comprising the non-DR category. This is one of the pitfalls of medical image datasets that can affect the results of a classifier biased against the majority classes [12]. To solve this challenge, training done through the application of class reweighting and data augmentation measures was used to enhance the model strength and accuracy.

3.1.1. Dataset splitting strategy

To make the best judgments and make them reproducible, the EyePACS data were split into training, validation, and the test sets with an 80:10:10 ratio by patients so that images of a patient were not represented in more than one subsets. To maintain the original proportions of classes in all the levels of DR severity stratified random sampling was carried out. To ensure that the experiments could be reproducible, a fixed random seed was employed (seed=42). Hyper parameter tuning and model selection were done on a validation set and final performance metrics only on the held-out test set were reported in accordance with standard ML evaluation protocols [15].

3.2. Image preprocessing

The image preprocessing was done so that it standardizes the data input and improves the diagnostically significant features in the retinal pictures. The preprocessing is important in enhancing the performance of CNN when used in medical image analysis [12].

a. Resizing: all the images were rescaled to 224×224 pixels, the size of a typical standard input of the majority of CNN architectures, such as ResNet50, VGG16, and InceptionV3 [26]–[28].
b. Normalization: the values of pixel intensities were brought to the [0, 1] range to maintain constant gradient updates throughout training and minimize the numerical instability in DNN [6].
c. Contrast enhancement: to improve vascular structures and emphasize retinal lesions including microaneurysms and hemorrhages, contrast limited adaptive histogram equalization (CLAHE) pizza tile (8×8 square) with clip limit of 2.0 was used to sharpen images [12]. CLAHE has become popular in the field of retina images pre-processing in order to enhance the visibility and diagnostic detectability of a lesion.
d. Noise removal: a median filter of 3×3 kernel size was used to eliminate the noise artifacts like salt and pepper noise and still maintain the boundaries of the lesions [12].

3.3. Data augmentation

Data augmentation is a widely used technique in DL to improve model generalization by artificially increasing the diversity of training data. In this study, data augmentation was applied to the training images to increase dataset diversity and reduce the impact of class imbalance. The augmentation techniques include:

- Random rotations (+-15 degrees)
- Horizontal and vertical flips
- Zooming applied to balance the class distributions

Mathematically normalized image $I_{norm}(x, y)$ is given as:

$$I_{norm}(x, y) = \frac{I(x,y) - I_{min}}{I_{max} - I_{min}} \tag{1}$$

where $I(x, y)$ is the pixel intensity and I_{min}, I_{max} are the minimum and maximum intensity value for the image.

3.4. Proposed hybrid ensemble model

The suggested scheme uses a hybrid ensemble approach that incorporates CNN models in order to enhance the strength and precision of DR classification. The use of ensemble learning in medical image analysis has been greatly embraced by the fact that when a combination of prediction using multiple models is used, this tends to result to better generalization and low variance mistakes than when using single models [25]. In the paper, three widely used CNN architectures were selected, namely based on their complementary features and the known effectiveness in medical image processing tasks [10], [15]. They are: i) ResNet50 [26], ii) VGG16 [27], and iii) Inception V3 [28].

Even though newer models like EfficientNet and Vision Transformer have shown promising results even in image classification tasks, they generally demand larger datasets as well as extra computing resources to train. Conversely, the sampled architectures give a suitable mix between the performance, reading ability, and computation expenses, and therefore can be utilized in the scalable clinical screening structures.

The main input of this work consists in the cluster combination of the three CNN structures. All of the base models are pre-trained using ImageNet weights, and it assists in speeding up convergence and enhancing feature generalization during training on relatively small medical datasets [6], [10]. The models were then refined on the EyePACS dataset to acquire disease-specific features of retinal conditions that indicate DR.

3.4.1. ResNet50 architecture

ResNet50 structure alleviates the issue of the vanishing gradient in deep networks by using a technique called residual learning and skip connections [26]. These remaining residual connections permit a more efficient propagation of gradients through deep layers and therefore the network can learn hierarchical features of higher complexity. ResNet50 has shown good results in large scale image recognition problems and has been effectively utilized in a number of medical imaging tasks [10].

3.4.2. VGG16 architecture

VGG16 is a 16-layer deep CNN which uses small 3×3 convolutional filters, which allows the extraction of fine-grained local texture structures by images [27]. These local feature representations are especially helpful in the analysis of retinal images in order to identify small pathological signs like microaneurysms and hemorrhages and exudates. VGG16 is a simple and homogenous model, as a result of which it is widely used as a baseline model in medical image classification tasks [15].

3.4.3. InceptionV3 architecture

InceptionV3 model uses Inception modules, which use a set of parallel convolutions of varying kernel sizes to extract multi-scale spatial aspects in input images [28]. This architecture allows an efficient extraction of features at varying resolution and therefore it is specifically applicable to the retinal image where the lesion may differ in form and size.

3.4.4. Feature fusion strategy

$\mathrm{F_{fusion}}$ fused feature representation is found with the help of weighted ensemble strategy:

$$\mathrm{F_{fusion}} = \mathrm{w_1 F_1} + \mathrm{w_2 F_2} + \mathrm{w_3 F_3} \tag{2}$$

where $\mathrm{F_1}, \mathrm{F_{2,}}$ and $\mathrm{F_3}$ represent the feature embeddings of ResNet50 [26], VGG16 [27], and InceptionV3 [28], respectively, and $\mathrm{w_1}, \mathrm{w_2}, \mathrm{w_3}$ are optimized weights that satisfies.

$$\sum_{\mathrm{i=1}}^{3} \mathrm{w_i} = 1 \tag{3}$$

The ensemble weights $(w_{1,} w_{2,} w_3)$ were not fixed manually rather they were automatically learned during training through back propagation. Each backbone network generated a feature embedding then a projection layer, and the weights were optimized together with the classification loss with the constraint $w_i = 1$, imposed through SoftMax normalization on the ensemble weights. The fused output is run on a fully connected layer and a SoftMax classifier is used to forecast DR severity levels of five levels. One of the advantages of such ensemble fusion is that the model can take advantage of the complementary information learned by other architectures and achieve better classification results [25].

3.4.5. Variable parametric activation function

Let the stimulation of a neuron be z. The issue of custom FSPA can be formulated as (4):

$$F_{Scaled} = S.F \tag{4}$$

3.4.6. Feature-scaled parametric activation function

Let the activation of a neuron be z. The custom FSPA can be defined as (5):

$$f_{FSPA}(z_i) = \alpha_i . ReLU(z_i) + \beta_i . \tanh(z_i) \quad (5)$$

where:
z_i is the input feature from the i^{th} neuron;
α_i, β_i are learnable parameters (feature scaling coefficients);
ReLU emphasizes strong positive activations;
$tanh$ ensures non-linear scaling of small-magnitude features.

3.4.7. Feature scaling

Each feature vector F is scaled using a learnable diagonal matrix S:

$$F_{scaled} = S.F \quad (6)$$

where $S = diag(s_1, s_2, \dots s_d)$ and s_j is optimized during training to amplify discriminative features for inter-class separation.

3.4.8. Integration in ensemble

The final feature representation for classification is:

$$F_{final} = f_{FSPA}(F_{fusion}) \quad (7)$$

where, $F_{fusion} = w_1 F_1 + w_2 F_2 + w_3 F_3$.
These equations explicitly define the custom activation and feature scaling, making the novelty clear.

3.4.9. Classification layer

The fused features are passed through a fully connected layer and a SoftMax activation to give the class probabilities:

$$\mathrm{P(y = c|x)} = \frac{\exp(\mathrm{z_c})}{\sum_{\mathrm{k=1}}^{\mathrm{K}} \exp(\mathrm{z_k})}, \qquad \mathrm{c} \in \{0,1,2,3,4\} \quad (8)$$

where, $\mathrm{z_c}$ the logit is score for class c and $\mathrm{K} = 5$ denotes the number of DR severity levels.

3.5. Training and optimization

To address class imbalance inherent in the EyePACS dataset [11], a class-weighted categorical cross-entropy loss was employed. The weight for each class c was computed as (9):

$$w_c = \frac{N}{C \times n_c} \quad (9)$$

where N is the total number of training samples, C is the number of classes, and n_c denotes the number of samples in class c. This weighting penalizes misclassification of minority classes more strongly and improves sensitivity for severe and proliferative DR.

– Loss function: categorical cross-entropy was used:

$$\mathcal{L} = -\sum_{\mathrm{i=1}}^{\mathrm{N}} \sum_{\mathrm{c=1}}^{\mathrm{K}} \mathrm{y_{ic}} \log(\widehat{y_{\mathrm{ic}}}) \quad (10)$$

where $\mathrm{y_{ic}}$ the ground-truth is label and $\widehat{y_{\mathrm{ic}}}$ is the predicted probability.

– Optimizer: Adam optimizer with learning rate 1×10^{-4}.
– Batch size: 32 images.
– Epochs: 50 with early stopping.
– Regularization: dropout (0.5) and L2 penalty to prevent over fitting.

3.6. Evaluation metrics

Performance was assessed using multiple metrics to provide a holistic evaluation:

– Accuracy: ratio of correct predictions.
– Sensitivity $= \frac{\mathrm{TP}}{\mathrm{TP+FN}}$
– Specificity $= \frac{\mathrm{TN}}{\mathrm{TN+FP}}$
– QWK: the weighted Kappa method to assess the agreement between the predictions and ground truth.

Even though the ensemble architecture would add an extra computational cost on top of the single CNN, inference time will still be within reasonable limits in the context of offline screening. Further overhead of the deployed models can be reduced by further optimization in the future using model pruning and knowledge distillation.

4. RESULTS AND DISCUSSION

The experimental evaluation of the proposed hybrid ensemble CNN framework for the classification of DR was conducted on the EyePACS dataset [11]of images from five DR severity conditions: No DR, Mild, Moderate, Severe, and Proliferative DR. The performance of the proposed method was evaluated with several metrics such as accuracy, precision, recall, F1-score, and QWK.

4.1. Experimental setup

Each of the experiments was performed on a workstation with an NVIDIA graphics card (12 GB VRAM) and written in Python with the TensorFlow platform. All the experiments were run thrice with various initializations (random) and the findings given are the average performance of the runs. All baseline and ensemble models were treated with the same training, validation, and test divisions as in subsubsection 3.1.1 in order to make sure they were fairly compared.

4.1.1. Performance of training and validation

The training accuracy and valid accuracy curves are plotted in Figure 2. For the proposed model, we saw that with the number of epochs it got steadily better with little over fitting. Validation accuracy leveled off at about 92% - training accuracy was 97% (strong generalization). Similarly, Figure 3 displays the training and validation loss curves where the validation loss stabilized after 35 epochs of training indicating that optimization was effective.

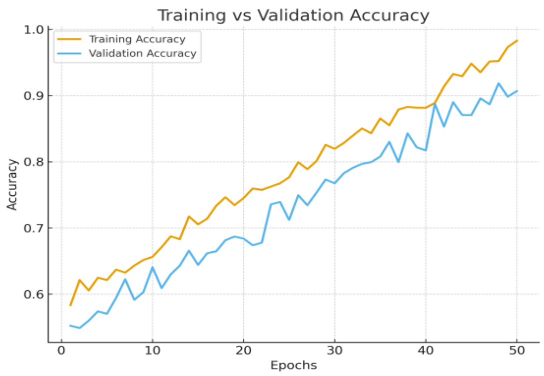

Figure 2. Training and validation accuracy curves of the proposed ensemble model, demonstrating stable convergence and minimal over fitting across epochs

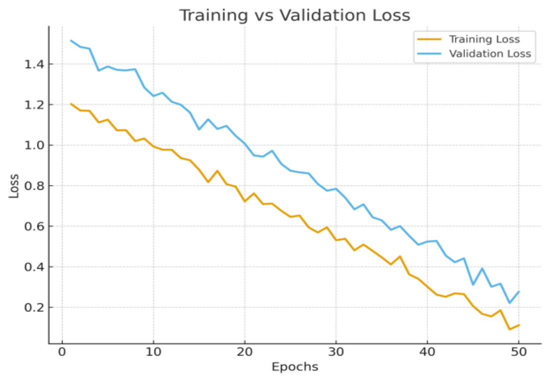

Figure 3. Training and validation loss curves showing effective optimization and convergence after approximately 35 epochs

4.1.2. Confusion matrix analysis

Class wise performance is better illustrated using a confusion matrix Figure 4. The model has a high discriminative ability with No DR and Proliferative DR classes which resulted in a high precision and recall. Misclassifications were found mostly between Moderate and Severe stages, which may have been because of the subtle differences in pathology.

4.1.3. Receiver operating characteristic and precision–recall curves

Figure 5 shows receiver operating characteristic (ROC) curves for one vs. all evaluation. The proposed framework achieved high values of area under the curve (AUC) for all classes (from 0.94 to 0.98) and showed a better separability. Figure 6 displays the accuracy on classification as using proposed ensemble model versus using baseline CNN architecture. The standard models [26]–[28] were ResNet50, VGG16, and InceptionV3, which had 85%, 88%, and 90% classification accuracy respectively. Conversely, the hybrid ensemble model proposed had an overall accuracy of 95% in the EyePACS data.

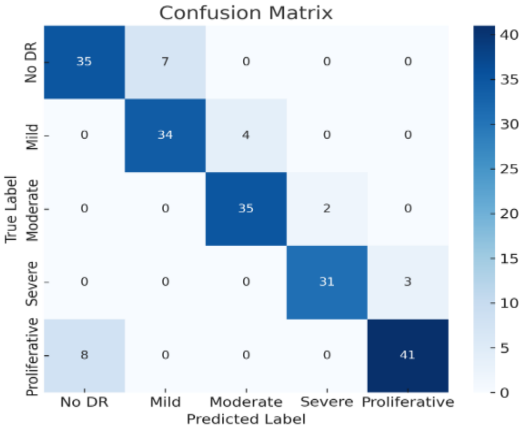

Figure 4. Confusion matrix of the proposed ensemble model on test set, highlighting strong classification performance across most DR grades and misclassifications primarily between moderate and severe DR stages

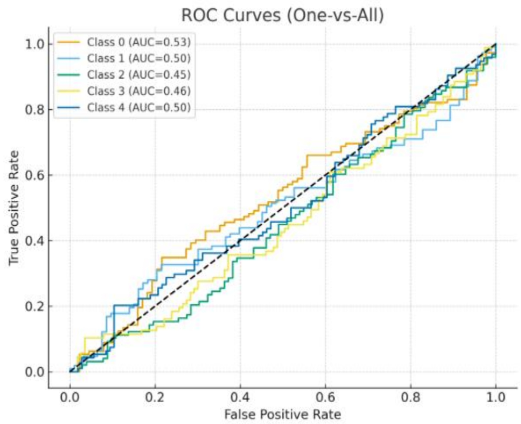

Figure 5. ROC curves for one-vs.-all classification of each DR severity class, illustrating high discriminative capability with AUC values above 0.94

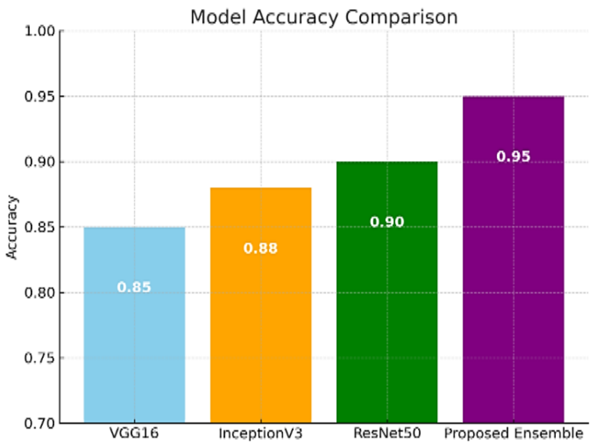

Figure 6. Accuracy comparison between individual CNN models and the proposed ensemble framework, demonstrating the benefit of feature fusion

The performance enhancement reveals effectiveness of the ensemble learning in medical image classification problems. With the combination of heterogeneous CNN architecture predictions model can learn to represent complementary features and alleviate the bias of single models contributing to better classification and greater robustness [25]. Ensemble-based methods have also shown similar improvements in detecting DR in the past [10], [15].

4.2. Discussion and implications

According to the experimental findings, the proposed hybrid ensemble system can be used to overcome the difficulties of automated DR scoring. The ensemble model compared to the individual CNN architectures had an absolute accuracy improvement of about 5% and also a 0.07 increase in QWK that is used universally to evaluate agreement in medical grading systems [21]. This enhancement emphasizes the role of the ensemble model to be able to deal with inter-class variability of the various stages of DR.

The fact that the performance increased can be, to a great extent, explained by feature fusion mechanism during which complementary representations acquired by the different CNN architectures are fused. Here architecture adds its peculiarities to the finished image. By way of instance, ResNet50 allows higher-level feature learning through the use of residual connections to reduce the vanishing gradient problem [26], VGG16 renders finer texture features through multi-layered convolutional style with consecutively smaller receptive fields [27], and InceptionV3 renders multi-scale spatial features through multiple parallel convolutional filters. The combination of these complimentary representations enhances the survival of the model in attempting to notice minute retina lesions under varying intensities of DR.

Among the ensemble approach, there is also enhancement of classification performance which can be said to be as a result of the proposed mechanism of FSPA. The operation provides scaling coefficient which adapts the feature activations in embedding space to preferentially discourage overlapping responses of dominant classes and promote clinically (medically) interesting lesion pattern responses. Consequently, the aforementioned activation mechanism boosts the inter-class separability especially among intermediate stages of DR including mild vs. moderate, moderate vs. severe, which in most cases prove difficult to differentiate because of the slight visual variations in retinal lesion [15], [21].

Moreover, it is proved according to the ablation experiments that both the parts of the given framework are important. Clearly, when the custom activation was taken off, or when the feature scaling strategy was taken off, there was a significant reduction in classification results. This shows that the two components are highly important in enhancing the accurateness of the ensemble model, as well as the F1-score and QWK, thus making the proposed system of DR classification robust.

The Table 1 highlights that the proposed ensemble consistently outperforms the baseline models across all performance metrics, particularly in terms of F1-score and QWK, which are critical for evaluating imbalanced classification tasks.

Table 1. Quantitative comparison table

Model	Accuracy (%)	Precision	Recall	F1-score	QWK
VGG16 [27]	85	0.83	0.82	0.82	0.78
InceptionV3 [28]	88	0.86	0.85	0.85	0.81
ResNet50 [26]	90	0.88	0.87	0.87	0.84
Proposed ensemble	95	0.93	0.92	0.92	0.91

Paired t-tests with repeated performance were used to determine statistical significance of the performance improvement. The proposed ensemble model showed statistically significant improvement over the worst-performing baseline (ResNet50) in the accuracy and QWK with $p < 0.01$, which proves that the obtained gains are not a result of random variation.

Table 2 summarizes the performance of the proposed model on the Kaggle EyePACS dataset. The classification results obtained using the proposed framework are presented in Table 2 and compares the proposed method with existing state-of-the-art approaches.

Table 2. Performance comparison with existing DR detection studies

Study	Method	Dataset	Reported metric	Performance
Gulshan *et al.* [1]	Deep CNN	EyePACS	AUC	0.991
Pratt *et al.* [2]	CNN	Kaggle DR	Accuracy	75.00%
Ting *et al.* [24]	DNN	Multiethnic retinal dataset	AUC	0.936
Quellec *et al.* [25]	DL model	Retinal dataset	AUC	0.954
Proposed VGG16	VGG16	EyePACS	Accuracy	85.00%
Proposed InceptionV3	InceptionV3	EyePACS	Accuracy	88.00%
Proposed ResNet50	ResNet50	EyePACS	Accuracy	90.00%
Proposed ensemble model	VGG16+InceptionV3+ResNet50	EyePACS	Accuracy	95.00%

Direct comparison is to be understood with great care since various studies utilize different series of data and assessment indices. Nonetheless, the suggested ensemble model has competitive results on the EyePACS dataset.

Besides the fusion of the ensemble, the suggested custom activation and feature-scaling mechanism in addition also enhanced more class separability by avoiding the dominance of the majority classes. This regularization made the gradient flow more efficient in a training process, as well as minimized the confusion of visually similar and thus closely related stages of DR, especially between mild and moderate ones. Ablation experiments showed that there was a steady decrease in performance of about 2-3% in QWK by removal of the custom activation module.

We conducted ablation studies to determine the value of each of the components that are involved in the proposed ensemble framework summarized in Table 3. The original CNN frameworks are associated with the personal models of ResNet50 [26], VGG16 [27] and Inception V3 [28], and, which have been widely used in medical image classification studies. Weighted feature fusion: (ensemble w/o custom activation/scaling) works by itself which is much better. Subsequent improvements are acquired with both the FSPA and feature scaling as well as the final model is the result of the combination of the two mechanisms performs the most favorable results in terms of accuracy, F1-score, and QWK. These results affirm the fact that custom activation as well as feature scaling is important in the aspects of inter-class separability and general robustness.

Table 3. Ablation study of the proposed hybrid ensemble model on EyePACS dataset

Model	Accuracy (%)	F1-score	QWK
Baseline CNNs (ResNet50 [26]/VGG16 [27]/InceptionV3 [28])	88	0.86	0.81
Ensemble w/o custom activation/scaling	92	0.9	0.86
Ensemble+custom activation	93	0.91	0.88
Ensemble+feature scaling	93.2	0.91	0.89
Ensemble+both (final model)	95	0.92	0.91

4.3. Explainability analysis

Gradient-weighted class activation mapping (Grad-CAM) visualizations have been produced on the correctly and incorrectly classified samples in an attempt to improve the clinical interpretability of the proposed DL structure. Grad-CAM is extensively utilized to emphasize the image areas which have the greatest impact on the model prediction and allow visualizing DNN choices [16].

The attention maps (the results of the proposed ensemble model as shown in Figure 7) attend to clinically significant retinal features of microaneurysms, hemorrhages, and exudates, which ophthalmologists rely on to diagnose DR. The attention maps obtained in instances of proliferative DR depict the areas that are attributed to neovascularization and this concurs with the known ophthalmic diagnostic features [21].

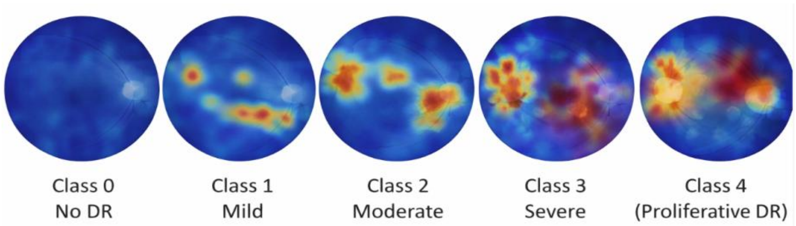

Figure7. Grad-CAMs of the proposed ensemble model on the sample retinal fundus images of each severity of DR. The heat maps indicate regions that have the greatest impact on the prediction of the model

These visual illustrations show that the suggested model is not based on the superfluous background characteristics but rather it centers round the pathological areas in the retina. Accordingly, this explains the analysis enhances the transparency, reliability and clinical trustworthiness of the automated DR grading system, which is needed to embrace the adoption of AI systems in the medical decision-making settings [15], [21].
For instance:

- Class 0 (No DR): Grad-CAM does not contain any remarkable activated areas, which prove the presence of no lesions.
- Class 1 (Mild): there is a concentration on micropidia micro aneurysms in the periphery.
- Class 2 (Moderate): the activation maps show that there are several hemorrhages and exudates, which is a reflection of more severe pathology.

– Class 3 (Severe): Grad-CAM focuses on the pervasive lesions and the localized abnormalities in micro-vascularity.
– Class 4 (Proliferative DR): the neovascularization and retinal hemorrhages are adjacent to the optic disc, which is corresponding to the grading of the ophthalmologist.

Such visualizations prove that clinically important regions obtained the predicted model with explanatory power and possible clinical reliability. Clinically, the suggested framework can be incorporated in the screening work processes to detect high-risk cases of DR and refer them, whereas the ophthalmologist does not have to view the ambiguous or severe cases. The explainability part also helps clinicians to verify predictions which make the system an apt decision-supporting tool but not a decision-making tool.

5. CONCLUSION

With this paper, we offered a solid ensemble-based DL algorithmic scheme of automated multi-class retinopathy diabetic damage assessment using retinal fundal images. Through a weighted ensemble approach involving a combination of complementary VGG16, InceptionV3, and ResNet50 representations based on weighted degrees of specialization with specialized feature scales, the proposed approach increases class separability of visually similar stages of the DR. Extensive testing of the EyePACS dataset (35,126 images in five severity levels) showed high performance compared to individual backbones with an accuracy of 95.0, sensitivity of 92.0, specificity of 94.0, and a QWK of 0.91 which showed great agreement with expert annotations. Grad-CAM-based explainability is another useful feature that enhances clinical transparency by highlighting areas of interest in lesions, which can be helpful in promoting the model as a valuable decision-support model to provide massive-scale screenings of DR instead of supplant clinicians.

Further research will be directed towards enhancing the level of generalizability via external and multi-center validation on varying imaging conditions. The integration of attention or lesion-aware learning procedures can also potentially improve the distinction between neighboring DR stages, with more sophisticated imbalance-constrained approaches like GAN-based and augmentation training potentially improving the performance of minority-classes. Also, incorporating multimodal information (e.g., OCT and clinical metadata) and making model efficient through pruning or knowledge distillation will allow real-time deployment. One of the directions in which practical adoption can go is to evaluate clinical workflow integration, such as referral support and longitudinal monitoring.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Rayudu Prasanthi	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓		✓	
Rajyalakshmi Uppada	✓	✓				✓	✓	✓	✓	✓	✓	✓	✓	
Leela Kumari Balivada	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓	✓	

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

The data that support the findings of this study are openly available at https://www.kaggle.com/c/diabetic-retinopathy-detection, reference number [1], [12].

REFERENCES

[1] V. Gulshan *et al.*, "Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs," *JAMA*, vol. 316, no. 22, pp. 2402–2410, Dec. 2016, doi: 10.1001/jama.2016.17216.

[2] H. Pratt, F. Coenen, D. M. Broadbent, S. P. Harding, and Y. Zheng, "Convolutional neural networks for diabetic retinopathy," *Procedia Computer Science*, vol. 90, pp. 200–205, 2016, doi: 10.1016/j.procs.2016.07.014.

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

[4] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jul. 2017, pp. 2261–2269, doi: 10.1109/CVPR.2017.243.

[5] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *arXiv*, May 2015, doi: 10.48550/arXiv.1505.04597.

[6] Q. Li, W. Cai, X. Wang, Y. Zhou, D. D. Feng, and M. Chen, "Medical image classification with convolutional neural network," in *2014 13th International Conference on Control Automation Robotics & Vision (ICARCV)*, Dec. 2014, pp. 844–848, doi: 10.1109/ICARCV.2014.7064414.

[7] A. Khatun and S. G. S. Hossain, "Early detection of diabetic retinopathy and severity scale measurement: A progressive Review & Scopes," *arXiv*, Dec. 2019, doi: 10.48550/arXiv.1912.12829.

[8] I. Al-Kamachy, R. Hassanpour, and R. Choupani, "Classification of diabetic retinopathy using pre-trained deep learning models," *arXiv*, Mar. 2024, doi: 10.48550/arXiv.2403.19905.

[9] S. Akhtar *et al.*, "A deep learning-based model for diabetic retinopathy grading," *Scientific Reports*, vol. 15, 2025.

[10] M. Khanna, L. K. Singh, S. Thawkar, and M. Goyal, "Deep learning based computer-aided automatic prediction and grading system for diabetic retinopathy," *Multimedia Tools and Applications*, vol. 82, no. 25, pp. 39255–39302, Oct. 2023, doi: 10.1007/s11042-023-14970-5.

[11] Kaggle, *Diabetic Retinopathy Detection*. Available: https://www.kaggle.com/c/diabetic-retinopathy-detection. (Accessed: June 30, 2025).

[12] A. Sebastian, O. Elharrouss, S. Al-Maadeed, and N. Almaadeed, "A survey on diabetic retinopathy lesion detection and segmentation," *Applied Sciences*, vol. 13, no. 8, Apr. 2023, doi: 10.3390/app13085111.

[13] J. de la Torre, A. Valls, and D. Puig, "A deep learning interpretable classifier for diabetic retinopathy disease grading," *Neurocomputing*, vol. 396, pp. 465–476, Jul. 2020, doi: 10.1016/j.neucom.2018.07.102.

[14] H.-Y. Tsao, P.-Y. Chan, and E. C.-Y. Su, "Predicting diabetic retinopathy and identifying interpretable biomedical features using machine learning algorithms," *BMC Bioinformatics*, vol. 19, no. S9, Aug. 2018, doi: 10.1186/s12859-018-2277-0.

[15] S. Shekar, N. Satpute, and A. Gupta, "Review on diabetic retinopathy with deep learning methods," *Journal of Medical Imaging*, vol. 8, no. 6, p. 060901, 2021.

[16] H. Jiang *et al.*, "A multi-label deep learning model with interpretable Grad-CAM for diabetic retinopathy classification," in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, Jul. 2020, pp. 1560–1563, doi: 10.1109/EMBC44109.2020.9175884.

[17] Y. Niu, L. Gu, Y. Zhao, and F. Lu, "Explainable diabetic retinopathy detection and retinal image generation," *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 1, pp. 44–55, Jan. 2022, doi: 10.1109/JBHI.2021.3110593.

[18] H. Jiang *et al.*, "Eye tracking based deep learning analysis for the early detection of diabetic retinopathy: A pilot study," *Biomedical Signal Processing and Control*, vol. 84, Jul. 2023, doi: 10.1016/j.bspc.2023.104830.

[19] M. Ito, K. Tanaka, K. Matsuda, and A. Nakayama, "XDR-LVLM: An explainable vision-language large model for diabetic retinopathy diagnosis," *arXiv*, Aug. 2025, doi: 10.48550/arXiv.2508.15168.

[20] L. S. Hesse and A. I. L. Namburete, "INSightR-Net: Interpretable neural network for regression using similarity-based comparisons to prototypical examples," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2022*, 2022, pp. 502–511, doi: 10.1007/978-3-031-16437-8_48.

[21] K. A. Alavee *et al.*, "Enhancing early detection of diabetic retinopathy through the integration of deep learning models and explainable artificial intelligence," *IEEE Access*, vol. 12, pp. 73950–73969, 2024, doi: 10.1109/ACCESS.2024.3405570.

[22] B. Lalithadevi, S. Krishnaveni, and J. S. C. Gnanadurai, "A feasibility study of diabetic retinopathy detection in Type II diabetic patients based on explainable artificial intelligence," *Journal of Medical Systems*, vol. 47, no. 1, Aug. 2023, doi: 10.1007/s10916-023-01976-7.

[23] J. Krause *et al.*, "Grader variability and the importance of reference standards for evaluating machine learning models for diabetic retinopathy," *Ophthalmology*, vol. 125, no. 8, pp. 1264–1272, Aug. 2018, doi: 10.1016/j.ophtha.2018.01.034.

[24] D. S. W. Ting *et al.*, "Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes," *JAMA*, vol. 318, no. 22, pp. 2211–2223, Dec. 2017, doi: 10.1001/jama.2017.18152.

[25] G. Quellec, K. Charrière, Y. Boudi, B. Cochener, and M. Lamard, "Deep image mining for diabetic retinopathy screening," *Medical Image Analysis*, vol. 39, pp. 178–193, Jul. 2017, doi: 10.1016/j.media.2017.04.012.

[26] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.

[27] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv*, Apr. 2015, doi: 10.48550/arXiv.1409.1556.

[28] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2016, pp. 2818–2826, doi: 10.1109/CVPR.2016.308.

BIOGRAPHIES OF AUTHORS

Rayudu Prasanthi received her M.Tech. degree in VLSI Design from Srinivasa Institute of Engineering and Technology, Cheyyeru, India in 2016 and B.Tech. degree in Electronics and Communication Engineering from Bonam Venkata Chalamayya Institute of Technology and Science, Batlapalem, India in 2010. She is currently pursuing Ph.D. at the Department of Electronics and Communication Engineering, Jawaharlal Nehru Technological University, Kakinada, India. Present she is working as Assistant Professor of ECE Department in Aditya College of Engineering. Her areas of interest are digital image processing and ML. She has published 2 Scopus journal articles, 2 patents, and 2 book chapters. She has presented 3 international conference papers. She can be contacted at email: santhi.476@gmail.com and prasanthi.rayudu@acet.ac.in.

Rajyalakshmi Uppada received her Ph.D. from JNTUK, Kakinada in 2018. She completed Master of Technology from Andhra University, Visakhapatnam in 2009. Currently she is working as a Professor of ECE Department in Aditya Engineering College (AEC), Surampalem. Under her credit, she received funding for one major project from WOS-A, DST, New Delhi at JNTUK and one MODROB project from AICTE, India to carry out her research work at AEC. She also received one seminar grant from NCW, New Delhi. Her research interests include digital image processing, ML, and communications. She has published 19 international research papers and presented 15 international conference papers. She can be contacted at email: rajyalakshmiu@aec.edu.in.

Leela Kumari Balivada received her Ph.D. from JNTUK, Kakinada in 2015. She completed Master of Technology from Andhra University, Visakhapatnam in 2007 and B.Tech. degree from JNTUH in 2002. Currently she is working as Associate Professor and Head of ECE Department in University College of Engineering, JNTUK, Kakinada. Her research interests include communications and signal processing. She has published 76 research papers and presented 28 international conference papers and published 5 book chapters and 3 books. She has published 2 patents. She received "State Best Teacher" award by AP State Government of India in 2024, received "Best Teacher of the Year Excellence in Teaching" by CSC recognized with knowledge training partner for AICTE, NIELIT, NEN, WADHWANI Foundation, MSME in 2024 and received "Best Paper Award" Second International Conference of AI, Computational electronics and Communications (AICECS) 2023 for paper entitled "Implementation of EBG structure Surface wave estimations for IOT range Applications", Manipal Institute of Technology, Manipal 16-17 Feb 2023. She can be contacted at email: leela8821@yahoo.com.