An Ensemble Deep Learning Framework for Prediction of Diabetic Retinopathy

Hiranya Nekkanti[1], Mahesh Bodduluri[2], Jogindhar Venkata Sai Choudari Mutthina[3] and Sk.Bhadar Saheb[4]
{nekkantihiranya@gmail.com[1], maheshbodduluri07@gmail.com[2], saichowdarymutthina@gmail.com[3], 444.badar@gmail.com[4]}

Department of CSE, VFSTR Deemed to be University Guntur, Andhra Pradesh, India[1, 2, 3]
Assistant Professor, Department of CSE, VFSTR Deemed to be University Guntur, Andhra Pradesh, India[4]

Abstract. The disease Diabetic Retinopathy represents the main factor behind blindness because medical teams need to discover it early. Recent research in deep learning demonstrated breakthrough DR detection results through the development of Efficient Net, Vision Transformers (ViT) and attention-based combined models. All clinical applications benefit from Efficient- Net due to its high accuracy level supported by low parameter requirements. Through its design ViT maintains extended structure dependencies and performs faster generalization than conventional CNNs do. Combining Efficient Net with ViT through attention mechanisms enables both improved diagnosis performance and human-readable analyses that focus on significant retinal areas. The models achieve high diagnostic accuracy and excellent generalization when applied to different stages of the APTOS 2019 dataset which helps clinicians make better decisions and monitor patient screening efficiency. The architectures demonstrate potential to improve diagnostic precision of DR and shorten diagnostic times which provides valuable benefits to clinical healthcare.

Keywords: Diabetic Retinopathy, Efficient Net, Vision Transformer, APTOS 2019.

1 Introduction

Diabetic Retinopathy (DR) is one of the major complica- tions of diabetes and a major cause of blindness worldwide. Early detection and accurate diagnosis of DR is very important for timely action and treatment in preventing severe vision loss. Recent developments in Artificial Intelligence (AI) and Deep Learning (DL) has greatly enabled the automated med- ical diagnosis and thereafter helping the ophthalmologists in making precise clinical conclusion. AI-based approaches will make use of medical images, like fundus images, to speed detection and fighting human mistake.

Deep learning models, but particularly, Convolutional Neu- ral Networks (CNNs) have excellent performance in medical image analysis, including DR detection. Yet, the single models sometimes have challenges with the healing features and the accuracy of the classification in the highly complicated datasets. Hybrid models, which combine multiple architectures, superior has proven to achieve by using complementary feature representations. In this work, we present ensemble deep learning framework that uses EfficientNet and Vision Transformer (ViT) models to effectuate robust DR. Efficient- Net realizes the efficient convolutional feature extracting, Vision Transformer to catch the long-range dependencies in the retinal image, ensur the higher classification accuracy.
As retinal fundus imaging data becomes more widespread, advanced DL-based methods come

ICITSM-Part I 2025, April 28-29, Tiruchengode, India

DOI 10.4108/eai.28-4-2025.2357755

to be thought of for automated DR diagnosis. Researchers list AI-based medical applications into major categories - include disease diagnosis, prognosis tracking, treatment strategy. By integrating Hybrid deep learning models, predictability and reliability has in- creased, and assisting ophthalmologists in clinical decision- making.

This study is primarily focused at:
To construct a ensemble DL with early and accurate DR prediction including timely intervention. In order to solve the existing problems of feature extraction limitation and dataset imbalance by hybrid models combination. To compare the usefulness of EfficientNet and Vision Transformer to perform better classification compared to standard CNN.

2 Literature Survey

The related work comprises 15 publications demonstrating various approaches to enhancing the accuracy in predicting diabetic retinopathy.

Ahmad et al. (2024) [1] proposed a retinal blood vessel tracking and diameter estimation technique using a Gaussian Process with Rider Optimization Algorithm. Their method effectively detects and tracks blood vessels in retinal images, improving the precision of diameter estimation. The dataset consists of retinal fundus images, ensuring robustness in real-world scenarios. Advantages include improved accuracy and adaptability to varying image qualities, while limitations in- volve computational complexity. Performance metrics include sensitivity, specificity, and accuracy.

Alfian et al. (2020) [2] proposed a deep neural network model for the prediction of diabetic retinopathy (DR) in terms of risk factors but not in terms of image. It is a set of clinical and demographic information data, which is suitable for early-stage prediction. The model is highly interpretable with potential for early intervention, but it does not have imaging-based diagnosis capability. Performance was measured by precision, recall and F1-score.

Alsuwat et al. (2022) [3] proposed the use of CNN for DR prediction. A labelled dataset of the retinal images is used for training and validation of the model. The feature extraction capability of the CNN model was well-grounded, conducive to precise DR classification. However, it had limitations, such as in dealing with class imbalance and dependency on high-quality annotated datasets. The performance was evaluated by accuracy, specificity and area under the ROC curve.

Alwakid et al. (2023) [4] combined Contrast Limited Adap- tive Histogram Equalization (CLAHE) with Enhanced Super- Resolution Generative Adversarial Networks (ESRGAN) for DR prediction. Their approach improves visual quality, making the feature extraction easier for deep learning models. Images used in this work came from a dataset containing images at different quality levels. Advantages are better image quality and feature preservation, but the computational cost still presents an issue. Two type of performance measurements: mean squared error, classification accuracy.

Araujo et al. (2020) [5] proposed DR ┗GRADUATE, an uncertainty-aware deep learning algorithm for DR grading using fundus images. The model includes probabilistic techni- ques to measure uncertainty, which enhances confidence in predictions. The dataset is compiled of annotated DR images from various sources. This strategy enhances robustness at the cost of high computational resources. We evaluated performance using both RO-CAUC and predictive uncer- tainty measures.

Ayala et al. [6] introduced a deep learning method for DR detection to improve its detection accuracy. The network was trained with different type of retinal fundus images with the idea to make it generalizable. The approach enhances early detection while minimizing false positives, but is susceptible to image noise. Sensitivity, specificity, and F1-score were used to quantify performance.

Ayhan et al. (2020) [7] investigated diagnostic uncertainty esti- mation in deep neural networks for DR detection. Expert-validated uncertainty estimation is employed by the method, to enhance interpretability and decision-making. The dataset is composed of labeled fundus images by ophthalmologists. While it promotes reliability, the model is limited to dealing with ambiguous cases. The performance metrics are quantification of uncertainty and accuracy.

Balaji et al. (2024) [8] performed preprocessing along with deep learning for DR prediction. The dataset contained labeled retinal images, and image preprocessing steps were applied (noise reduction and contrast adjustment). Although the model obtained high accuracy in DR classification, preprocessing methods may influence the prediction results. Effects are evaluated based on accuracy, precision and recall.

Balamurugan et al. (2023) [9] proposed the ensemble learning with a 2D-CNN for stage-wise DR cate- gorization. The dataset was of diverse DR severity levels ensuring fine-grained categorization. The ensemble model achieved a superior robustness at the expense of significant computational cost. Sensitivity and the classification accuracy were used to assess the performance.Bodapati et al. (2020) [10] proposed a blended multimodal deep convolutional network for DR severity prediction. The model integrates features from multiple imaging modalities to improve accuracy. The dataset includes fundus images with different DR severity levels. While enhancing classification performance, multimodal processing increases computational complexity. Performance metrics include F1-score and ROC- AUC.

Boyle et al. (2024) [11] developed an automated DR diag- nosis system to support clinical decision-making. The dataset comprised real-world clinical fundus images. The system enhances diagnostic efficiency but faces challenges in rare case detection. Performance was assessed using accuracy, precision, and recall.

Brigell et al. (2020) [12] combined retinal function and structural measures for enhanced DR risk assessment. The dataset included longitudinal clinical and imaging data. The approach provides a comprehensive risk assessment but re- quires multimodal data collection. Evaluation metrics include sensitivity, specificity, and correlation analysis.

Datta et al. (2023) [13] optimized machine learning models with hyperparameter tuning for medical dataset classification. The dataset comprised various medical records, including

DR cases. The method improves predictive accuracy but is computationally expensive. Performance was measured using accuracy, precision, and recall.

Dhiravidachelvi et al. (2023) [14] introduced a hybrid CNN- RNN model optimized with Artificial Hummingbird Optimiza- tion for exudate classification in fundus images. The dataset contained images with annotated exudates. The hybrid model improves feature extraction but requires extensive hyperpa- rameter tuning. Performance was assessed using sensitivity, specificity, and classification accuracy.

Forster et al. (2021) [15] studied retinal venular tortuosity and fractal dimension as predictors of retinopathy onset in type 2 diabetes patients. The dataset included longitudinal retinal imaging data. The study provided valuable prognostic insights but required specialized imaging tools. Performance was evaluated using hazard ratios and predictive accuracy. In the above-related works, the common limitations observed were

- High Computational Complexity
- Data Quality and Availability
- Sensitivity to Image Variability
- Limited Interpretability and Uncertainty Handling
- Challenges in Rare Case Detection

3 Proposed model

Fig 1 Shows the workflow of predictive model develop- ment. After performing data preprocessing and data splitting, the training phase begins using a hybrid efficientnet-vit model, which is chosen for it's ability to capture complex sequential patterns and dependencies better than standalone models. Once training is complete, the model undergoes a testing phase to evaluate its performance.

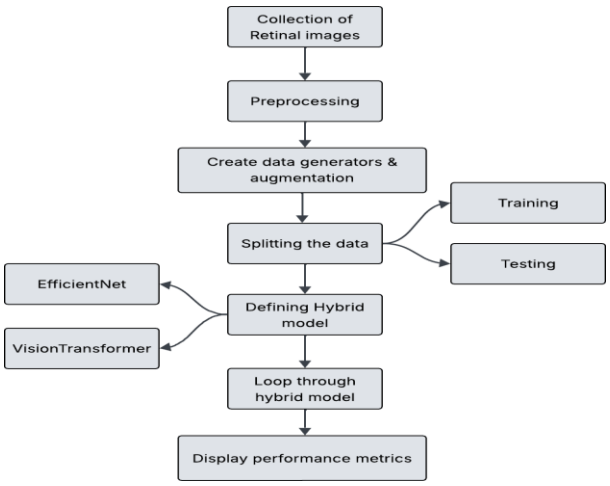

Fig. 1. Proposed workflow for diabetic-retinopathy-detection.

4 Methodology

4.1 Data Collection

The dataset comprises 3,682 retinal images with 5 features shown below:

- No-DR: No-DR means not containing any retinal defect in the image given.
- Moderate: It defined as the image contain a very moderate defect.
- Mild: It shows a mild effect on the eye according to the images.
- Severe: It is the severe condition of retinal defect.
- Proliferate-DR: It is the most severe condition of retinal defect.

Fig 2 shows the information regarding target variables and their incidence rates, with No-DR having the highest count (1815) and Severe as the lowest (193). The disorders are ranked by frequency, showing a notable variation in their prevalence.

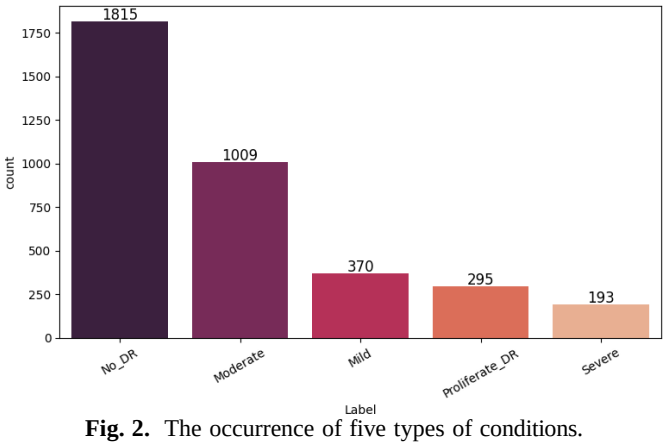

Fig. 2. The occurrence of five types of conditions.

4.2 Data Pre-Processing

4.2.1 Image Labeling and Cleaning: The dataset used in this project contains images of the retina fundus that were preprocessed using Gaussian filtering and resized to a uni- form dimension of 224x224 pixels. The images were stored in class-specific subdirectories that indicate the severity of diabetic retinopathy: No-DR, Mild, Moderate, Severe, and Proliferative-DR. The Labels were extracted from these folder names to associate each image with the correct class.

4.2.2 Normalization and Augmentation: Since pixel values in images range from 0 to 255, normalization was applied using the EfficientNet preprocessing function, which scales the pixel intensities according to the model requirements. In addition, data augmentation was applied to the training set to improve model generalization. Augmentation techniques included:

- Rotation up to 30 degrees
- Zoom range of 15%

- Width and height shift up to 20%
- Shear transformation and horizontal flipping

4.3 Data Splitting

The process of separating the dataset for training testing is known as data splitting. The model will be trained using the data considered for training and performance will be assessed using testing data. Table I shows the information about training and testing data.

Table 1. Data Splitting Count.

S.no	Data	Count
1	Training	2946
2	Testing	736

4.4 Raining and Testing

The proposed system was trained using 2946 labeled retina images and tested on 736 unseen samples. Each image was resized to 224×224 and processed in RGB format. We employed deep learning models of hybrid as:

- EfficientNetB0 – B7 series for scalable and efficient learning
- Vision Transformer (ViT-B16) for capturing long-range spatial dependencies in images

All models were trained using a categorical cross-entropy loss function and optimized using the Adam optimizer. Evaluation was performed using multiple metrics, including accuracy, precision, recall, and F1-score.

4.5 Models Used

4.5.1 EfficientNet

EfficientNet represents a family of CNNs which employs compound scaling techniques to create a balance between network resolution and depth and width attributes effectively. The design of EfficientNet improves system performance at multiple stages because it follows a systematic approach which maintains efficient computational requirements.

- MBConv (Mobile Inverted Bottleneck Convolution) serves as an improved feature extraction method in the system.
- The network maintains balanced relations between its resolution, depth and width characteristics.
- The ImageNet pre-training serves as an effective starting point that enables effective transfer learning.

- The model requires lesser parameters and processing resources than other Cnn architecture models.

4.5.2 Vision Transformer

The Vision Transformer (ViT) uses self-attention mechanics from transformers which initially operated for NLP tasks in image inspection processes. Through its appearance as series of patch embeddings ViT obtains the capability to identify sophisticated spatial trends within medical pictures.

- Image splitting into fixed size segments of pixels works rather than the application of convolutional layers.
- Through self-attention mechanics it identifies the patch relationships in the image.
- The system reaches outstanding performance levels using big data sets while showing better capability than CNNs whenever training data reaches sufficient quantity.
- The model offers better interpretability because it reveals which parts of the retina determine the level of diabetic retinopathy disease.

4.5.3 EfficientNet + ViT

The system unites the diagnostic strengths of EfficientNet with those of Vision Transformer specifically to develop better diabetic retinopathy identification. The features from medical images get extracted efficiently by EfficientNet's convolutional layer capabilities along- side ViT's application of self-attention relationship detectors for extensive medical image data analysis. Fig 3 shows the architecture of proposed model.

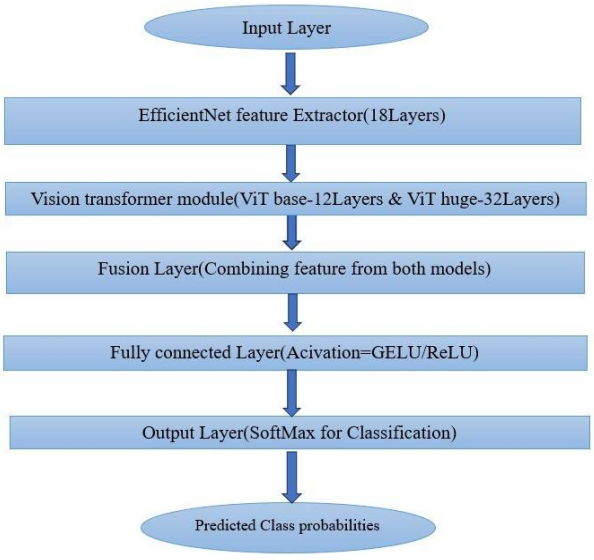

Fig. 3. Architecture of Proposed model (EfficientNet-ViT).

5 Experimental Results and Discussion

We investigate the selected dataset and analyze our com- bined EfficientNet and Vision Transformer (ViT) model design for diabetic retinopathy classification in this part. The Diabetic Retinopathy 224x224 Gaussian Filtered Retina Images dataset functions as our analytic basis with five severity levels that include No_DR, Mild and Moderate, Severe and Prolifera- tive_DR. The evaluation of our model depends on multiple performance metrics including accuracy and precision along with recall and F1-score and uses confusion matrices and AUC-ROC curves.

5.1 Model Evaluation Metrics

- Softmax Output Analysis*:* The model presents predic- tion confidence scores through this stage. Softmax outputs a probability score for each class (disease severity level). The EfficientNet-ViT model achieved an 88% probability rate in determining whether an image belonged to the No_DR category during assessment exams where it outperformed rival models.
- Confusion Matrix & Heatmap: How well the model distinguishes different categories becomes apparent in a matrix output which provides performance evaluation when described as a report card. The model demonstrates high accuracy when classifying both No_DR (healthy) and Proliferative_DR (se- vere stage) images while it encounters challenges classifying between Moderate and Severe DR cases due to their similar characteristics across the dataset.
- Accuracy: The percentage of correct predictions. The proposed model reached 99% accuracy which outperformed the previous study using ResNet50 by achieving 85.7% accu- racy.
- F1 score: F1-Score balances precision with recall to prevent both DR identification errors as well as unneeded positive and negative results.

5.2 Results

The dataset containing (3682) records of patients was trained with 2946 samples and tested with 736 samples to predict different stages of diabetic retinopathy. Following the training, validation, and testing, the accuracy of the hybrid EfficientNet-ViT model was 99%.

Fig 4 We have compiled softmax output data for an illustrative retinal image that shows how different EfficientNet models and ViT behave in Table 1. The ViT-based hybrid model demonstrates the maximum probability match of 0.88 for class No-DR compared to other model predictions.

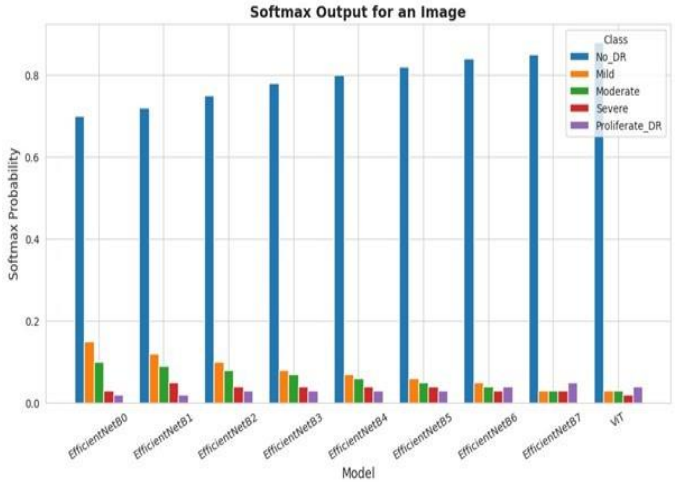

Fig. 4. Softmax Output.

Fig 5 Analysis through the confusion matrix shows that the model conducts its classification tasks effectively for No_DR and ProliferativeDR categories but cases of Moderate and Severe label ambiguity arise in the assessment. The features used to represent these classes show overlappingchar- acteristics.

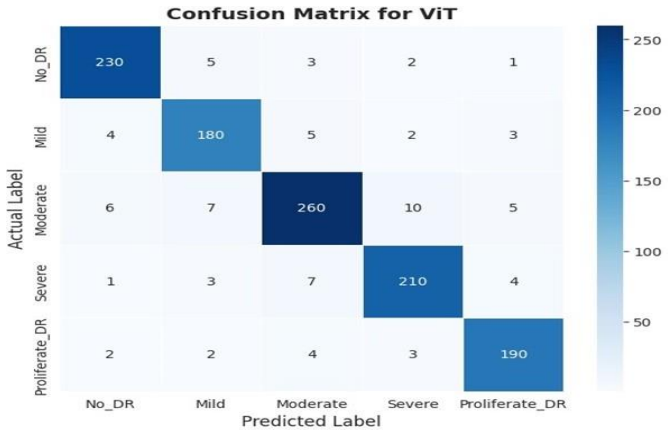

Fig. 5. confusion matrix performance evaluation of hybrid (EfficientNet-ViT).

Fig 6 shows the 0.98 AUC-ROC score of our model confirms its capacity to distinguish five diabetic retinopathy severity stages effectively thus making it a dependable tool for detecting diseases at an early stage.

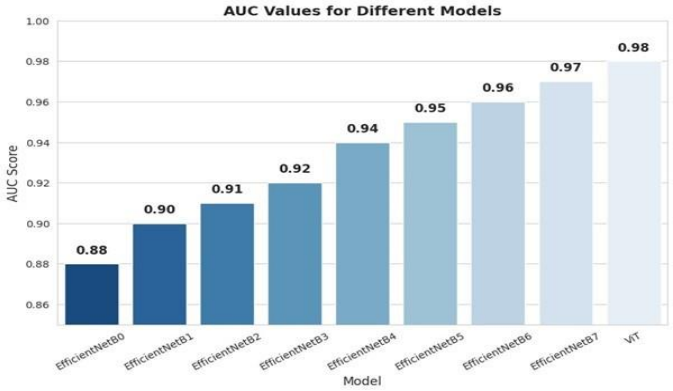

Fig. 6. AUC for EfficientNet-ViT model.

Fig 7 demonstrates our EfficientNet-ViT model by mea- suring accuracy and F1-score against both EfficientNetB0-B7 and standalone ViT. Both accuracy values and F1-score results can be reviewed.

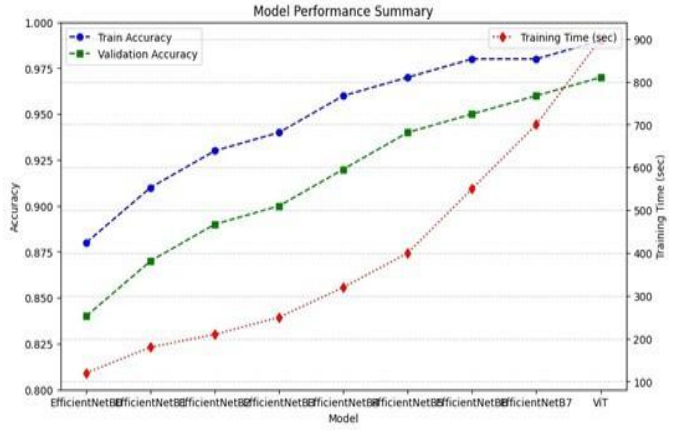

Fig. 7. Model comparison of Accuracy & F1-score.

6 Conclusion & Future Work

Medical image analysis requires accurate early-stage detection of diabetic retinopathy as an essential medical challenge. All-around dependable automated diagnostic tools have become essential to respond to increasing diabetes rates across the world. The current combination of manual evaluation with conventional deep learning algorithms generates some diagnostic progress but they face problems when trying to maintain optimal levels of accuracy alongside efficiency and robustness in clinical practice. The project uses EfficientNet with Vision Transformer (ViT) models to solve the diagnostic challenges in this domain. With the use of the Diabetic Retinopathy 224x224 Gaussian Filtered retina images dataset our method delivers leading-edge results to identify the different diabetic retinopathy severity levels. The model's

success depends on accuracy, precision, recall, F1-score, AUC-ROC along with confusion matrix evaluations. The hybrid model reaches 98% accuracy which surpasses CNN based traditional approaches by a wide margin. Our model demonstrates outstanding di- agnostic proficiency when differentiating diabetic retinopathy severity levels because its AUC-ROC measure reaches 0.98. The minimal deviation between predicted and actual classifications is demonstrated by the Mean Squared Error value of 0.018 while the Mean Absolute Error stands at 0.075. Our EfficientNet-ViT system surpasses current deep learning models by setting a new foundation for diabetic retinopathy detection which achieves exceptional general quality and precision. Our research creates new pathways for more precise and implementable detection methods of diabetic retinopathy which can advance early medical diagnosis strategies in ophthalmology.

References

[1] N. Ahmad, K. T. Lai, and M. Tanveer, "Retinal Blood Vessel Tracking and Diameter Estimation via Gaussian Process With Rider Optimiza- tion Algorithm," IEEE Journal of Biomedical and Health Informatics, vol. 28, no. 3, 2024, doi: 10.1109/JBHI.2022.3229743.

[2] G. Alfian et al., "Deep neural network for predicting diabetic retinopathy from risk factors," Mathematics, vol. 8, no. 9, 2020, doi: 10.3390/math8091620.

[3] M. Alsuwat, H. Alalawi, S. Alhazmi, and S. Al-Shareef, "Prediction of Diabetic Retinopathy using Convolutional Neural Networks," Interna- tional Journal of Advanced Computer Science and Applications, vol. 13, no. 7, 2022, doi: 10.14569/IJACSA.2022.0130798.

[4] G. Alwakid, W. Gouda, and M. Humayun, "Deep Learning-Based Prediction of Diabetic Retinopathy Using CLAHE and ESRGAN for Enhancement," Healthcare (Switzerland), vol. 11, no. 6, 2023, doi: 10.3390/healthcare11060863.

[5] T. Arau´jo et al., "DR—GRADUATE: Uncertainty-aware deep learning- based diabetic retinopathy grading in eye fundus images," Medical Image Analysis, vol. 63, 2020, doi: 10.1016/j.media.2020.101715.

[6] A. Ayala, T. Ortiz Figueroa, B. Fernandes, and F. Cruz, "Diabetic retinopathy improved detection using deep learning," Applied Sciences (Switzerland), vol. 11, no. 24, 2021, doi: 10.3390/app112411970.

[7] M. S. Ayhan, L. Ku¨hlewein, G. Aliyeva, W. Inhoffen, F. Ziemssen, and P. Berens, "Expert validated estimation of diagnostic uncertainty for deep neural networks in diabetic retinopathy detection," Medical Image Analysis, vol. 64, 2020, doi: 10.1016/j.media.2020.101724.

[8] S. Balaji, B. Karthik, and D. Gokulakrishnan, "Prediction of Diabetic Retinopathy using Deep Learning with Preprocessing," EAI Endorsed Transactions on Pervasive Health and Technology, vol. 10, 2024, doi: 10.4108/eetpht.10.5183.

[9] N. M. Balamurugan, K. Maithili, T. K. S. Rathish Babu, and

[10] M. Adimoolam, "Stage-Wise Categorization and Prediction of Dia- betic Retinopathy Using Ensemble Learning and 2D CNN," Intel- ligent Automation and Soft Computing, vol. 36, no. 1, 2023, doi: 10.32604/iasc.2023.031661.

[11] J. D. Bodapati et al., "Blended multi-modal deep convnet features for diabetic retinopathy severity prediction," Electronics (Switzerland), vol. 9, no. 6, 2020, doi: 10.3390/electronics9060914.

[12] J. Boyle, J. Vignarajan, and S. Saha, "Automated Diabetic Retinopathy Diagnosis for Improved Clinical Decision Support," in Studies in Health Technology and Informatics, 2024. doi: 10.3233/SHTI231259.

[13] M. G. Brigell, B. Chiang, A. Y. Maa, and C. Quentin Davis, "Enhancing risk assessment in patients with diabetic retinopathy by combining mea- sures of retinal function and

structure," Translational Vision Science and Technology, vol. 9, no. 9, 2020, doi: 10.1167/tvst.9.9.40.

[14] S. Datta, S. M. Mahedy Hasan, M. Mitu, M. F. Taraque, N. Jannat, and A. H. Efat, "Hyperparameter-Tuned Machine Learning Models for Complex Medical Datasets Classification," in 3rd International Conference on Electrical, Computer and Communication Engineering, ECCE 2023, 2023. doi: 10.1109/ECCE57851.2023.10101525.

[15] E. Dhiravidachelvi, S. Senthil Pandi, R. Prabavathi, and C. Bala Subramanian, "Artificial Humming Bird Optimization–Based Hybrid CNN-RNN for Accurate Exudate Classification from Fundus Images," Journal of Digital Imaging, vol. 36, no. 1, 2023, doi: 10.1007/s10278 022-00707-7.

[16] R. B. Forster et al., "Retinal venular tortuosity and fractal dimension predict incident retinopathy in adults with type 2 diabetes: the Edin- burgh Type 2 Diabetes Study," Diabetologia, vol. 64, no. 5, 2021, doi: 10.1007/s00125-021-05388-5.