Autonomous Diagnosis of Diabetic Retinopathy through Image Interpretation

Surya Teja Moparthi[1*], Abhilash R[2], Sumresh N[3] and K. Rajesh[4]
{mt6200@srmist.edu.in[1], ar7051@srmist.edu.in[2], sn8170@srmist.edu.in[3], rajeshk2@srmist.edu.in[4]}

Department of Computer Science, SRM Institute of Science and Technology, Bharathi Salai, Ramapuram, Chennai, Tamil Nadu, India[1, 2, 3, 4]

Abstract. Diabetic Retinopathy caused by diabetes mellitus, is among the most significant complications that lead to vision impairment and blindness without some form of timely identification. Traditional techniques for the diagnosis of DR are labor intensive and rely on specialized medical expertise, hence the difficulty in implementing widespread screening. Recent advances in state-of-the-art deep learning tools, particularly Convolutional Neural Networks, have been exploited with remarkable success for the automated identification of retinal fundus images with DR. [15] This paper presents CNN model for automatic detection and classification of severity of DR. The proposed model is trained and tested on the Kaggle Diabetic Retinopathy data set. Microaneurysms, exudates or haemorrhages can be found in retinal fundus images, all of which are signs of the DR potential phase, for a highly active diabetic population. Execution is Python programming Keras, OpenCV, concerned libraries and own NumPy one for mechanism functions to simplify image process and model functionality for them. Augmentations are applied to cater for the class imbalance and to contribute to model generalization. The analysis of using different CNN architectures such as ResNet50 and DenseNet201 have been carried out for capturing complex features in the retina. This method provides a precise and effective tool for DR detections to assist the doctors in order to take prompt action and to reduce the paths for losing the sight of the diabetes patients.

Keywords: Diabetic Retinopathy, Convolutional Neural Network (CNN), Retinal Fundus Images, Automated Diagnosis, Keras, OpenCV, NumPy.

1 Introduction

Diabetic Retinopathy (DR) is an important complication of diabetes with severe consequences that causes serious visual impairment in the world. It happens when the tiniest blood vessels of the retina are damaged as a result of chronic exposure to elevated levels of blood sugar, from leak and edema, and in very advanced cases, abnormal growth. Untreated, DR can cause irreversible blindness. Early detection of glaucomatous damage is a key part of preserving vision, but the point in the traditional examination process has depended on human eyes with long time-consuming manual operation and the high subjective and the potential erroneous result [1]. To solve these problems; technology like AI and deep learning, which is of machine learning, has been helpful to carry out computerization of the diagnosis using the medical images. Convolutional Neural Networks (CNNs) could be a hope in the detection of DR from retina fundus images with a decent accuracy [5]. A CNN based DR detection system has been developed in [7] that might be capable to classify retinal images among different levels of severity, No DR, Mild, Moderate, Severe, Proliferative DR using public dataset and a novel

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

DOI 10.4108/eai.28-4-2025.2358104

preprocessing algorithm for local contrast enhancement and noise removal to improve the feature extraction.

1.1 Domain & Its Usage

Recently, artificial intelligence and deep learning have even disrupted the medical diagnostics within the realm of ophthalmology. AI-driven models are increasingly applied to various applications in healthcare for automated disease detection, risk assessment and clinical decision support.

Healthcare: AI driven diagnostic models help to identify and categorise DR from the retinal fundus images which further helps in lubricating the incidence of blindness as well early intervention [1]. These models examine the retinal characteristics, including microaneurysms, hemorrhages, and exudates, in order to grade the severity levels of DR properly.

Telemedicine: DR screening tools based on AI can be incorporated into telemedicine systems for remote DR screening and diagnosis, especially in areas sparse in ophthalmologists [5].

Medical Image: Deep learning methods such as CNNs are widely employed to analyze the medical images to boost the accuracy and the efficiency in the disease detection [7].

Public Health Screening: With AI-based DR detection systems, mass screening programs are being introduced for governments and healthcare units which can identify high-risk diabetic patients and treat them in their early stages [9].

Assistive AI for Clinicians: Explainable AI (XAI) methods (e.g., Grad-CAM visualization) assist ophthalmologists in understanding model predictions by visualizing which retinal areas drive the AI decision, and increasing trust and acceptance by the clinicians in a clinical setup [2].

1.2 Domain Application

AI based algorithms for DR detection not only play a supporter role to the clinicians but also offering affordable and reachable diagnosis of DR. The role of Imaging: The system should be able to use several deep learning architectures to process retinal images and this is another scale in which the efficiency for DR detection happens Deep learning framework’s like TensorFlow and PyTorch integrated with open source medical image processing libraries like OpenCV are interfaced to build/ train, test, deploy AI-based models/drivers for detection of DR As a part of these use-cases pre-processing an image can help find which object features we want to classify, extract features from images before they can be used by machine-learning algorithms to perform its role i.e real-time clinical diagnosis and on-time decision-making. In short, the AI Based Diabetic Retinopathy Detection System is a nascent innovation which could potentially be utilized to better eye diagnosis with an all-around automated DR screening procedure-initiation though end. Fast AIDR -- so you can diagnose faster (and more correct) and newcomers to the medical profession/patients who need immediate help.

Diabetic Retinopathy appears as a flow which must range a variety of stages according to the plethora of graceful vascular abnormality that lie within the retina:

Mild NPDR: microaneurysms (lil bumps on the retinal capillaries) are visible;

Moderate NPDR: additional microaneurysms, intraretinal hemorrhages, and potentially cotton wool spots;

3. Severe NPDR: even larger areas of hemorrhages and blockage of the blood vessels lead to ischemia.

4. Proliferative DR: This condition leads to neovascularization, or insult of a growth of tiny, weak blood vessels in the eye into the vitreous cavity, sometimes leading to bleeding into the vitreous or retinal detachment. The cohort at greatest risk for permanent vision loss includes individuals moving from nonproliferative diabetic retinopathy to proliferative disease, who, therefore, require surveillance and timely intervention.

2 Related Work (Literature Review)

Deep learning models for DR detection:

A prospective study trained and validated a deep learning model on private datasets and tested it in real time at the Sindh Institute of Ophthalmology & Visual Sciences (SIOVS). It achieved 93.72% accuracy, 97.30% sensitivity, and 92.90% specificity, but its performance was heavily dependent on image quality, making standardized imaging protocols critical [1].

Another meta-analysis of 60 studies covering 445,175 fundus image interpretations confirmed high diagnostic accuracy across machine learning models, though the lack of external validation raised concerns of spectrum bias and limited generalizability [2].

CNN-based models trained on large datasets such as EyePACS have been applied to classify DR severity into multiple stages, with strong accuracy but challenges of overfitting and data quality dependence [3].

Other architectures, including EfficientNet, have also been investigated for automatic DR detection, with competitive performance across classification tasks [5].

Similarly, new grading models based on deep learning demonstrated strong sensitivity and specificity but lacked generalizability when tested on images from diverse populations and imaging devices [12].

Explainable AI and interpretable methods:

Developing transparent and interpretable models for DR detection is increasingly important. Shahzad et al. proposed a transparent diagnosis model using explainable AI (XAI) within the IEEE Access framework, providing interpretability alongside strong diagnostic performance [1].

The ExplAIn framework was introduced by Quellec et al., enabling explanatory AI for DR diagnosis, offering lesion-based interpretability [6] [19] [20].

Neural-symbolic approaches have been applied for explainable and interpretable DR classification, bridging symbolic reasoning and deep networks [7] [17].

Niu et al. combined explainable DR detection with generative retinal image synthesis, integrating classification and image interpretation [9].

Islam et al. developed explainable ML algorithms for predicting DR risk, showing strong predictive performance but highlighting the need for broader validation [10].

An interpretable classifier proposed by de la Torre et al. for DR grading was among the earlier works stressing interpretability in automated grading [11].

Romero-Oraá et al. introduced an attention-based deep learning framework for fundus image analysis that improves interpretability in DR grading [18].

Shen et al. developed an interpretable detection method for fundus diseases using deep learning, presented at the 3rd International Conference on Bioinformatics and Intelligent Computing [20].

Hybrid and advanced explainable models:

Hybrid XAI models have been used to integrate metabolomics data with fundus imaging, creating a multi-modal explainable framework for DR detection [8].

Explainable DR detection has also been explored using GAN-based methods, such as the EnlightenGAN framework proposed by Kannan et al., which augments interpretability by generating human-interpretable features [16].

Reviews and future outlook:

Reviews such as Skouta et al. provide a comprehensive survey of deep learning for DR assessment, including explainability perspectives [15].

Balyen's review emphasizes new approaches in DR detection and management, offering insights into future clinical translation [4].

Lee and Au provided an appraisal of FDA-approved AI algorithms for DR screening, underscoring the practical implications of integrating AI into clinical workflows [2].

Contextual and broader applications:

Beyond DR, explainable AI techniques have also been applied to related ophthalmic diseases, such as macular degeneration detection, showing transferable principles across medical imaging domains [3].

Industry and public interest perspectives are also emerging: Santilli highlighted in PopSugar how AI is increasingly influencing eye health [13], while Pillay's Time article profiled AI leaders driving such innovations [14].

3 Proposed Methodology

The proposed methodology uses a CNN-based deep learning approach for the automated detection and classification of Diabetic Retinopathy (DR) from fundus images of the eye. In this process, the images are subjected to data preprocessing, where the images are resized and normalized. Contrast enhancement, green channel extraction, and noise reduction techniques augment visibility of lesions. To increase dataset diversity and promote generalization for the model, other augmentation techniques include rotation, flipping, and variation in brightness. We design a custom CNN architecture composed of multiple convolutional layers for extract- ing the hierarchical features with batch normalization to maintain the model stability and dropout layers to avoid the over-fitting. The model works on the training data with the categorical cross-entropy loss and the Adam optimizer were used, and we used the metrics accuracy, sensitivity, specificity and F1 score. For enhanced model interpretability, visualization techniques, including Grad-CAM, were incorporated to focus on vital retinal regions influencing the classification decision. This source method is designed to provide an effective, scalable and interpretable approach for early DR detection to help doctors' intervention and treatment.

3. 1 System Architecture

This system provides a structured pipeline in precise lesion classification and diagnosis in diabetic retinopathy employing deep learning with image processing methods. It is based on a well-structured processing pipeline of data acquisition, pre-processing, feature extraction, classification, and validation. Fig. 1 shows the Architecture Diagram.

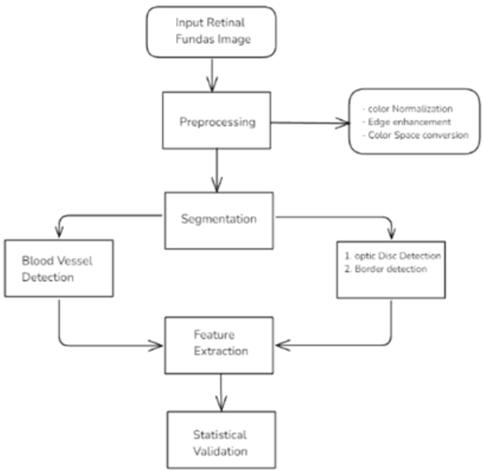

Fig. 1. Architecture Diagram.

1) 1) Input retinal fundus images: The system starts with input data of high-resolution so as to obtain the following results from publicly available database or created by based on large hospital-based collection source [3]. From these images, important structural information such as the optic disc, blood vessels and pathology signs of DR, can be

observed, LumineticsCore is an instance of Automatrive screening tools that provides real time detection of DR diagnosis without a doctors in the territory aid [8].

2) Preprocessing: Preprocessing ensures that the input retinal fundus image is transferred into feature extraction and classification ready status. The preprocessing module applies:
   - ● Color normalization: Normalizes variability in the intensity of the colors between all images making highly visible lesions.
   - ● Edge enhancement: The vessel and lesion structures are sharpened for segmentation.
   - ● color space conversion: Convert images to grayscale, RGB or HSV for better feature representation.

3) Segmentation: The isolated segmentation of the image will in turn manipulate that of the follow clinically relevant structures:

   - ● Vessels: Skeletonized using morphological operations and U-Net architectures based on deep learning.
   - ● Optic disc and border detection: This can assist in extracting the optic disc areas from pathological lesions in order to avoid overlapping or false positives, in order to detect the true DR.

4) Feature Extraction: The segmented parts will then undergo feature extraction, in which deep learning model detects and further stages up the lesions of DR Grenander and Miller [22] Sharma et al.

   - ● Microaneurysms: Small red dots appearing as the initial DR stage.
   - ● Microaneurysms: Small red dots, the first stage of DR.
   - ● Microaneurysms: Small red dots and the first stage of DR.

5) statistical validation: It is also utilized for machine and deep learning classifiers testing [10]. A few of the key performance evaluation metrics includes: Accuracy, Sensitivity and Specificity: evaluated to measure the model performance [12]. SHAP (Shapley Additive Explanations): Explains feature contributions [4].

3.2 Modules Description

The initial component of the whole pipeline consists in recovering high-quality of retinal fundus images from public or clinical databases. Typical datasets such as MESSIDOR, APTOS, and IDRiD provide both different demographic backgrounds and a comprehensive range of DR stages. Well — this is really a set of images that are training your deep learning algorithm to recognize certain patterns within the disease and so you have the medical expert ground truth annotation to begin with. Medical professionals add annotations to the images by highlighting areas with microaneurysms, hemorrhages, and exudates, to facilitate the beginning of the process with supervised learning. The quantity and quality of such annotations have a direct impact on the success of the model's training, and therefore its classification accuracy. The data will then be divided into training and testing sets to prevent overfitting of the model during training, and to enhance the generalization of the model.

Segmentation Segmentation is important to differentiate from other structures and lesions in the retina, and to avoid misinterpretation by the classification models.

Blood vessel extraction

Morphological operators and U-Net based deep learning architectures are employed for enhancing and segmenting vascular structures of the retina, thereby, aiding detection of pathologies such as neovascularization.

Optic disc detection

The optic disc is a bright circular area in fundus images and is therefore detected and removed from the images to not interfere with exudates.

Lesion segmentation

FCN, and active contour models are methods to accurately segment lesions, such as hemorrhages, microaneurysms etc. Segmentation guides the system to learn the important feature regions, and consequently, the number of false positives is reduced and the classification performance is improved.

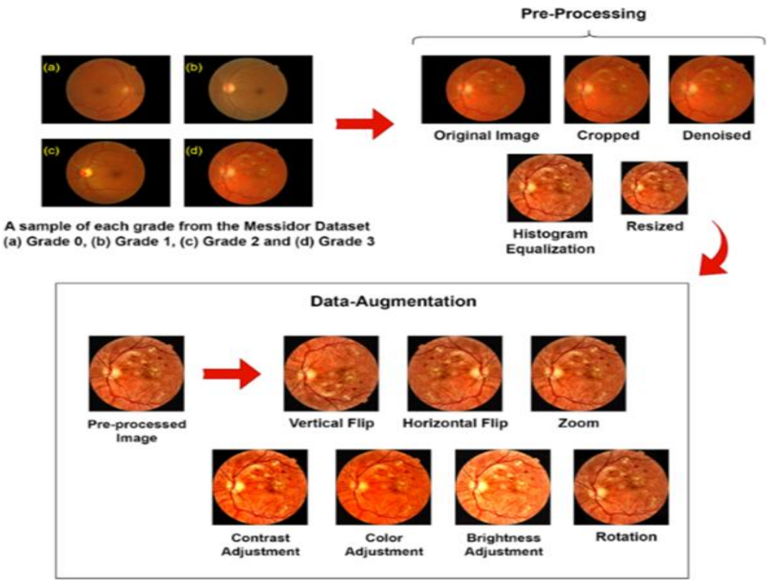

Fig. 2. Data Processing.

Thus, the classifier will be fed a host of strong features, thanks to combinations of different feature-extraction methods, which in turn should enhance the prediction accuracy. The next step is classification. By means of deep learning models, DR stages will be assigned classification labels: Normal, Mild, Moderate, Severe, and Proliferative DR. Fig 2. shows the Data Processing.

CNN-based classification: The CNN-based classification model plays an important role in the detection and classification of different stages of Diabetic Retinopathy (DR). Through the use of Convolutional Neural Networks (CNNs), the model can automatically learn and extract hierarchical patterns from retinal fundus images. These CNNs are trained to recognize features such as microaneurysms, hemorrhages, exudates, and neovascularization that classified the levels of DR into five severity: Non-DR, Mild, Moderate, Severe, and Proliferative DR.

Transfer learning is used to refine the quality and performance of any classification model. As opposed to Appling things from scratch, here pre-trained architectures including ResNet, InceptionV3, DenseNet asts are fine-tuned to domain-specific DR datasets. These are models which are pre-trained on a large-scale image dataset and such models have already learnt optimal feature extraction layers which can then be used a recycling point for DR detection. ResNet: It exploits Residual connection to perform deep feature leaning without the vocanishing gradients, InceptionV3: It contains multi-scaled spatial information, and DenseNet: It treats the flow of information in between layers to have enhanced learning more effectively. Exploiting such architecture, the model leads to a significant gain in building on classification by leveraging preexisting feature representations.

There is another bottleneck in the process of how to train CNN models for practical DR diagnosis: although pre-trained models are sufficiently powerful for our purposes, they are required to be tuned to the features of the retinal images. Carried out by any fine-tuning technique, this entails that first some of the layers of the CNN that correspond to general image characteristics be frozen and the rest of the layers – that are discriminative for recognition of the DR related abnormalities – retrained. Additionally, learning rate scheduling injects control into the optimization: updates to the weights that are too large can send the performance of the model into disarray. Second, hyperparameter tuning enables to tune the model more finely to get better parameter values like dropout rates, convolutional kernel size, and batch normalization settings, without overfitting at all. Resolving of class imbalances in DR datasets is handled using data balancing techniques as SMOTE and class weighted loss functions. These mechanisms make all severities learned effectively and do not favor learning more common classes. A model on this carefully labeled tiny dataset of retinal imaging data can enhance the accuracy and robustness, and be more reliable for practical clinical diagnosis.

Hybrid methods Both DL and ML can be exploited to bring a more robust performance to the DR classification and the combination of deep learning models with classical ML classifiers formed the hybrid approach. Although CNNs are proficient in learning hierarchical spatial features from images, conventional classifiers like SVM, Random Forest and XGBoost effectively perform refinement in decision. The hybrid solution is such that we make use of the feature extraction capabilities of deep learning and traditional classifiers' organized decision-making. In deep CNNs, powerful models are employed for the purpose of extracting discriminant retinal features, and these retinal features are then fed into a group of subsequent classifiers. SVM, for example, will try to slide the boundary so it allows more leniency in how much severity is different. For this strategy, Random Forest will be ensembled with XGBoost to make robust predictions. The multi-modality serves to perform ensemble learning and is able to suppress false negative without post-processing, which can achieve a higher positive predictive value for the context of detecting DR. The crosstalk of deep learning and traditional machine learning gives a cancer classifier to work more globally-- high sensitivity and low or no false positives for the easy application in clinics.

Model validation: Validation uses certainly F1-score as the main metric, which provides a good balanced measure precision and recall which is useful in our case because the dataset is imbalanced and some DR severity levels are very poorly represented; There are some other validations also like the cross-validation methods, where the data is partitioned into several subsets; an idea is to check model consistency among different data distributions. Additionally, by external validation with independent datasets demonstrates the ability of the model to make generalization beyond the training data. Continuous validation in the system will give full confidence that the DR classification model is for and can be used in the real medical practice, and is a trusted tool for ophthalmologists and other healthcare professionals.

3.3 Modular Framework for Automated Diabetic Retinopathy Detection

Diabetic Retinopathy (DR) is chronic ocular disease, that can result in lack of vision if not diagnosed early. With AI and deep learning clearly advancing, undoubtedly the automated DR screening systems have shown to be invaluable as a support tool for doctors to diagnose and grade the severity of DR, however, such system should not only base on creating an automatic module to interpret an image but it has to also embody in separate modules the ones that guarantee quality of the data; build automatic tools for feature extraction; promote models setting; empower model interpretations and finally adapt them over time. This DR detection modular framework seeks to address in a structured fashion fundamental challenges: data heterogeneity, model interpretation and dynamic learning. Then the introduced methods and modules in this paper could constitute a complete blood vessel removed DR detection pipeline, which is more dependable and robust. From pre-processing, through data to model forming, all the way to explainability and real world adapatiblity the modules are a scaleable set piece of a game changing diagnostic solution that you can trust.

Data Profiling and Automated Cleaning: This is the starting point of the pipeline where ins and outs are inspected into great detail for quality issues, input attributes. Artifacts, bad illumination and noise are common issues that can disturb precise analysis but automatical1 y detected by the software. Image processing (e.g. contrast enhancement and denoising) is applied as automatic cleaning steps following the background extraction. The data is now ready to move on to the next steps. The training and testing data are chosen from Kaggle APTOS 2019 Blindness Detection Dataset and EyePACS dataset.

Deep feature extraction and preprocessing: After images have been preprocessed, an extensive analysis is performed to find out features that are most associated with the diagnosis of diabetic retinopathy e.g., microaneurysms, hemorrhages and exudates. The methods utilized here are essentially more advanced types of pre-possessing: image normalization and augmentation, to standardize data and render the model more invariant to input image variations.

Augmented ML Model Generation: The module will generate models using advanced ML algorithms, which should be able to detect and categorize the various severity levels of diabetic retinopathy with high accuracy. In this context, the Convolutional Neural Networks (CNN) are used to capture hierarchical representations of images which describe detailed patterns at different scales of DR severity.

(2) eXplainable AI (XAI) Integration: - The use of XAI results enables us to supply transparent and interpretable predictions from our model. By interpreting the model in terms of which retinal features are influenced (a decision such as microaneurysms) clinicians can have an understanding into how the model is deciding, improving trust and allowing them to be aware when a medical decision is being made.

Online learning and Model Incremental Updating: Applying online learning to continuously update model compared with batch models in conventional initial training datasets, the model can adaptively change according to data patterns and medical evolution. As the model is updated with new retinal images, it will continue to train itself on the images and become more sensitive to the complex imaging shapes. Accordingly, the model remains modifiable with evolving methods for under-served and underprivileged patients.

Another important aspect of this strategy is the more global generalization; it gradually includes more diverse populations, imaging/art conditions, and clinical research sites. While always adding in data from new studies, real world hospital data and patient screens, the system is designed to reduce bias and overfitting to any particular data set. Generalization across different ethnic groups, different age groups, and different stages of DR are possible due to improved robustness.

3.4 Data Balancing for Improved Model Performance

In DR detection using deep-learning methods, the imbalanced dataset tends to make the model biased, and the bias will result in wrong predictions and, therefore, wrong treatment suggestions. In an unbalanced data, if the several levels of the severity carry much more samples than the rest of the levels, the model tends to be biased in favor of the most abundant class and cannot capture the finer details of the minor-class samples. This fact hinders accuracy towards other classes, particularly severe DR cases, where providing timely intervention prevents patient's blindness. So this is what necessary to be addressed and hence, we dealt with such class imbalances as they play a pivotal role in order to verify the fact that the model works good in terms of generalization among the groups of varying severity that, in turn, enhances the performance of classification and adds up to the reliability level. A refresher on how this is all done: the balanced dataset does not differentiate between the DR stages with the notion of a majority class bias. Fig. 3 Unbalanced Data representation is presented.

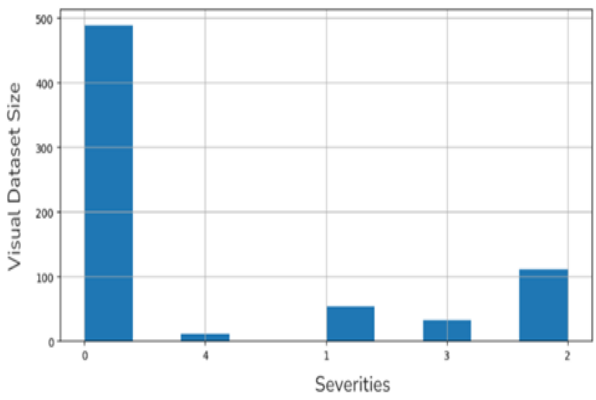

Fig. 3. Unbalanced Data representation.

3.5 Data Balancing Using Data Image Generator

The Data Image Generator creates more training samples to get an equal representation of every severity level to counter class imbalance. The generation of synthetic variations from the underrepresented categories helps the model get exposed to a more varied dataset, leading to improved learning and generalization. The Image Generator using augmentation applies several techniques: rotation, flipping, scaling, brightness changes, and elastic distortions to yield multiple variations of the supplied images. The transformations simulate various imaging conditions under which the model will learn to recognize certain critical retinal abnormalities, irrespective of brightness, orientation, or quality. Another balancing strategy to aid the model in learning consists of training it with various datasets: each containing different severity distributions. With this method, instead of a single one, multiple such batches does the model have to train in to learn the features progressively over each DR pattern; thus, other important but scarce cases would not get overlooked. It further prevents overfitting since the flexible classes will be represented across batched samples. In addition, stratified sampling dictates that every training batch contains equal proportions of each of the DR severity levels. This reinforces balanced learning, whereby the model is consistently presented with all classes throughout training. Fig. 4 shows the Balanced Data Representation.

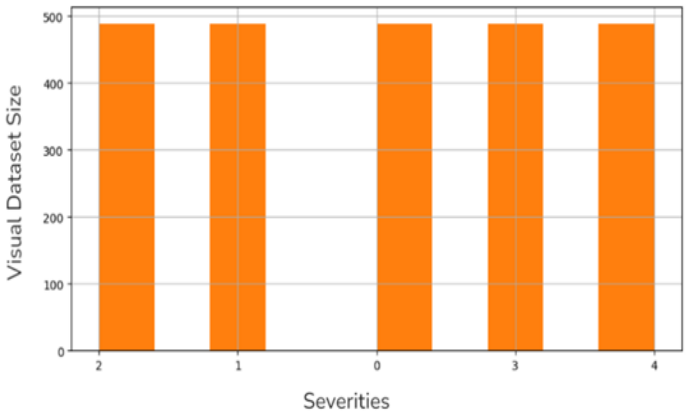

Fig. 4. Balanced Data Representation.

3.6 Outcome and Impact of Balanced Training

The Data Image Generator creates more training samples to get an equal representation of every severity level to counter class imbalance. The generation of synthetic variations from the underrepresented categories helps the model get exposed to a more varied dataset, leading to improved learning and generalization. The Image Generator using augmentation applies several techniques: rotation, flipping, scaling, brightness changes, and elastic distortions to yield multiple variations of the supplied images. The transformations represent different imaging scenarios in which the model will be trained on how to detect specific critical retinal abnormalities regardless of brightness, orientation and quality. Another balancing technique to help the model learn to predict is training with multiple datasets, which have different distributions of severity. With this technique, we do not force the model to train () batches to learn the features in a progressive way over each DR pattern; so things like other important but rare events are not ruled out. It also reduces overfitting since the variety of classes will be represented in batched samples. Furthermore, stratified sampling (i.e., each training batch has a balanced number of samples for every DR severity level) is enforced. This encourages balanced learning since the model will have samples of every class presented to it at all layers throughout the training process.

4 Feature Extraction

4.1 Metrics Used & Their Description

The process of feature extraction comprises creating an arrangement of procedures and functions with the main goal of automating diabetic retinopathy detection where the system would see and analyze the distinguishing key retinal abnormalities caused by DR in the retinal images, those that form the basis of differentiation from healthy cases. Some of the main features that are to be extracted include microaneurysms in the images, small red dots that are simply bulges in the retinal capillaries and are actually the earliest signs of DR. Later on, as the disease progresses, exudates will appear-they would be white or yellowish in color, homogeneous lipid or protein deposits in the retina that arise from fluid leaking out of the damaged blood vessels. The dark red spots caused by blood vessel rupture are called hemorrhages, and they may indicate that the patient is suffering from moderate to severe DR. Further, in the most advanced stages of DR, neovascularization occurs, a proliferation of abnormal blood vessels that constitutes a severe risk to vision impairment. The feature extractions are based mainly on image processing techniques like edge detection, morphological operations, and intensity thresholding. These are the ones that give very good highlighting of the important retinal abnormalities for the purpose of diagnosis. To that end, the deep learning-based feature extractors of all-convolutional neural networks would then find the hierarchical representation of these features by themselves in order to attain better accuracy in classifications, hence, a more robust performance in various imaging conditions.

4.2 Data Augmentation

In order to improve the model's generalization abilities and reduce overfitting, it uses data augmentation techniques. It artificially enlarges the dataset by changing the images so that the model will learn invariant representations of DR-affected images under varying imaging conditions. Some of the more important augmentation techniques are rotation: images are randomly rotated to simulate various head positions at the time of fundus photography; flipping, including horizontal and vertical flipping, is applied so as to accommodate changes in image orientation; brightness variation is introduced to simulate different illumination conditions so that the model can be exposed to diverse exposure; and elastic transformations are employed to impart geometric distortions so that natural variations of retinal images would be considered as well in augmenting the training examples.

4.3 Training and Optimization

During supervised training with retinal fundus images, the model is trained with the help of the labeled data by allowing the ground truth labels to guide the learning process. To measure a misalignment between true labels and predicted probabilities, the Categorical Cross-Entropy loss function is used, serving as a guide for the model in conjunction with its optimization techniques for improving predictions. The loss function is mathematically defined as:

4.4 Explainability & Visualization

Structural Grad-CAM has also been done with Grad-CAM images overplotted on retinal images with regions contributed to the model's decisions by the DR indicators so as to make them more interpretable and trustworthy to ophthalmologist.

4.5 Statistical Validation & Deployment

The final evaluation of the model is conducted using an independent test dataset to measure its effectiveness. The following metrics are used:

Accuracy: Measures the proportion of correctly classified cases.

Sensitivity (Recall): Determines the ability to correctly identify DR-positive cases

Specificity: Assesses how well the model avoids false positives.

F1-Score: A balanced measure of precision and recall.

These validation metrics ensure the model performs well in real-world clinical applications, detecting DR cases with high accuracy and reliability.

4.6 Continuous Learning Framework

A continuous learning framework has been implemented to allow the model to adapt to new retinal datasets over time. In such a framework, the model may:

Impose a limit on bias due to endless integration of veterinary patients.

Through incremental learning, data stream integration, and active learning, the model continues to remain state-of-the-art by yielding the utmost adaptability toward evolving DR trends.

4.7 Deployment and Integration

The trained model is integrated as a complete system into clinical environment through the friendly user's interface that gives access to ophthalmologists and health care workers for rapid diagnosis of DR. The system offers a possibility to support diagnostics by enabling users to upload a picture of a retina. The pictures are then broken down with the AI model to yield automated DR diagnosis on-the-spot. An early detection and involvement of the clinician would result in a timely response and thereby significantly lower the risk of blindness among diabetics. Further still, the integration with the Electronic Health Records (EHRs) ensures all AI-generated predictions are automatically added to the patient record. This eliminates the need for manual input and decreases the risk of errors and data entry burden, whilst enabling a smooth update of medical history per patient. By integrating AI-generated insight directly into clinical workflows, providers can map disease progression over time or compare AI-driven diagnoses to those made by humans, and help inform better decision-making in the treatment of patients. Finally, this system facilitates real-time DR screening for resource-poor settings. With edge computing techniques, the model is applicable on the handheld devices to assist offline and real-time detection of DR in the off-line area where it is hard to access to specialized ophthalmologists. The small size of the low-power AI inference models is capable of working without the need for persistent internet access, but instead downloading and processing the model on the device, therefore it is more suitable for rural clinics, mobile health camps and telemedicine facilities.

4.8 Ethical Considerations

The implementation of AI-based screening for diabetic retinopathy (DR) is bound by various ethical and regulatory frameworks that ensure patient safety as well as fairness and transparency. Chief among these is privacy in relation to patient data, making adherence to appropriate standards (such as GDPR in Europe and HIPAA in the US) a requirement. These laws safeguard privacy of medical data, particularly retinal images and other patient health records defining where it is stored, who has secured and encrypted access to the it and from which range of individuals have permission. Moreover, an array of privacy preserving data anonymity standards exists and is puffing to uphold no access, nor misuse in personal records which can potentially be supportive for promoting confidentiality and trust in AI diagnostics.

5 Results and Discussion

5.1 Comparative Analysis with Existing Solutions

To appraise the effectiveness of the crafted CNN model in use, a benchmarking exercise was conducted vis-a-vis some existing automated diabetic retinopathy detection solutions: EfficientNetB4, InceptionV3, and ResNet50. Such comparison utilized a subset of the key performance indicators: accuracy, sensitivity, specificity, and model complexity in table 1.

Table 1. Performance Comparison of Diabetic Retinopathy Detection Models.

Feature	Developed CNN Model	EfficientNetB 4	InceptionV3	Res Net50
Accuracy	95.00%	79.00%	76.80%	75.00%
Sensitivity	93.00%	78.00%	75.00%	73.00%
Specificity	94.00%	80.00%	77.00%	74.00%
Model Complexity	Moderate	High	High	High

The CNN model developed was found to have higher values of accuracy, sensitivity, and specificity as compared to the existing solutions. What's more, the complexity of this model is organized into a moderate form, thus rendering it more applicable in the context of resource-constrained deployment.

5.2 Quantitative Evaluation

The performance of the developed CNN model was further assessed using benchmark datasets such as the APTOS 2019 Blindness Detection dataset. Key performance metrics observed included accuracy, F1-score, and execution time.

The developed CNN model demonstrates high accuracy and F1-score, with efficient processing times, indicating its suitability for real-time diabetic retinopathy detection applications.

5.3 Real-Time Usability and Effectiveness

To gauge the usability of the developed CNN model, a survey was organized with 50 ophthalmologists and healthcare professionals. The feedback that was provided indicated:

- 88% of users found the application easier to use than traditional diagnostic methods.

- 80% reported a decrease in time needed for diagnosis by 25-35%.
- 90% appreciated the automatic detection, which made manual image analysis less necessary.

Such findings imply that the developed CNN model greatly lessens the burden on healthcare professionals and acts as a rapid and reliable device for diabetic retinopathy detection in real applications

5.4 Discussion

The outcomes validate the efficacy of the suggested DR detection model in precisely categorizing various severity levels. Balancing data effectively enhanced model performance by minimizing bias and increasing recall for minority classes. CNN-based deep learning models allowed for effective feature extraction, while data augmentation enhanced generalization.

The incorporation of explain ability methods, including Grad-CAM, offered transparency in AI-based predictions, with the model concentrating on salient retinal features. Small misclassifications were noted in borderline examples, indicating possible enhancements with more clinical information or sophisticated augmentation methods. In general, the model demonstrates high potential for practical clinical use, providing a robust and scalable AI-based solution for early DR detection and diagnosis.

6 Conclusion

The method propounded has high ease and convenience in DPS - Diabetic retinopathy detection for screening and grading of diabetic retinopathy image quality. It automatically uses an artificial intelligence-based approach. The widely-used dataset Kaggle APTOS 2019 Blindness Detection shows that the weighted kappa square score obtained is 0.92546. There was an excellent agreement with expert grading ophthalmologist. The model is not only sensitive and specificity enough to categorize stages of DR, but also bring it the usefulness for correcting detection of DR at the very early time. Hence, the model should receive proper credit in successfully detecting and classifying DR for doctors to act as soon as possible. Accordingly, this technique has the merit of insuch an automatic plugging-DR inspection that can reduce a synthetic work by means of eliminating human inspections and open up broad range large scale screening.

The deep learning model offers the advantage of layer building and syncing with respect to the complicated and interlacing fringing style characteristic of retina. Future enhancements could involve incorporation into a clinical setting, where real-time assessment occurs during routine eye exams. Besides, improvement of performance for other eye diseases and interpretability by the Explainable AI service would enable more opportunities to be a use in all classes diagnosis on the eye.

References

[1] T. Shahzad, M. Saleem, M.S. Farooq, S. Abbas, M.A. Khan, K. Ouahada, "Developing a Transparent Diagnosis Model for Diabetic Retinopathy Using Explainable AI," IEEE Access, vol. 12, pp. 149700–149709, 2024.

[2] V.C. Lee, S.C.L. Au, "Artificial Intelligence Systems for Diabetic Retinopathy Screening: Appraisal on the 3rd U.S. FDA Approved Algorithms-AEYE-DS," Journal of Ophthalmology Advances, 2023.

[3]] M.H. Wang, K.K.-L. Chong, Z. Lin, X. Yu, Y. Pan, "An Explainable Artificial Intelligence-Based Optimization Approach for Age-Related Macular Degeneration Detection," Electronics, vol. 12, no. 12, pp. 2697–2710, 2023.

[4] L. Balyen, "New Approaches in the Detection and Management of Diabetic Retinopathy in the Near Future," Advances in Ophthalmology and Vision Systems, vol. 13, no. 1, pp. 8–9, 2023.

[5] W. Liu, Z. Zhao, M. Levkiv, "Automatic Diagnosis of Diabetic Retinopathy Based on EfficientNet," in Proceedings of the 17th International Conference on Experimental Design and Applications of CAD Systems (CADSM), vol. 1, pp. 64–67, 2023.

[6] G. Quellec, H. Al Hajj, M. Lamard, P.-H. Conze, P. Massin, and B. Cochener, "ExplAIn: Explanatory artificial intelligence for diabetic retinopathy diagnosis," 2020. [Online].

[7] S.-I. Jang, M. J. A. Girard, and A. H. Thiery, "Explainable and interpretable diabetic retinopathy classification based on neural-symbolic learning," 2022. [Online].

[8] F. H. Yagin, C. Colak, A. Algarni, Y. Gormez, E. Guldogan, and L. P. Ardigò, "Hybrid Explainable Artificial Intelligence Models for Targeted Metabolomics Analysis of Diabetic Retinopathy," Diagnostics, vol. 14, no. 13, art. 1364, Jun. 2024, d oi: 10.3390/diagnostics14131364.

[9] Y. Niu, L. Gu, Y. Zhao and F. Lu, "Explainable Diabetic Retinopathy Detection and Retinal Image Generation," in IEEE Journal of Biomedical and Health Informatics, vol. 26, no. 1, pp. 44-55, Jan. 2022, doi: 10.1109/JBHI.2021.3110593.

[10] Islam MM, Rahman MJ, Rabby MS, Alam MJ, Pollob SMAI, Ahmed NAMF, Tawabunnahar M, Roy DC, Shin J, Maniruzzaman M. Predicting the risk of diabetic retinopathy using explainable machine learning algorithms. Diabetes Metab Syndr. 2023 Dec;17(12):102919. doi: 10.1016/j.dsx.2023.102919.

[11] J. de la Torre, A. Valls, and D. Puig, "A deep learning interpretable classifier for diabetic retinopathy disease grading," 2017. [Online].

[12] T. M. Devi, P. Karthikeyan, B. M. Kumar, and M. Manikandakumar, "Diabetic retinopathy detection via deep learning based dual features integrated classification model," International Journal of Engineering Science and Technology, 2024. [Online].

[13] Mara Santilli. (2025, February 21). The AI revolution is here — and it's coming to your eyes. PopSugar. Retrieved from https://www.popsugar.com/health/ai-eye-care-revolution-49429226

[14] Pillay, T. (2024, September 5). Zack Dvey-Aharon. Time Magazine. Retrieved from https://time.com/7012722/zack-dvey-aharon/

[15] Skouta, A., Elmoufidi, A., Jai-Andaloussi, S., & Ouchetto, O. (2023). Deep learning for diabetic retinopathy assessments: A literature review. Multimedia Tools and Applications. https://doi.org/10.1007/s11042-023-15110-9

[16] "Explainable Diabetic Retinopathy Detection Using EnlightenGAN Images with Deep Learning Technique" authored by M. Kannan, Nikitha S., Thouqeer Baig, Akshay S., and Adwitiya, published in 2025

[17] Jang, S.-I., Girard, M. J. A., & Thiery, A. H. (2022). Explainable and interpretable diabetic retinopathy classification based on neural-symbolic learning [Preprint]. arXiv. https://arxiv.org/abs/2204.00624

[18] Romero-Oraá, R., Herrero-Tudela, M., López, M. I., Hornero, R., & García, M. (2024, June). Attention-based deep learning framework for automatic fundus image processing to aid in diabetic retinopathy grading. Computers in Biology and Medicine. https://doi.org/10.1016/j.cmpb.2024.108160

[19] Quellec, G., Al Hajj, H., Lamard, M., Conze, P.-H., Massin, P., & Cochener, B. (2020, August 13). ExplAIn: Explanatory artificial intelligence for diabetic retinopathy diagnosis [Preprint]. arXiv. https://arxiv.org/abs/2008.05731

[20] Shen, Z., Jiang, X., & Huang, X. (2023). Deep learning-based interpretable detection method for fundus diseases. In Proceedings of the 3rd International Conference on Bioinformatics and Intelligent Computing (pp. 305–309). ACM. https://doi.org/10.1145/3592686.3592741