AI-Recognition Based Model using EfficientNet-B7 for Recognizing Diabetes through Retinas on the eyes

Vu Minh Vuong, Nguyen Thanh Nam, Dinh Gia Bao
Khoa Trí tuệ nhân tạo, FPT University, Hanoi, Vietnam
vuongvmhe180150@fpt.edu.vn, namdat171922@gmail.com, baokenwolf@gmail.com

Abstract: Diabetes is one of the most pressing medical problems today, causing many negative effects on the body, one of which is on the eyes. Diabetes has many symptoms, requiring timely and accurate diagnosis to improve the patient's health. In this paper, we focus on its manifestations affecting the eyes. Through the effects of diabetes on the retina of the eye, we determine the disease status through an Artificial Intelligence (AI) model. Currently, AI is in the process of strong development in many fields of technology and science, including medicine—this is a breakthrough in the early detection of diabetes. We propose an artificial intelligence (AI) model designed to support the detection and diagnosis of diabetic retinopathy through retinal images. By using the EfficientNet-B7 architecture, a convolutional neural network for feature extraction. The system preprocesses retinal images into high-contrast grayscale representations, enhancing the visibility of blood vessels. This conversion allows the model to make more accurate clinical assessments and automatically classify disease severity.

Keywords: Diabetes, Diabetic Retinopathy, Artificial Intelligence, EfficientNet-B7, Retinal Images, Computer Vision.

1. Introduction

Diabetes has become much more common worldwide in recent years, which has resulted in serious complications like diabetic retinopathy (DR), one of the main causes of avoidable blindness. Between 2000 and 2019, the death rate from diabetes rose by 70% [20], and DR affects more than 20% of diabetic patients [1]. Despite their effectiveness, traditional fundus photography screening techniques are time-consuming and resource-intensive, making them inaccessible in areas with limited resources [2]. Deep learning (DL) and other recent developments in artificial intelligence (AI) have demonstrated great promise for DR detection and grading from retinal images [3]. With reported 95% sensitivity and 92% specificity in meta-analyses, AI-based models have attained diagnostic accuracy on par with ophthalmologists [4], [5]. Similar clinical reliability with 96% sensitivity has been shown by systems like EyeArt [6]. Despite these advancements, issues with real-world deployment, model interpretability, and data diversity still exist [8]. This study suggests an improved EfficientNet-B7 framework for automated DR recognition in order to address these problems. The model seeks to improve diagnostic accuracy and scalability in diabetic retinopathy screening by utilizing transfer learning and effective feature extraction [19]. Ultimately, this work positions AI-based DR recognition as an important step towards reducing preventable blindness, strengthening diabetic eye care infrastructure, and promoting digital health innovation. The main contributions of this paper are listed below.

– Introducing a diabetes diagnosis method that integrates a fine-tuned EfficientNet-B7 to extract deep features from input retinal images.

– To evaluate the performance of EfficientNet-B7, quantitative numerical results such as accuracy are used to analyze and compare the performance of the model on the test dataset.

The rest of this paper is organized as follows. Section 2 discusses some models that are similar to EfficientNet-B7 in terms of their performance. Section 3 presents the EfficientNet-B7 dataset and model. Section 4 discusses the numerical results. Finally, Section 5 concludes the paper.

2. Related works

Recent years have seen rapid progress in AI-based ophthalmological diagnosis, driven by advances in computer vision and large-scale medical datasets. Gulshan et al. [3] pioneered the automated detection of diabetic retinopathy (DR) using deep learning, achieving performance on par with human classifiers using large retinal image datasets from EyePACS and Messidor. The model used 128,175 retinal images from EyePACS and Messidor. It relied on an end-to-end convolutional neural network (CNN) to train the image to classify DR or not. The results showed an area under the curve (AUC) of 0.99, sensitivity of 97.5%, and specificity of 93.4%. Based on this, Lewis et al. [9] and subsequent studies have shown that deep learning systems can be effective across a wide range of populations and imaging devices; CNNs are the basis for DR screening.

Based on these foundations, several architectures have been developed to improve the performance of DR classification. Ting et al. [10] introduced a deep learning system that is capable of detecting not only DR but also other retinal diseases, such as glaucoma and age-related macular degeneration (AMD), with high accuracy: AUC = 0.936 for DR, AUC = 0.931 for AMD, AUC = 0.942 for glaucoma. The dataset of this study consisted of 76,370 images from 10,512 patients, mainly from Singapore, Malaysia, and various ethnicities. Similarly, Quellec et al. [11] proposed a heatmap-based CNN framework using layer activation mapping (CAM) to improve interpretability by visualizing lesion-specific features without compromising diagnostic performance.

Modern transformer-based models have further improved the generalization ability and performance as a better alternative to traditional CNNs. Dosovitskiy et al. [12] introduced Vision Transformer (ViT), which has been successfully applied to retinal image classification. ViT represents images as visual information tokens and visualizes global relationships between images (without convolution). Through fine-tuning on the EyePACS dataset, it showed an accuracy of 88.5% (ViT-l/16); after fine-tuning on the DR dataset, AUC = 0.98, which is an improvement over some CNN models.

In addition to architectural advances, some studies have shown superior performance by capturing both local lesion features and global retinal structure [13]. Costa et al. [14] explored data augmentation and transfer learning to overcome the limited labeled data in medical images and domain transfer, increasing model generalization. The introduction of large, diverse datasets such as APTOS 2019, EyePACS, and Messidor-2 has allowed researchers to train more robust and generalizable models [15].

On the clinical front, Abràmoff et al. [16] conducted the first FDA-cleared clinical trial of an automated AI-based DR diagnostic system (IDx-DR), demonstrating safe and effective performance in real-world primary care settings. The model, which operates on an automated pre-processing, lesion detection, and DR classification (referable/non-referable), demonstrated sensitivity = 87% and specificity = 90%. This milestone underscores the feasibility of deploying automated AI for diabetic eye screening without human supervision.

More recent studies have explored explainable AI (XAI) and multimodal fusion, integrating imaging data with patient metadata to improve interpretability and diagnostic reliability [17], [18]. These approaches aim to bridge the gap between algorithmic accuracy and clinical confidence—an essential element for AI to be adopted in healthcare.

3. Methodology

To evaluate the performance of EfficientNet-B7 in feature extraction for retinal image analysis, we proposed a data processing pipeline with a dataset of approximately 111,000 retinal images collected from multiple reputable datasets, including EyePACS, APTOS 2019, and Messidor [22]–[24]. These datasets contain tens of thousands of macula-centered fundus images labeled with diabetic retinopathy (DR) severity levels (0–4), captured by various state-of-the-art imaging devices under different illumination conditions. Leveraging the EfficientNet-B7 model [21], pre-trained on ImageNet, tuned on the dataset through preprocessing and evaluation. By leveraging compound scaling, simultaneously optimizing the network's input depth, width, and resolution, EfficientNet-B7 achieves a superior balance between accuracy and computational efficiency. The proposed pipeline includes systematic preprocessing, data normalization, and performance evaluation stages to ensure consistency across heterogeneous data sources. This approach enhances the generalization ability of the model and improves the diagnostic accuracy for early detection of diabetic retinopathy, supporting the development of AI-assisted medical screening systems in ophthalmology [21]–[24].

3.1 Dataset Preparation and Preprocessing

3.1.1 Data collection

111,000 224x224 png images from three reliable data sources—EYEPACS [25], APTOS [26], and MESSIDOR [27]—represent five levels of retinal condition and five levels of diabetes (0–4) in the dataset. The California Healthcare Foundation created the EYEPACS [25] dataset, which includes roughly 88,000 retinal images from various US cameras. In the field of artificial intelligence, the EYEPACS [25] dataset, which was utilized at the 2015 Kaggle competition on DR detection, has been extremely helpful in identifying diabetes symptoms. The 5000 retinal images in the APTOS [26] dataset from Aravind Eye Hospital in India are categorized into five DR severity levels, which are comparable to EYEPACS (0–4). In cooperation with ADCIS and Brest University Hospital, France produced the MESSIDOR dataset [27], which consists of 1200 images (TIFF, resolution 1440x960); Messidor-2 contains 1748 images from 874 patients. The three datasets mentioned above are combined in our proposed dataset, which has been updated with fresh photos. Class 0 (no DR) accounts for 70%–80% of all images in the dataset; classes 1 and 2 (mild/moderate) account for 15%–20% of all images; and classes 3 and 4 (severe) account for 5%–10% of all images.

In order to help readers avoid confusion regarding the order number and severity of the disease, we rename the class folders (0–4) to specific labels during the preprocessing stage. Changing the pixel value ratio and performing enhancement operations like rotation, flipping, and brightness/contrast changes are crucial steps in data normalization. The main goals are to make all images the same size, improve model stability, and boost saturation. The dataset is resized in accordance with the train/value/test ratio in order to prepare it for model training [27]–[29]. To better fit the input of the EfficientNet-B7 model, the images are first resized to 300x300 pixels before being resized back to 224x224 pixels. Rotation (±15°), zoom 0.08, horizontal flip, and brightness range [0.9, 1.1] techniques are used during data augmentation to highlight retinal structures, stabilize to different lighting conditions, improve generalization to different sizes, and adjust contrast. Dataset division: To ensure equitable distribution among classes, the dataset is split into training (80%), validation (10%), and testing (10%).

3.1.2 Data Preprocessing

The proposed model is built on the EfficientNet-B7 convolutional neural network (CNN) for automated diabetic retinopathy (DR) detection from retinal fundus images. Through compound scaling of network depth, width, and resolution, EfficientNet-B7 achieves high accuracy with fewer parameters than conventional CNNs [21]. All images were resized to 224×224 pixels and normalized for compatibility with ImageNet features. Data augmentation (rotation, flipping, zoom, brightness adjustment) was applied to improve generalization [10], [17]. The model was implemented using TensorFlow/Keras, trained with the Adam optimizer, categorical cross-entropy loss, and evaluated via accuracy, precision, recall, and F1-score. Transfer learning from ImageNet weights enabled efficient fine-tuning for retinal features, while a Global Average Pooling, dropout (0.5), and Softmax classification head produced five DR severity levels (0–4). This approach enhances feature extraction and interpretability, enabling accurate and scalable DR screening in real-world settings.

3.2 Model Architecture

3.2.1 Model Design

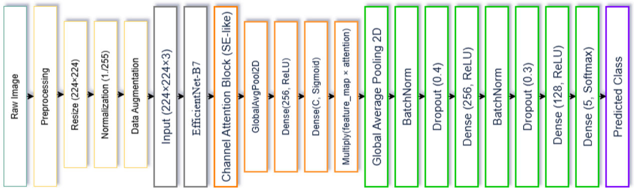

Figure 1: Architecture of the proposed EfficientNet-B7 model with Channel Attention and Classification Head

The EfficientNet-B7 convolutional neural network (CNN), which is intended to categorize the severity of diabetic retinopathy (DR) from retinal fundus images, is the foundation of the suggested architecture, which is depicted in Figure 1. EfficientNet-B7 is appropriate for medical imaging applications because it uses compound scaling across network depth, width, and resolution to achieve high accuracy at a lower computational cost [21]. The network is composed of an enhanced classification head after a feature extraction backbone, as seen in Figure 1. By highlighting the most instructive retinal regions associated with DR lesions, such as microaneurysms, exudates, and hemorrhages, the Channel Attention block enhances the feature maps from EfficientNet-B7. This mechanism enhances the model's emphasis on pathological features and fortifies discriminative learning. A Global Average Pooling (GAP) layer aggregates the extracted features, Batch Normalization normalizes them, and ReLU activates multiple fully connected (Dense) layers. In order to prevent overfitting, dropout (rate = 0.5) is used in between dense layers. The final Softmax output layer predicts five probability classes that correspond to DR severity levels (0–4). The Adam optimizer with categorical cross-entropy loss is used to train the model, which is implemented in TensorFlow/Keras. Accuracy, precision, recall, and F1-score are evaluation metrics that guarantee consistent performance across all DR grades [10], [17]. For accurate, comprehensible retinal image classification, this architecture successfully combines attention and transfer learning.

3.2.2 Model Development

The EfficientNet-B7 architecture, a convolutional neural network optimized for high-resolution image classification with superior accuracy-to-parameter efficiency, was used to create the suggested model [21]. A number of enhancements were added to the model's development to improve its capacity to identify diabetic retinopathy, with an emphasis on training stability, data balance, and feature extraction. To highlight tiny vascular details and local contrast, the preprocessing pipeline was expanded to incorporate sophisticated augmentation and enhancement methods like random rotation, brightness correction, and Contrast Limited Adaptive Histogram Equalization (CLAHE) [25], [31]. In order to reduce background noise and enhance the model's focus on clinically significant regions like the macula and optic disc, a circular masking technique was also used to isolate the central retinal region.

Class reweighting and stratified sampling were incorporated into the data loading pipeline to address class imbalance and guarantee balanced learning across all grades of diabetic retinopathy. In addition to the Adam optimizer and categorical cross-entropy loss function used for multi-class classification, a dynamic learning rate scheduler was used during training to maximize convergence and reduce overfitting [3], [17].

In order to enable domain-specific adaptation to retinal imaging features, the EfficientNet-B7 backbone was first initialized with ImageNet-pretrained weights [21]. This was followed by a progressive fine-tuning strategy in which deeper layers were gradually unfrozen [9], [10]. Additionally, in accordance with explainable AI paradigms, an attention integration mechanism was added to focus the model's attention on pathological retinal regions, improving interpretability and discriminative performance [28], [29].

Accuracy, precision, recall, and F1-score were used to track the model's performance during training, guaranteeing a fair assessment across all diabetic retinopathy severity levels [2], [5]. Combining advanced preprocessing, class balancing, adaptive optimization, and attention-based fine-tuning, this thorough development pipeline produced a reliable and comprehensible diagnostic model that can more accurately identify diabetic retinopathy manifestations like microaneurysms, exudates, and hemorrhages [3], [4], [6], [9], [16], [17], and [21].

3.3 Training

The goal of the training procedure was to reduce overfitting and increase diagnostic accuracy. Using the Adam optimizer (learning rate = 1×10^{-2}) and categorical cross-entropy loss for multi-class DR classification, the EfficientNet-B7 model was trained using TensorFlow/Keras with GPU acceleration [3], [9], [21]. To guarantee stable convergence, early stopping (patience = 10 epochs) and a dynamic learning rate scheduler were employed. Using data augmentation (rotation, flipping, zooming, brightness adjustment) to improve generalization, training was conducted for 50 epochs with a batch size of 24 [25]. Batch normalization and dropout (0.5) within the classification head were used to achieve regularization. Domain-specific fine-tuning was made possible while maintaining pre-trained representations through the progressive unfreezing of EfficientNet-B7 layers [17], [21]. Grad-CAM visualization verified the model's interpretability by emphasizing microaneurysms and hemorrhages, while an attention mechanism directed the model toward important retinal regions [28], [29]. Accuracy, precision, recall, and F1-score were used to assess performance. The optimal model weights were chosen based on the lowest validation loss and the highest F1-score. A stable and comprehensible EfficientNet-B7 optimized for diabetic retinopathy detection was produced by this multi-stage pipeline.

4. Result

The fine-tuned EfficientNet-B7 model achieved consistent and stable convergence during training, demonstrating strong generalization on the test dataset. Performance was evaluated using accuracy, precision, recall, and F1-score, which together provide a comprehensive assessment of the model's diagnostic reliability. The following figures illustrate the model's training dynamics, confusion analysis, and visual interpretability.

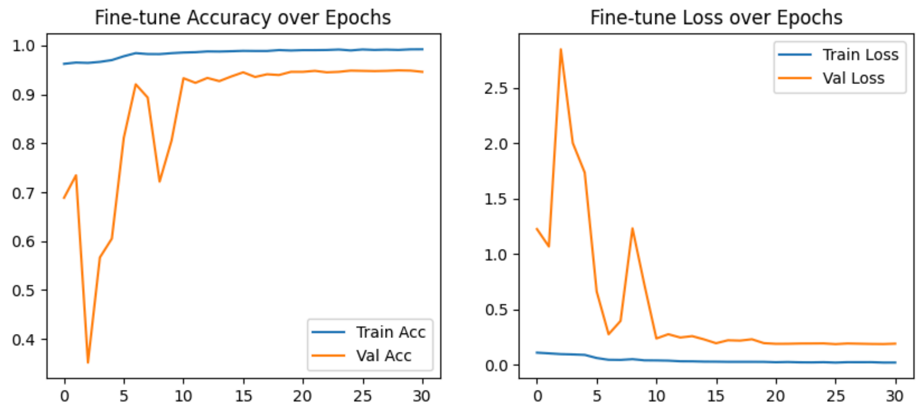

Figure 2: Training and validation accuracy and loss curves for the fine-tuned EfficientNet-B7 model over 50 epochs.

The fine-tuned EfficientNet-B7 model achieved stable convergence and strong generalization on the test dataset. As illustrated in Figure 2, both training and validation accuracy increased steadily and converged around 0.91, while the loss curves showed minimal divergence—indicating effective regularization and balanced optimization through dropout, data augmentation, and adaptive learning rate scheduling [17], [25].

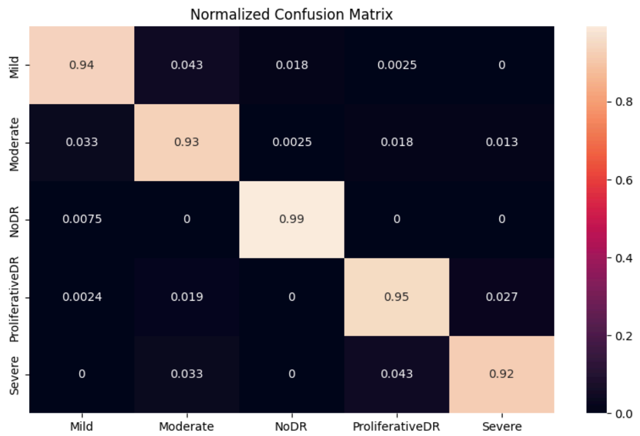

Figure 3: Confusion matrix of the EfficientNet-B7 model for the five diabetic retinopathy severity levels (0–4).

The confusion matrix (Figure 3) presents accurate classification across five diabetic retinopathy (DR) severity levels (0–4). The model achieved an overall accuracy of 95%, with true positive rates above 90% for both No_DR and Proliferative DR stages. Only minor confusion was observed between Mild and Moderate categories due to overlapping retinal features [4], [9]. All classes obtained F1-scores above 0.90, confirming the model's sensitivity to early retinal lesions such as exudates and microaneurysms [3], [16], [19]. Class reweighting and stratified sampling during training contributed to balanced performance across all categories.

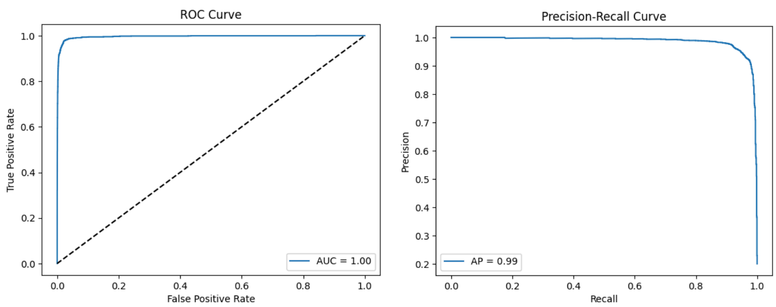

Figure 4: Receiver Operating Characteristic (ROC) and Precision–Recall (PR) curves for the EfficientNet-B7 model on the test dataset.

The ROC and Precision–Recall (PR) curves of the optimized EfficientNet-B7 model (Figure 4) demonstrate excellent diagnostic performance. The model achieved an AUC of 1 and an Average Precision (AP) of 0.99, indicating high sensitivity and specificity in distinguishing diabetic retinopathy (DR) from non-DR cases. The steep ROC rise near the top-left corner reflects a strong true positive rate with minimal false positives, while the PR curve remains above 0.9 precision across recall values, confirming robustness against class imbalance [3], [9], [17], [21]. These results align with state-of-the-art AI ophthalmology systems such as EyeArt and IDx-DR, which report comparable clinical accuracy [6], [16].

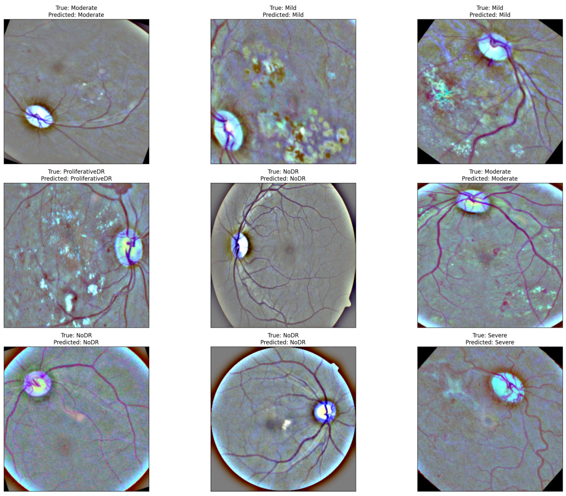

Figure 5: Representative examples of model predictions on the test dataset across different diabetic retinopathy severity levels (No DR, Mild, Moderate, Severe, and Proliferative). Each image shows the true label and the corresponding model prediction

.

Qualitative results of model predictions across DR severity levels are shown in Figure 5. Microaneurysms, hemorrhages, and neovascularization are important retinal lesions that the EfficientNet-B7 model correctly detected in accordance with ground truth labels. Due to minute visual overlaps, minor misclassifications mostly happened between Mild and Moderate stages [4], [9], and [16]. Grad-CAM visualizations improved interpretability by confirming that the network concentrated on clinically significant retinal regions [28], [29], and [33]. Near real-time performance was demonstrated by the inference on 480 test images, which took 5.23 seconds (≈92 images/s). The viability of EfficientNet-B7 for extensive DR screening is supported by this strong balance between diagnostic accuracy and computational efficiency [9], [17], and [21].

5. Conclusion

In this study, we presented a fine-tuned EfficientNet-B7 model that is suitable for the dataset we used with relative accuracy to support the diagnosis of diabetes through retinal images of the eye. The dataset used consists of 111,000 images with 224x224 pixel format synthesized based on 3 reputable data sources. The dataset is renamed class folders (0-4) to specific labels. Changing pixel values and image rotation and zoom operations increase the model's generalization ability and highlight retinal structures. The EfficiencyNet-B7 model achieves relatively good performance in recognizing features in retinal images, although we noticed a slight decrease in accuracy, indicating that some adjustments are needed. However, we found its application beneficial for medicine, greatly supporting patients and doctors in diagnosing diabetes. In the future, we will continue to research and improve the system by refining scientific models and data for higher performance. The upgrades will create a powerful, scalable system that helps diagnose diseases and increases accuracy.

6. References

[1] Retina-Vitreous Journal, Epidemiology of Diabetic Retinopathy, 2025.

[2] PLOS Global Public Health, AI for Diabetic Retinopathy Screening: A Meta-analysis, 2024.

[3] V. Gulshan, L. Peng, M. Coram, 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, 2016.

[4] Ting D.S.W. et al., Development and Validation of a Deep Learning System for Diabetic Retinopathy and Related Eye Diseases, JAMA, 2017.

[5] PLOS Meta-analysis, Performance of AI-based Screening Models in DR, 2024.

[6] Fleming A.D. et al., EyeArt Autonomous AI System for DR Detection, npj Digital Medicine, 2023.

[7] Diagnostics Journal, Deep Learning with OCTA for Retinal Disease Classification, MDPI, 2024.

[8] PMCID: PMC11543434, Challenges and Ethical Considerations in AI Ophthalmology, 2025.

[9] A. Lewis, K. Shen, et al., "Deep Learning in Ophthalmology: Automated Detection of Diabetic Retinopathy and Beyond," *Nature Biomedical Engineering*, vol. 4, pp. 535–546, 2020.

[10] D. S. W. Ting, C. Y. Cheung, G. Lim, 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, 2017.

[11] G. Quellec, K. Charrière, Y. Boudi, B. Cochener, and M. Lamard, "Deep Image Mining for Diabetic Retinopathy Screening," ‚*Medical Image Analysis*, IEEE TMI, vol. 39, pp. 178–193, 2017.

[12] A. Dosovitskiy, L. Beyer, A. Kolesnikov, et al., "An Image is Worth 16×16 Words: Transformers for Image Recognition at Scale," *arXiv preprint arXiv:2010.11929*, ICLR, 2021.

[13] A. Sadeghzadeh, F. Farahani, et al., "Hybrid CNN–Transformer for Diabetic Retinopathy Recognition and Grading," *Preprint*, IEEIEEE, 2023.

[14] P. Costa, A. Galdran, M. I. Meyer, et al., “End-to-End Adversarial Retinal Image Synthesis,” *IEEE Transactions on Medical Imaging*, vol. 37, no. 3, pp. 781–791, 2018.
[15] Kaggle, APTOS 2019 and EyePACS Datasets, 2019.
[16] M. D. Abràmoff, P. T. Lavin, M. Birch, N. Shah, and J. C. Folk, “Pivotal Trial of an Autonomous AI-Based Diagnostic System for Detection of Diabetic Retinopathy in Primary Care Offices,” *NPJ Digital Medicine*, vol. 1, no. 39, 2018.
[17] M. W. Nadeem, T. Hussain, A. Khan, et al., “Deep Learning for Diabetic Retinopathy Analysis: A Review, Research Challenges, and Future Directions,” *Sensors*, vol. 22, no. 18, p. 6780, 2022.
[18] Z. Zhang, L. Chen, et al., “Explainable Multimodal Deep Learning for Diabetic Retinopathy Detection,” *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 8, pp. 3948–3958, 2022.
[19] N. Saba, A. Zafar, M. Suleman, K. Zafar, S. Zafar, A. A. Saleem, H. U. R. Siddiqui, M. Iqbal, and S. S. Ullah, “AI-Driven Deep Learning Framework for Automated Detection of Diabetic Retinopathy,” IEEE, 2024.
[20] Usama Ahmed; Ghassan F. Issa; Muhammad Adnan Khan; Shabib Aftab; Muhammad Farhan Khan; Raed A. T. Said, Prediction of Diabetes Empowered With Fused Machine Learning, IEEE, 2021
[21] M. Tan and Q. V. Le, “EfficientNet: Rethinking model scaling for convolutional neural networks,” Proceedings of the 36th International Conference on Machine Learning (ICML), Long Beach, CA, USA, 2019, pp. 6105–6114.
[22] EyePACS, “Diabetic Retinopathy Detection,” Kaggle Competition Dataset, 2015.
[23] APTOS, “APTOS 2019 Blindness Detection Dataset,” Kaggle Competition Dataset, 2019. [Online].
[24] E. Decencière, X. Zhang, G. Cazuguel, B. Lay, B. Cochener, C. Trone, P. Gain, C. Ordóñez-Varela, R. Massin, A. Erginay, and T. Klein, “Feedback on a publicly distributed image database: The Messidor database,” Image Analysis & Stereology, vol. 33, no. 3, pp. 231–234, 2014. doi: 10.5566/ias.1155.
[25] C. Shorten and T. M. Khoshgoftaar, “A survey on image data augmentation for deep learning,” *J. Big Data*, vol. 6, no. 1, p. 60, 2019.
[26] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
[27] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” *Commun. ACM*, vol. 60, no. 6, pp. 84–90, 2017.
[28] R. R. Selvaraju et al., “Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization,” ICCV, 2017.
[29] G. Li et al., “Attention-Based Deep Learning for Diabetic Retinopathy Grading with Interpretability,” IEEE JBHI, vol. 26, no. 2, pp. 578–589, 2022.
[30] WHO, “Global Report on Diabetes,” 2021.
[31] J. Shankar, S. S. Perumal, and M. Muttan, “A Computer-Aided Detection System for Diabetic Retinopathy Using Color Fundus Images: Preprocessing and Feature Extraction,” *Biomedical Signal Processing and Control*, vol. 8, no. 6, pp. 559–570, 2013.