International Journal of Electrical and Electronics Engineers
Volume 18, Issue No. 02, July-December 2026
ISSN (O) 2321-2055
ISSN (P) 2321-2045

Vision Transformer Framework for Diabetic Retinopathy Detection

G. Arunalatha

Assistant Professor, Department of Computer Science and Engineering,
Perunthalaivar Kamarajar Institute of Engineering and Technology (PKIET),
Karaikal. Puducherry.
Email: vigneshgayu1121@gmail.com

Abstract:

Diabetic Retinopathy (DR) is a diabetes-related eye disease caused by prolonged high blood glucose levels, which damage the blood vessels in the retina and may lead to vision loss. Early detection is essential to prevent disease progression and blindness. This study focuses on developing an automated diabetic retinopathy detection system using a Vision Transformer (ViT)-based deep learning approach. The proposed system analyses retinal fundus images and classifies them. Image preprocessing techniques are applied to improve image quality, while the Vision Transformer uses a self-attention mechanism to extract important retinal features and identify abnormalities such as microaneurysms, haemorrhages, and exudates. The model performance is evaluated using accuracy, precision, recall, and F1-score. This approach aims to improve diagnostic accuracy, support early detection, reduce manual screening efforts, and assist ophthalmologists in preventing diabetes-related vision loss.

Keywords: *Fundus, Retinography, NPDR, CNN.*

a. Introduction

The fundus describes the inside of the back of the eye. It includes the retina, optic disc, macula, fovea, choroid and retinal vessels. It is an important structure involved in vision as it absorbs light that enters the eye and converts it to electrical signals which are then transmitted to the brain via the optic nerve. The retina is a sensitive layer that lines the inside of the eye and detects light, colour and fine detail. The optic disc is where the optic nerve exits the eye. The macula is the centre part of the retina for detailed vision. The centre of the macula is called the fovea and is where our vision is most clear. The retinal vessels are important to deliver oxygen and nutrients to the retina, so the retina functions properly. Fundus testing is an important way to look inside the eye. It can be done using an ophthalmoscope or fundus camera.

Fundus cameras provide non-invasive high quality colour images of the retina. They allow health professionals to identify and monitor different diseases affecting the retina. Fundus cameras are used widely to both diagnose and screen for a number of eye conditions such as diabetic retinopathy, hypertensive retinopathy, glaucoma, age related macular degeneration, retinal haemorrhage and macular oedema. Fundus imaging can be used to detect diabetic retinopathy at an early stage. Diabetic retinopathy is caused by damage to the retinal vessels due to prolonged high blood sugar levels. Fundus imaging allows us to detect characteristic changes in

International Journal of Electrical and Electronics Engineers
Volume 18, Issue No. 02, July-December 2026
ISSN (O) 2321-2055
ISSN (P) 2321-2045

the retina such as microaneurysms, haemorrhages, hard exudates, cotton wool spots and new blood vessel formation. Identifying these characteristics can help us determine how severe the disease is, monitor progression and provide early treatment to avoid vision loss. Digital fundus images can be analysed using computer aided diagnostic software and deep learning technology. This means we can automatically detect and grade different retinal diseases. This improves screening accuracy, reduces the pressure on ophthalmologists and provides early diagnosis. In rural areas or areas with less access to specialists this can be very beneficial. Fundus imaging is a vital tool to assess the retina, screen for disease, make a diagnosis and reduce preventable blindness.

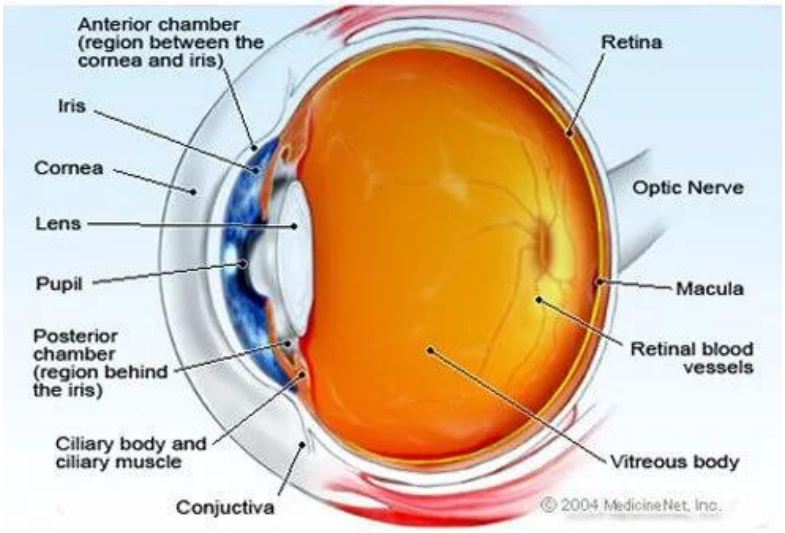


Figure1: Human Eye

a. Human eye structure:

The retina, a delicate membrane lining the posterior interior of the human eye, constitutes one of the ocular system's most vital components, serving as the innermost, light-sensitive stratum essential for visual acuity. Much like photographic film within a camera, this tissue functions by receiving light that penetrates the pupil and converting it into electrical signals transmitted via the optic nerve to the brain. Comprising millions of microscopic photoreceptor cells known as rods and cones, the retina translates incident light into neural impulses for cerebral processing; while rods exhibit heightened sensitivity to variations in light and darkness, cones are specifically tuned to detect color. Collaborating to facilitate both contrast and color perception are the rods and cones. Situated centrally within the retina, the macula governs sharp, central visual acuity. As the eye's most critical component, the retina remains inherently vulnerable to a diverse array of pathologies

b. Diabetic Retinogaphy:

The retina is the light-sensitive tissue situated at the back of the eye, essential for vision. To function optimally, it requires a steady supply of oxygen and nutrients from the surrounding blood vessels. Damage to the retina can impair visual ability and lead to vision loss. Common retinal disorders include retinal detachment, diabetic retinopathy, and age-related macular degeneration. Retinal detachment occurs when the retina separates from its normal position, disrupting its oxygen and nutrient supply, potentially harming retinal cells. This condition

represents a medical emergency that requires immediate treatment. Common symptoms include a sudden increase in floaters, flashes of light, and a shadow or dark veil covering parts of the visual field.

Diabetic retinopathy occurs when consistently high blood sugar levels damage the small blood vessels in the retina. These vessels may swell, leak fluid or blood, become blocked, or stimulate the formation of abnormal new vessels. These changes can cause retinal hemorrhages, exudates, macular edema, and neovascularization, leading to blurry vision or severe vision loss. Age-related macular degeneration affects the macula, the central part of the retina responsible for sharp, detailed vision. It leads to progressive damage to the macula, resulting in blurred central vision and blind spots, particularly in older adults. While it can significantly diminish central vision, it rarely causes complete blindness.

Chronic high blood sugar can gradually damage eye structures, resulting in retinal issues like hemorrhages (small bleeding areas), exudates (deposits from leaking blood vessels), and neovascularization (abnormal new blood vessel growth). Diabetic retinopathy is a vision-threatening condition due to prolonged hyperglycemia. Damage to the retinal blood vessels can allow fluid and blood to seep into the retina, causing swelling and blurry or cloudy vision. High blood sugar levels can also alter the fluid balance and shape of the lens, leading to temporary vision changes. These lens-related fluctuations may improve with normalized blood sugar; however, damage to the retinal blood vessels is often irreversible. In advanced stages, fragile new blood vessels can form on the retinal surface, rupture, and bleed, further impairing vision and creating complications.

Diabetic retinopathy can affect both type 1 and type 2 diabetes patients, with the risk heightening due to poor blood sugar control and longer diabetes duration. Other risk factors include high blood pressure, elevated cholesterol, a family history of diabetic retinopathy, and certain ethnic groups, notably African American and Hispanic/Latino populations. Pregnancy can worsen existing diabetic retinopathy in those with pre-existing diabetes, while gestational diabetes increases the likelihood of developing type 2 diabetes later in life. Maintaining blood sugar levels within target ranges, controlling blood pressure and cholesterol, scheduling regular eye exams, and seeking timely treatment can contribute to preventing or delaying progression and reducing the risk of vision loss. Diabetic retinopathy (DR) is a leading cause of blindness in working-age individuals and usually affects both eyes. It has two main stages: non-proliferative diabetic retinopathy (NPDR) and proliferative diabetic retinopathy (PDR).

NPDR is the early stage characterized by damaged retinal blood vessels that may leak, become blocked, or close entirely. Leakage can lead to retinal swelling, including macular edema, a common cause of vision loss in diabetics. Other changes include exudates and macular ischemia, which occurs when blood flow to the macula is reduced or blocked, compromising central vision. PDR is the later stage, marked by the emergence of abnormal new blood vessels on the retina. These fragile vessels can bleed into the vitreous, the gel-like substance inside the eye, resulting in dark floaters or significant vision loss. Associated scar tissue can pull on the retina, damage the macula, or trigger retinal detachment. PDR can threaten both central and peripheral vision. The likelihood of developing and advancing through these stages rises with longer diabetes duration and consistently high blood glucose

International Journal of Electrical and Electronics Engineers
Volume 18, Issue No. 02, July-December 2026
ISSN (O) 2321-2055
ISSN (P) 2321-2045

2. Literature Survey:

In[1], neural networks were used to detect diabetic features from the fundus images using a set of fundus images. In this test, 147 diabetic and 32 normal images were captured from a fundus camera, and analyzed using a back propagation neural network. The network was trained using digital filtering techniques and different network variables to recognize features in the retinal image. The system was assessed by selecting 200 diabetic and 101 normal images randomly to evaluate the network's performance for the detection of diabetic retinopathy against an ophthalmologist.

With the advancements in machine learning [2], the prediction of eye diseases was performed based using deep convolutional neural networks(DCNNs) on fundus photography. The lightweight deep learning architecture called MobileNetV2 was utilized for learning to distinguish four common eye diseases such as Glaucoma, Pathological Myopia, Maculopathy and Retinitis Pigmentosa.

Retinal diseases can lead to irreversible visual impairment, if they were not diagnosed in earlier stage. Single disease-based deep learning algorithms had been developed for the detection of diabetic retinopathy, age-related macular degeneration, and glaucoma. A deep learning platform (DLP) capable of detecting multiple common referable fundus diseases was developed for conditions (39 classes) by using 249,620 fundus images marked with 275,543 labels from heterogenous sources.

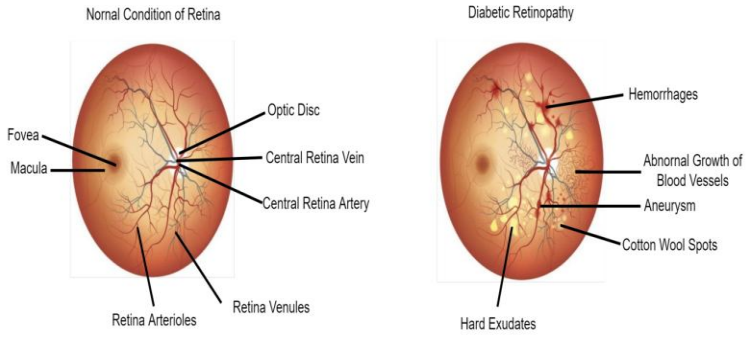


Figure 2. Illustration of the entire retina image. The fundus image on the left depicts the normal retina condition, and the fundus image on the right depicts a retina with diabetic retinopathy symptoms.

Researchers have found that nonproliferative diabetic retinopathy (NPDR) was present in 25% of patients 5 years after they were diagnosed with diabetes, 60% at 10 years, and 80% at 15 years. Patients with NPDR generally present with hemorrhages of varying sizes, microaneurysms (MAs), hard exudates, soft exudates (cotton wool spots) intraretinal microvascular abnormalities (IRMAs), and venous looping or beading. STAGE 1: MILD NPDR: These patients have at least one MA but no other findings. stage 2: MODERATE NPDR: These patients have hemorrhages or MAs in one to three retinal quadrants and/or cotton wool spots, hard exudates, or venous beading. STAGE 3: SEVERE NPDR; These patients have intraretinal hemorrhages (> 20 in each quadrant), venous beading in two or more quadrants, or an IRMA in one or more quadrants. STAGE 4: PROLIFERATIVE DIABETIC RETINOPATHY: These patients had NPDR that has progressed to PDR, and they exhibit either neovascularization of the disc/elsewhere or vitreous/preretinal hemorrhage.

A deep learning system to identify referable diabetic retinopathy was presented[4] in collaboration with Digifundus Ltd, an ISO 9001:2015 certified provider of diabetic retinopathy screening and monitoring services

in Finland. Digifundus Ltd provided anonymized retinal image dataset of 14624 patients with diabetes, including 41122 graded retinal color images. By inducing mydriasis with tropicamide 5 mg/ml eye drops, the fundus images were taken at Two 45 degree, centered on fovea and optic disc with Canon CR2 retinal camera from the patient's both eyes.

Smartphone-based fundus imaging (SBFI) [5] allows mobile fundus examination for low cost using an adapter on a smartphone. Design Evaluation of diagnostic performed on Three hundred and eighty one eyes of 193 patients with diabetes at outreach eye clinics in South-India.

The digital monochromatic nonmydriatic fundus photography[6] in the diagnosis of diabetic retinopathy is evaluated on 197 patients sequentially by three different techniques. The dilated ophthalmoscopy was done by an experienced ophthalmologist by capturing 7 standard color mydriatic stereo fields and a single digital monochromatic nonmydriatic image incorporating the disc and macula. Stereo color photographs served as the reference standard and compared to either ophthalmoscopy performed by a physician. The performance parameters such as Sensitivity and specificity of the three methods were compared.

Diabetic Retinopathy (DR) [7] was done by adapting transformer-based learning models to capture the crucial features of retinal images to understand DR severity better using ensembling image transformers, where four models, namely ViT (Vision Transformer), BEiT (Bidirectional Encoder representation for image Transformer), CaiT (Class-Attention in Image Transformers), and DeiT (Data efficient image Transformers), were used to infer the degree of DR severity from fundus photographs.

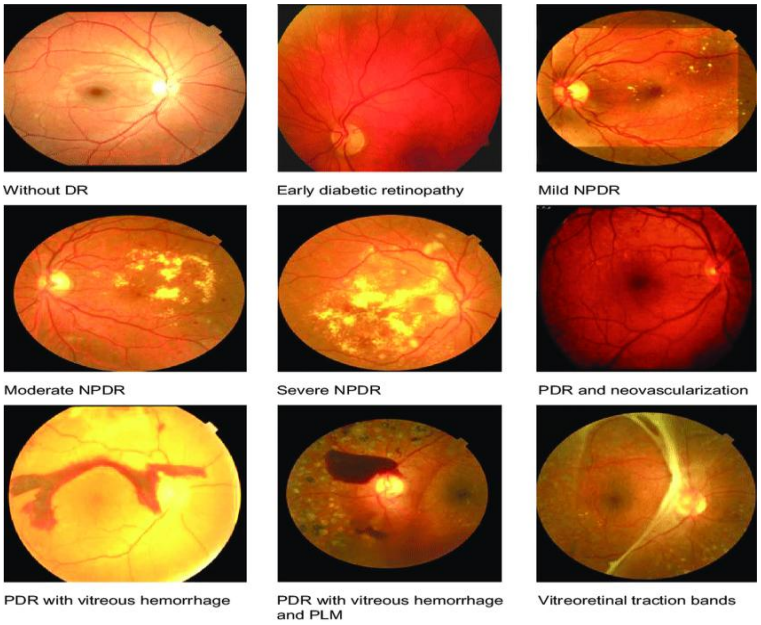


Figure 3: Retinal diseases

3. Proposed work

I. Image Preprocessing

Gaussian filtering serves as a preprocessing method for images to diminish noise and enhance the smoothness of retinal fundus images. It utilizes a Gaussian kernel on adjacent pixels, mitigating slight intensity fluctuations and improving overall image uniformity. This technique can assist in minimizing noise before extracting features

International Journal of Electrical and Electronics Engineers
Volume 18, Issue No. 02, July-December 2026
ISSN (O) 2321-2055
ISSN (P) 2321-2045

with the proposed Efficient Vision Transformer framework. Nevertheless, overly aggressive smoothing can hide intricate retinal features and minor lesions linked to diabetic retinopathy. As a result, a suitable kernel size is chosen to strike a balance between noise reduction and the retention of clinically important details. The efficacy of Gaussian filtering is assessed by comparing classification outcomes with and without this preprocessing

II. Vision Transformer Architecture

The proposed model uses a lightweight pretrained Vision Transformer as its feature-learning backbone. A pretrained architecture is preferred because labeled medical image datasets may be relatively small compared with the datasets typically required to train large transformer models from scratch.

The input image is divided into non-overlapping patches. Each patch is converted into a vector representation and projected into an embedding space. Positional embeddings are added to preserve information about the spatial arrangement of patches. The resulting sequence is processed through transformer encoder layers containing multi-head self-attention, feed-forward networks, residual connections, and layer normalization.

The self-attention mechanism enables the model to learn relationships between different retinal regions. This may help the model represent both local abnormalities and broader retinal context. The final image representation is passed to a classification head, which generates five output scores corresponding to the DR severity grades. For efficient implementation, a small or compact pretrained model, such as a ViT-Tiny or ViT-Small variant, may be evaluated. The exact architecture will depend on available hardware, pretrained weights, memory constraints, and measured inference time. The final paper must identify the precise model variant and report its configuration.

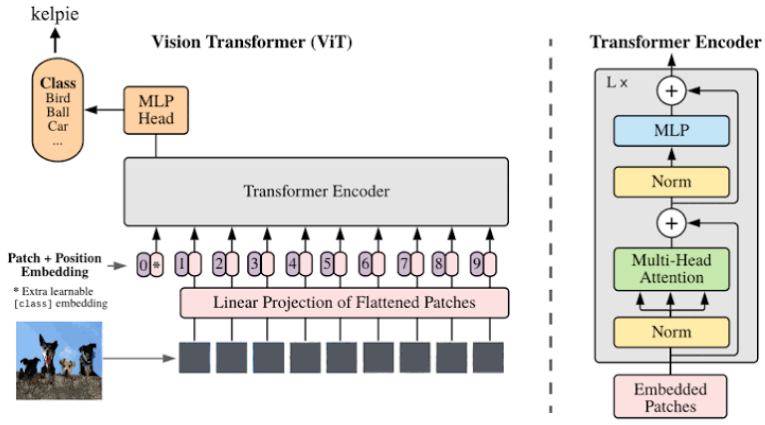


Figure 4: Vision Transformer

The Vision Transformer divides each preprocessed image into small patches and converts them into numerical embeddings. Positional embeddings are added to preserve information about the spatial arrangement of these patches. The Transformer encoder uses multi-head self-attention to learn relationships between different retinal regions and capture relevant features, including blood vessel patterns, microaneurysms, hemorrhages, and exudates. These learned representations are then passed to the classification layer to predict the severity of diabetic retinopathy. A lightweight Vision Transformer architecture is selected to reduce computational complexity while maintaining effective feature representation. The following features are extracted. The proposed framework uses the feature representations learned by the Vision Transformer rather than requiring a

separate handcrafted feature-extraction stage. During training, the model learns image representations that support discrimination between the five severity categories. The classification head maps the learned representation to five output logits. A softmax operation converts these logits into class probabilities. The class with the highest predicted probability is selected as the model's predicted severity grade.

4. Datasets:

a) APTOS 2019 Blindness Detection Dataset

The APTOS 2019 Blindness Detection dataset was released through a competition organized by the Asia Pacific Tele-Ophthalmology Society in collaboration with Aravind Eye Hospital. It contains color retinal fundus photographs collected under varying imaging conditions and is widely used for automated diabetic retinopathy classification. The labeled dataset contains approximately 3,662 images, each assigned a severity grade ranging from 0 to 4. These grades represent no diabetic retinopathy, mild, moderate, severe, and proliferative diabetic retinopathy, respectively. The images exhibit variations in illumination, focus, resolution, and camera characteristics, making the dataset suitable for investigating image preprocessing and model generalization. In this study, APTOS 2019 will be used to train and evaluate the proposed Vision Transformer framework. The labeled images will be divided into training, validation, and test subsets using stratified sampling to preserve class representation.

b) 3.2.2 EyePACS Dataset

The EyePACS dataset is a large collection of retinal fundus photographs obtained from diabetic eye-screening programs. It was used in the 2015 Kaggle Diabetic Retinopathy Detection competition and contains images captured under different clinical and imaging conditions. The labeled training dataset commonly used from this competition contains 35,126 images classified into five diabetic retinopathy severity grades. The dataset includes substantial variations in image quality, illumination, and resolution, as well as an imbalance in the number of samples across severity categories. These characteristics make EyePACS useful for studying the robustness of deep learning models and the effects of class-imbalance handling. In this research, EyePACS may be used as an additional training dataset or for evaluating generalization, provided that dataset access, label consistency, and an appropriate independent evaluation protocol are established.

c) 2.3 Messidor-2 Dataset

The Messidor-2 dataset contains retinal fundus examinations collected for research into computer-assisted diabetic retinopathy assessment. The official database includes 874 examinations comprising 1,748 retinal images, with images acquired using a non-mydriatic fundus camera. It provides a useful source of retinal images for investigating the performance of automated image-analysis methods under different acquisition conditions. However, the official Messidor-2 database does not itself provide a definitive diabetic retinopathy severity ground-truth annotation for every image. Researchers must therefore verify the source and reliability of any grading labels used in their experiments. In this study, Messidor-2 may be considered for external evaluation if suitable, documented labels and a compatible grading protocol are available.

5. Evaluation Metrics:

i. Accuracy

Accuracy is a fundamental evaluation metric for assessing the overall performance of a classification model. It is the ratio of the correctly predicted instances to the total instances in the dataset

$$Accuracy=TP+TN/TP+FP+TN+FN$$

ii. Precision

Precision evaluates the accuracy of the positive prediction made by the classifier. In precision specifies how many were actually positive in the total number of instances.

$$Precision = (TP) / (TP) + (FP)$$

iii. Recall

The recall is also known as sensitivity or true positive rate. It is the ratio of the number of true positive predictions to the total number of actual positive instances in the dataset. Recall measures the ability of a model to identify all relevant instances.

$$Recall = (TP) / (TP) + (FN)$$

iv. F1-Score

F1 score is the harmonic mean of precision and recall. It provide a single metric that balances the trade-off between precision and recall. It is especially useful when the class distribution is imbalanced.

$$F1\ Score = 2 \times [(Precision \times Recall)/ (Precision + Recall)]$$

6. Conclusion:

The Vision Transformer (ViT)-based approach provides an effective method for the automated detection and classification of diabetic retinopathy using retinal fundus images. By utilizing a self-attention mechanism, the model can identify important retinal features and detect abnormalities associated with different stages of diabetic retinopathy. The proposed system helps reduce the time and effort required for manual screening while supporting early diagnosis and accurate disease classification. It can assist ophthalmologists in identifying retinal changes and making timely treatment decisions. Overall, the Vision Transformer offers a promising deep learning approach for diabetic retinopathy screening, improving the efficiency of retinal disease detection and helping to prevent vision loss. Further improvements using larger datasets and advanced model optimization techniques may enhance its accuracy and reliability.

References:

[1] Gardner,G.,Keating,D.,Williamson,T.,Elliott,A.,1996. Automated detection of diabetic retinopathy using an artificial neural network: a screening tool. Br. J. Ophthalmol. 86,940– 944.

[2] Guo, Chen & Yu, Minzhong & Li, Jing. (2021). Prediction of Different Eye Diseases Based on Fundus Photography via Deep Transfer Learning - jcm-10-05481-2. Journal of Clinical Medicine. 10. 10.3390/jcm10235481.

[3] Cen, LP., Ji, J., Lin, JW. et al. Automatic detection of 39 fundus diseases and conditions in retinal photographs using deep neural networks. Nat Commun 12, 4828 (2021). https://doi.org/10.1038/s41467-021-25138-w

[4] Sahlsten, Jaakko & Jaskari, Joel & Kivinen, Jyri & Turunen, Lauri & Jaanio, Esa & Hietala, Kustaa & Kaski, Kimmo. (2019). Deep Learning Fundus Image Analysis for Diabetic Retinopathy and Macular Edema Grading.

[5] Wintergerst, Maximilian & Mishra, Divyansh & Hartmann, Laura & Shah, Payal & Konana, Vinaya & Sagar, Pradeep & Berger, Moritz & Murali, Kaushik & Holz, Frank & Shanmugam, Mahesh & Finger, Robert. (2020). Diabetic Retinopathy Screening Using Smartphone-Based Fundus Imaging in India. Ophthalmology. 127. 10.1016/j.ophtha.2020.05.025.

[6] Danny Y. Lin, Mark S. Blumenkranz, Rose Brothers, and Digital Diabetic Screening Group, The Role of Digital Fundus Photography in Diabetic Retinopathy Screening Diabetes Technology & Therapeutics 1999 1:4, 477-487

[7] Adak, Chandranath & Karkera, Tejas & Chattopadhyay, Soumi & Saqib, Muhammad. (2023). Detecting Severity of Diabetic Retinopathy from Fundus Images using Ensembled Transformers.