



OPEN

Visual impairment prevention by early detection of diabetic retinopathy based on stacked auto-encoder

Shagufta Almas¹, Fazli Wahid^{1,2}, Sikandar Ali¹, Ahmed Alkhyyat^{3,7,8}, Kamran Ullah⁴, Jawad Khan⁵ & Youngmoon Lee⁶✉

Diabetic retinopathy (DR) presents a significant concern among diabetic patients, often leading to vision impairment or blindness if left untreated. Traditional diagnosis methods are prone to human error, necessitating accurate alternatives. While various computer-aided systems have been developed to assist in DR detection, there remains a need for accurate and efficient methods to classify its stages. In this study, we propose a novel approach utilizing enhanced stacked auto-encoders for the detection and classification of DR stages. The classification is performed across one healthy (normal) stage and four DR stages: mild, moderate, severe, and proliferative. Unlike traditional CNN approaches, our method offers improved reliability by reducing time complexity, minimizing errors, and enhancing noise reduction. Leveraging a comprehensive dataset from KAGGLE containing 35,126 retinal fundus images representing one healthy (normal) stage and four DR stages, our proposed model demonstrates superior accuracy compared to existing deep learning algorithms. Data augmentation techniques address class imbalance, while SAEs facilitate accurate classification through layer-wise unsupervised pre-training and supervised fine-tuning. We evaluate our model's performance using rigorous quantitative measures, including accuracy, recall, precision, and F1-score, highlighting its effectiveness in early disease diagnosis and prevention of blindness. Experimental results across different training/testing ratios (50:50, 60:40, 70:30, and 75:25) showcase the model's robustness. The highest accuracy achieved during training was 93%, while testing accuracy reached 88% on a training/testing ratio of 75:25. Comparative analysis underscores the model's superiority over existing methods, positioning it as a promising tool for early-stage DR detection and blindness prevention.

Keywords Diabetic retinopathy, Disability, Deep learning, Stacked auto-encoder, Dropout

Diabetic retinopathy (DR) is a rapidly growing disease among the masses around the globe. The origin of the disease is the presence of a high level of sugar in the blood of the person who is diabetic. This increase in glucose affects many organs including the heart, gums, kidneys, eyesight, bones, etc. The perpetual rise in glucose levels damages retinal vessels, which may lead to permanent blindness¹. In earlier stages, retinal vessels only swell, but in severe cases, the retinal vessels rupture and leak blood into the eye.

In the working-age population around the world, diabetic retinopathy is the major cause of vision loss and blindness, but it is also becoming more common in the elderly. However, not all individuals with diabetes mellitus (DM) will develop DR it is a complication that arises in some patients with prolonged or poorly controlled diabetes. DR is divided into two types; proliferative DR (PDR) and non-proliferative DR (NPDR), with a patient presenting with either NPDR or PDR at a given time, not both simultaneously². The estimated number of people worldwide with diabetes is projected to increase by almost 600 million by 2040 compared to

¹Department of Information Technology, The University of Haripur, Haripur 22620, Pakistan. ²Collage of Science and Engineering, School of Computing, University of Derby, Derby DE22 3AW, UK. ³College of Technical Engineering, The Islamic University, Najaf 54001, Iraq. ⁴Department of Biology, The University of Haripur, Haripur 22620, Pakistan. ⁵School of Computing, Gachon University, Seongnam 13120, Republic of Korea. ⁶Department of Robotics, Hanyang University, Ansan 15588, Republic of Korea. ⁷Department of Computers Techniques Engineering, College of Technical Engineering, The Islamic University of Al Diwaniyah, 58001, Al Diwaniyah, Iraq. ⁸Department of Computers Techniques Engineering, College of Technical Engineering, The Islamic University of Babylon, 51002 Babylon, Iraq. ✉email: youngmoonlee@hanyang.ac.kr

415 million in 2015. NPDR is the earlier stage of DR and classified into four stages: normal, mild, moderate, and severe. Although NPDR may initially present without noticeable symptoms, it tends to worsen over time^{3,4}. Because DR is classified as a microangiopathy, even in its early stages, all retinal vessels exhibit some degree of alteration due to diabetes-induced changes in small blood vessels. In the mild stage, microaneurysms are detected which are micro-scale swellings in retinal blood vessels that, in turn, causes swelling in the macula. In moderate stage, swelling increases, which impedes blood flow, hence resulting in deposits in the macula. Following the severe stage, huge blockage starts occurring in retinal vessels, causing hindrance in blood flow. Due to this hindrance, the body receives signals to form new vessels. In the later stage of DR, known as PDR, these newly formed blood vessels are fragile and prone to leaking, often resulting in severe vision impairment. Diabetic macular edema (DME), characterized by fluid buildup in the macula, can occur in both NPDR and PDR stages and may significantly impact vision⁵.

Diabetic Retinopathy is a worldwide health issue, which if not treated in time may result in lifelong blindness. According to a survey conducted by the World Health Organization, 366 million people will be predicted sight-threatened due to diabetes in 2030^{6,7}. To diagnose the disease at its earlier stages, many systems have developed worldwide to aid ophthalmologists in early detection and classification using deep learning⁸. Deep learning convolutional neural network has been extensively used for medical image analysis, but it has certain limitations. While preprocessing and post-processing can enhance CNN performance, it may add time to model training depending on the dataset and task. Additionally, CNNs require a large amount of data to achieve reliable accuracy, as well as preprocessing steps that can lead to information loss during image handling, feature extraction, and segmentation^{6,9}. For this reason, we have proposed a stacked auto-encoder mechanism that eludes feature extraction and segmentation, which enhances classification accuracy and speed^{10,11}.

This research provides a summarized review of state-of-the-art techniques and tools used in the detection of diabetic retinopathy stages and elaborates on the variation in retinal vessels due to diabetes and the impact of these variations on human vision. The proposed system will detect different stages of Diabetic retinopathy. Images of the retina from eye will be captured and passed on to the proposed system. If there are no red spots on the retina, it will result in a normal retina. However, in the presence of red spots, which may indicate either microaneurysms or small intraretinal hemorrhages, mild Diabetic retinopathy will be detected. In the moderate stage, hard exudates, and cotton wool spots are present. In the severe stage, multiple hemorrhages occur, and leakage of fluid from vessels occurs. The proliferative stage involves the formation of new vessels or may say neovascularization. This research proposes a system that gives more accurate results to ensure the best eye care and in-time treatment so that diabetic patients may be prevented from permanent blindness. The main idea of this research is to detect particular stages of diabetic retinopathy using a stacked auto-encoder, a deep learning approach hence introducing a mechanism, which can help ophthalmologists to achieve better results than existing systems. This will help in early diagnosis of the disease stage and prevent the patient from blindness if treated in time. In Fig. 1, the stages of diabetic retinopathy and normal case are shown and visualized through their related images.

Why use stacked auto encoders

In the realm of diabetic retinopathy (DR) detection, several established methods have shown promise, including traditional machine learning models (such as Support Vector Machines and Random Forests) and deep learning techniques (such as Convolutional Neural Networks, or CNNs). While these methods have made significant strides in automating the detection process, they still face notable limitations, especially in terms of early-stage detection, computational efficiency, and adaptability to different datasets. Our proposed technique aims to address these challenges and bring substantial improvements in the following ways:

Early-stage detection

Current methods, including CNNs, tend to perform well when detecting advanced stages of diabetic retinopathy, where clear and prominent features (e.g., large hemorrhages or exudates) are present. However, they struggle with detecting subtle signs of DR in its early stages, such as microaneurysms and small hemorrhages. These early-stage features are often overlooked because of their relatively small size and the difficulty in differentiating them from normal variations in the retina. Our technique, using stacked auto-encoders, is designed to capture these subtle features more effectively. By learning hierarchical representations of the data in an unsupervised manner, the stacked auto-encoders can extract features even from limited or noisy data, making them particularly suitable for early-stage DR detection.

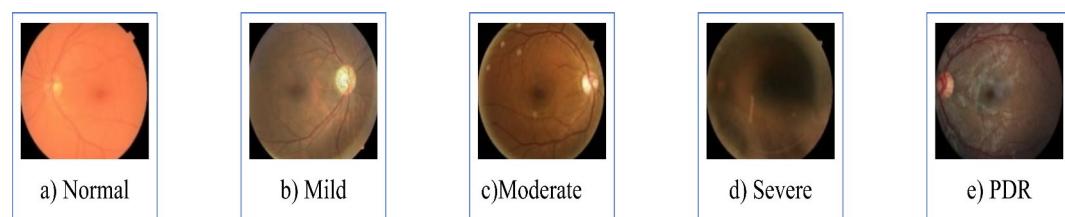


Fig. 1. Stages of Diabetic Retinopathy. Panel (a) shows a normal retina, while panels (b) to (d) illustrate NPDR stages. Panel (e) represents PDR¹².

Computational efficiency and resource constraints

Traditional deep learning models, particularly CNNs, often require substantial computational resources, large labeled datasets, and high-performance hardware to achieve optimal performance. This makes them less practical in real-world settings, especially in resource-constrained environments such as rural clinics or mobile health units, where access to powerful hardware is limited. In contrast, our approach, which utilizes stacked auto-encoders, offers a more lightweight and computationally efficient alternative. The stacked auto-encoder can operate effectively with smaller datasets and fewer computational resources, making it a practical solution for widespread adoption in diverse healthcare settings.

Adaptability to diverse data

Existing methods, especially those based on deep learning, are typically trained on large, diverse datasets from specific populations. However, they often struggle when deployed on new, unseen datasets or in different clinical settings due to variations in the data (e.g., different imaging equipment, lighting conditions, or patient demographics). The unsupervised nature of the stacked auto-encoder allows it to generalize better to different types of retinal images, even when labeled data is sparse. This increased adaptability is a significant advantage, particularly in cases where it is difficult to collect large, annotated datasets from diverse populations.

Improved feature extraction with limited data

While CNNs are highly effective in extracting features, they require a significant amount of labeled data for training. This is not always feasible in medical applications, where labeled data may be scarce or expensive to obtain. Stacked auto-encoders, however, can learn powerful feature representations in an unsupervised manner, enabling them to work well even with limited labeled data. This is particularly valuable in DR detection, where annotated medical images may be limited.

Major contributions

The main contributions of the proposed model compared to other state-of-the-art models include:

1. The proposed model achieves higher accuracy and performance metrics compared to other models. This indicates its effectiveness in accurately detecting and classifying DR stages, potentially leading to better patient outcomes and early intervention.
2. Unlike some existing models that rely on traditional convolutional neural networks or machine learning algorithms, the proposed model utilizes enhanced stacked auto-encoders. This architecture offers advantages in terms of reduced time complexity, minimized errors, and improved noise reduction.
3. The proposed model is trained and tested on a comprehensive dataset containing a large number of retinal fundus images across four DR stages and one healthy stage. This broader classification scheme allows for more detailed analysis and detection of DR stages compared to models focusing on fewer stages.
4. By accurately detecting and classifying DR stages, the proposed model has the potential to facilitate early diagnosis and intervention, thereby helping to prevent vision loss and improve patient outcomes. The rest of the paper is organized as follows: Section “[Literature review](#)” gives the literature review and problem statement. Section “[Proposed methodology](#)” explains the proposed methodology, including the improved stacked auto-encoder for diabetic retinopathy detection. The experimental results and discussion are presented in Section “[Experiments and Discussion](#)”. A comparative analysis of the proposed method with already existing models is introduced in Section “[Comparative Analysis](#)”. Finally, Section “[Conclusions](#)” summarizes the research conclusion and future directions.

Literature review

Several studies have explored the detection and classification of diabetic retinopathy using machine learning and deep learning techniques on various datasets. One such approach was deployed on three datasets: MESSIDOR, DRIVE, and DIARETDBI. The technique consists of three modules: feature learning via unlabeled retinal images, SFCN (Self-supervised framework clustering), and a classification module. As this is an unsupervised learning technique, data has been classified into clusters, formed based on probabilities using the Fuzzy C-Means algorithm. Two clusters are formed: DR and No DR. As this technique has been applied to three datasets, the achieved accuracies are as follows: Accuracy MESIDOR: 87.6%, DRIVE: 81.7%, DIARETDBI: 84.7%¹³.

In another study, a technique based on the presence of Micro aneurysms which are red lesions present on the retina facet was proposed. For this reason, the machine learning algorithm Principle component analysis is used for binary classification. It involves the following steps: Preprocessing where the colored fundus images are converted into green channels because it can enhance the texture of the lesion from the background, using Gaussian correlation coefficients, Micro aneurysm candidates are filtered. The candidates encompass the following aspects: brightness, contour, normal distribution, and pigment. The machine learning method sparse PCA is applied to classify the candidate features of Micro aneurysms from non-Micro aneurysms. The Database used is from the Retinopathy Online Challenge platform. Keeping in view, this technique takes too long to perform as it requires a number of steps before actual processing¹⁴. Another study proposed Micro aneurysm detection using three machine learning classifiers and the use of a Principal Component analysis algorithm for binary classification. Two datasets are under consideration in this research, and the next few phrases will discuss their purpose: DIARETDBI and Retinopathy Online Challenge. Preprocessing took place before passing through the classifiers, patches were extricated, and proffered as direct input to Random Forest, Neural Network, and Support Vector Machine. PCA and RF were used to reduce input variables in the dataset. The performance was measured compared to deep learning in terms of AUC (SVM: 98.5%, NN: 97.9%, RF: 97.2%) and F-measure (SVM: 92.6%, NN: 89.4%, RF: 88.8%). In addition, classifiers have been trained on DIARETDBI and tested

on Retinopathy Online Challenge, and got improved results which showed the generalization of technique on diverse datasets. Additionally a study by¹⁵ put forwarded the stages of diabetic retinopathy. Diabetes occurs due to abnormal flow of glucose in the bloodstream, which has adverse effects on the retina of the eye as well. The disease may vary according to the destruction made in retinal vessels. The main purpose of this research was to detect and identify the stage of DR. This requires blood vessel detection, following the extraction of vessel features. Furthermore, the stage identification is done by measuring the parameters and area of blood vessels. The Random forest technique is used for the classification of stages, and three stages have been identified, i.e. Normal, moderate, and non-proliferative DR. RF shows accuracy for normal stage is 90% while for the rest of the two stages accuracy is 87.5%.

Several studies have also leveraged deep learning techniques to detect multiple stages of diabetic retinopathy. For instance, a deep learning technique using an ensemble model approach was proposed to detect four stages of diabetic retinopathy and a healthy stage. Convolutional Neural Network models are used to achieve better accuracy than state-of-the-art methods and to fully dig out the feature information for efficient results. The architecture required the preprocessing stage where classes were augmented and then passed through an ensemble model. The KAGGLE dataset was used for training and testing purposes. Collectively these ensemble models gave these performance metrics Accuracy = 80.8%, Recall = 51.5%, Specificity = 86.72%, Precision = 63.85%, F1 score = 53.74%. This model outperformed traditional methods in multi-class classification, but the authors noted that performance could be further improved¹⁶.

A computer-aided diagnostic technique for detecting abnormalities in retinal fundus images and detecting the presence of abnormal patterns in retinal fundus imagery was proposed by¹⁷. Their suggested technique aims at enhancing image data, noise reduction, detecting blood vessels as well as identifying the input image, extracting exudates and Micro aneurysms (MA), feature extraction, and labeling different stages of diabetic retinopathy as mild, moderate, severe NPDR (Non-Proliferative Diabetic Retinopathy) and PDR (Proliferative Diabetic Retinopathy). The use of new imaging transformation methods such as discrete wavelet transform, curvelet transform, and wavelet transform, has been explored by¹⁸ for enhancing retinal images. As a result, their research compares these three optical alterations. Shanthi et al.¹⁹ focus on the classification of Diabetic Retinopathy images using a Convolutional Neural Network with Rectified Linear Unit layers added to it along the Pooling layer and softmax activation function. This model uses the MESSIDOR dataset and classifies 3 stages of disease and the accuracy of each stage is 96%, 95%, and 96%.

Another study emphasizes building a system capable of detecting microaneurysms using a fusion of text and images by deep learning techniques. Performance measures obtained are 99.7% precision and 87.8% recall²⁰. A separate framework developed by³ diagnoses non-proliferative diabetic retinopathy and classify it into mild, moderate, and severe non-proliferative diabetic retinopathy using deep convolutional neural algorithms. Additionally,¹⁵ proposes an approach to diagnose diabetic retinopathy based on the presence of red lesions while dissecting the images in a longitudinal view. The framework evaluates into two classes; no lesion and small red retinal lesion. The system achieves 80% sensitivity. Another model, developed by²¹ which can classify the data into two classes; DR and No - DR by using CNN. The model achieves an accuracy of 95% for two-class classification.

The classification of 2 classes of diabetic retinopathy by using three different CNN Architectures was explored. The accuracies achieve on three architectures VGG16, VGG 19, Inception V3 are 74%, 80%, and 75% respectively²². A framework to detect and classify diabetic retinopathy using dense deep feature fusion of CNN has been proposed. The proposed system²³ detects the hard exudates which state-of-art methods fail to do and achieves an accuracy of 97%.

A hybrid method for detecting and classifying diabetic retinopathy from images was proposed by²⁴. The images are passed through pre-processing which includes histogram equalization, and handed over to deep CNN for the classification of the severity of DR. The MESSIDOR Dataset with 400 images is used, average values are calculated, and obtain 97% accuracy, 94% recall, 98% specificity, 94% precision, F-Score 94%. Another framework enhances the region of interest through a local contrast enhancement approach. Features are extracted on every candidate lesion and determine which images are healthy and which are unhealthy. Three different datasets have been used: E-ophtha, Local, DIARETDB, and MESSIDOR. The results obtained are 0.99 AUC and 98.9% Accuracy on E-ophtha, 1.00 AUC, and 100% Accuracy on local, DIARETDB and MESSIDOR datasets²⁵.

Pre-trained Convolutional Neural Network models, including AlexNet, VGG-16, and SqueezeNet, were proposed, achieving accuracies of 93.46%, 91.82%, and 94.49%, respectively²⁶. The CNN model is modified into 5 fully connected layers model which gives the following results: sensitivity 98.94%, specificity 97.87%, and accuracy 98.15%.

An automated classification system for diabetic retinopathy (DR) was developed to analyze fundus images under varying illumination conditions and fields of view, utilizing machine learning models such as CNN, VGG-16, and VGG-19. The system demonstrated 80% sensitivity, 82% accuracy, 82% specificity, and an Area Under the Curve (AUC) of 0.904, effectively classifying images into five severity grades from 0 to 4. The EyePACS dataset from the 2015 Kaggle competition was used for training and evaluation²⁷. Another approach²⁸ involved training a DenseNet model on approximately 3,662 high-resolution fundus images from the APTOS 2019 Blindness Detection competition on Kaggle, which included data from the Aravind Eye Hospital. The model achieved an accuracy of 96.11% by leveraging DenseNet's feature extraction capabilities and activation functions. A hybrid deep learning model combined transfer learning with Inception-ResNet-v2 and an additional custom CNN block to detect diabetic retinopathy. This method was evaluated on the Messidor-1 and APTOS 2019 datasets from Kaggle, achieving test accuracies of 72.33% and 82.18%, respectively²⁹. VGG-NiN model, which integrates VGG16, spatial pyramid pooling (SPP), and network-in-network (NiN) layers was proposed by³⁰. The SPP layer enabled scale-invariant image processing, while the NiN layers improved nonlinearity and classification

accuracy. When evaluated on the Kaggle EyePACS dataset, the model achieved a Micro-AUC of 95.0 and a Macro-AUC of 84.0. DenseNet-169 was also applied for early DR detection, incorporating datasets from Kaggle. The DenseNet-169 model achieved 90% accuracy, while a regression model tested alongside it attained an accuracy of 78%³¹.

Although many machine learning and deep learning algorithms have been applied for retinopathy detection, there is still much room for improvement that needs to be addressed. For instance, Machine Learning algorithms involve many steps before performing classification, which increases the time complexity of the system. The steps involve image acquisition, feature extraction, reduced images, and then classification. Such pre-processing steps can increase manual workload, typically managed by technicians or automated systems rather than ophthalmologists, because fundus images require noise reduction, contrast, and brightness to achieve consistent results. Disease detection and classification entails adept physicians. To redeem the accurate stage of the disease requires proficient feature extraction and bifurcation of the fundus images as the features show similarity in retinal structures. CNN depends upon multiple layers (Convolutional layer, pooling layer, fully connected layer) which slow down the performance of automated systems. It requires huge data in the training set to outdo the results³².

Building sizeable datasets for CNN in the medical field costs high and surges the workload of developers³³. CNN has a complex mechanism that leads to data loss while reducing images. While CNNs generally perform better on preprocessed and augmented datasets, pre-trained CNN models may face limitations in generalizing to new datasets due to fixed (frozen) layers, which restrict adaptation when applied to data outside of the original training set³⁴.

The proposed approach differs from these models in several key aspects:

1. *Utilization of Enhanced Stacked Auto-encoders:* While many existing models rely on traditional convolutional neural networks or other machine learning algorithms, the proposed approach leverages enhanced stacked auto-encoders. This choice offers advantages in terms of reduced time complexity, minimized errors, and improved noise reduction compared to CNNs.
2. *Comprehensive Dataset and Five-Stage Classification:* The proposed approach utilizes a comprehensive dataset from KAGGLE, containing a large number of retinal fundus images across four stages of diabetic retinopathy and one healthy stage. This broader classification scheme allows for more detailed analysis and detection of DR stages compared to models focusing on fewer stages.
3. *Quantitative Evaluation and Performance Metrics:* Unlike some of the referenced models that may lack detailed quantitative measures, the proposed approach rigorously evaluates its performance using metrics such as accuracy, recall, precision, and F1-score. This ensures a thorough assessment of the model's effectiveness in DR detection and classification.
4. *Superior Accuracy and Performance:* The results of the proposed approach demonstrate superior accuracy and performance compared to existing models, as evidenced by the quantitative measures presented in the study. This indicates that the proposed approach may offer more accurate and reliable results for the detection and classification of DR stages, potentially leading to better patient outcomes and early intervention. Figure 2 is an existing model for detecting and grading diabetic retinopathy through fundus images.

The model explains all the preprocessing and post-processing steps other than the actual CNN architecture. We may observe an ample data loss during pre-processing and post-processing steps which minimize the accuracy of the results and maximize the error margin. Moving forward, keeping in mind all the limitations aforementioned, there is a need to propose such a model that processes in minimal steps, improves accuracy, and cuts down on error, unlike state-of-art methods.

Proposed methodology

The proposed methodology for the stacked auto-encoder system involves three stages as shown in Fig. 3, which are image acquisition, model development, and model evaluation. In the first stage, images are obtained from various sources and integrated into a single database. The second phase belongs to model development where the acquired data trains the model and then tests the model. Third, Model Evaluation is based on several performance metrics with the help of a confusion matrix. The confusion matrix consists of Sensitivity, Accuracy, Specificity, F1 Score, and Precision.

Image Acquisition is the first step in our proposed methodology where fundus images are retrieved from a reliable source. As our research focuses on the classification of Diabetic Retinopathy into four stages (mild, moderate, severe, proliferative) along with one healthy (normal) stage, we will split the data into five levels (0–4). After acquiring the images, we will move toward the model development. Since the main purpose of the study is to classify the images into multiple classes, we require multiple layers to handle complex data. Auto-Encoders comprise encoders and decoders. Encoders compress the raw data while decoders attempt to recreate the data from the compressed version provided by encoders. Encoders consist of dense layers of neurons which start decreasing as we move into the model and start increasing as we move along the decoder's part. Encoders and decoders both are connected through a compression layer which is called the bottleneck of the model, here the data compression occurs³⁵.

Enhanced stacked auto-encoders for disease detection

The stacked auto-encoder system is used as a feature extractor or maybe a dimensionality reduction network. It compresses the raw data using multiple layers or encoders and proceeds to the bottleneck of the architecture or hidden layer, then again, the compressed data starts decoded increasingly on the output layer. Decoders are symmetries of encoders because the compressed data recreates itself in as many dimensions as on the first

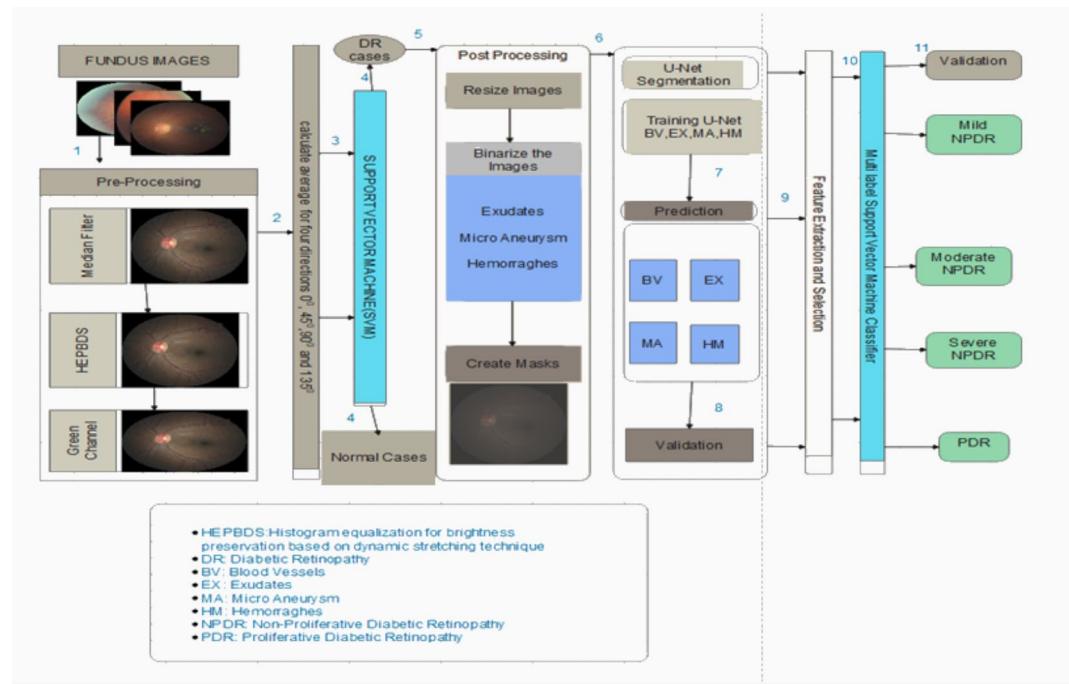


Fig. 2. CNN architecture and steps involved in DR detection and stages.

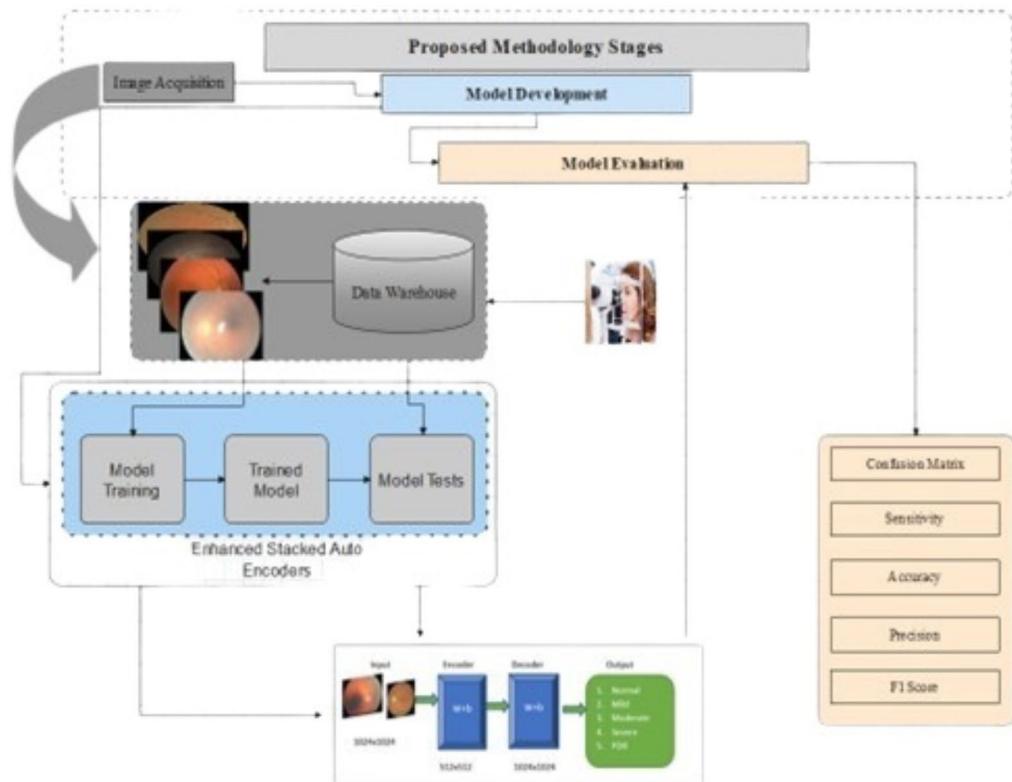


Fig. 3. Stages of proposed methodology.

encoder. Stacked Auto-Encoders have been widely used for numerous disease detection including brain tumors, breast cancer, chronic kidney disease, cervical cancer, Parkinson's disease etc. This deep learning technique is best suited for binary as well as multi-class classification purposes, which makes it stand out in biomedical research.

Brain is a complex organ that regulates the normal functioning of the whole body. Any abnormality may lead to fatal consequences. To diagnose the malfunction of the brain, deep learning technique stacked auto-encoder has been used on MRI brain image datasets and achieved accuracy 96.6% and 95% on two different datasets. The models classified images into five classes: normal, infectious, tumor, degenerative, and stroke³⁶.

Parkinson's illness (PD) may well be a quiet neural disorder that affects a range of people. This has continuously growing stages to halt the entire neural activities of any person. There are many techniques projected to observe and predict medical symptoms and measurements. The medical measurements provided by wholly different experiments ought to be effectively handled to supply concrete results on the detection of disease. Using four different variants of auto-encoders predicts 5–10% diseases in accuracy than state-of-art methods³⁷.

Breast cancer is common among females all around the world. It can be cured if diagnosed at its early stage but if not detected it is fatal. Many machine learning and deep learning techniques have been used to detect the stages, but the stacked sparse auto-encoder outperforms other techniques and achieves 98% accuracy with improvement in performance metrics: precision, sensitivity, and specificity³⁸.

Kidney diseases are fast-growing and are of great concern in the present era. If diagnosed at early stages, can be treated and a patient may come to normal life but if it is not diagnosed in proper time kidneys may fail which decreases the life expectancy of the population. Stacked Auto-encoder proves to be the best technique in an automated diagnosis system for kidney diseases. Accuracy achieved is 100% by using a stacked auto-encoder in contrast with SoftMax classifier for the classification of diseases³⁹.

Aforementioned diseases are a few of the examples that showed better results in terms of accuracy and other performance metrics when using stacked auto-encoders than state-of-art methods. The technique is proposed for diabetic retinopathy detection for multi-classification of disease. In the proposed methodology of enhanced stacked auto-encoders for diabetic retinopathy detection, the auto-encoder architecture is stacked, multiple layers of encoders are used sequentially, followed by corresponding decoders. The auto-encoder is trained in a step-by-step manner, with each auto-encoder in the stack trained separately. The output of one auto-encoder is fed as input to the next one, forming a cascaded structure. Finally, the entire stacked network is trained to fine-tune the parameters⁴⁰. Encoder 1 processes the input images, transforming them into a lower-dimensional representation while retaining essential features. This compressed representation is then further refined by Encoder 2, which operates on the output of Encoder 1, extracting deeper and more abstract features from the data. Following this, Encoder 3 continues the process, refining the representation even further to capture finer details and patterns. Within each encoder, the rectified linear unit (ReLU) activation function is utilized to introduce non-linearity and capture complex patterns in the data. Additionally, the final encoder incorporates a SoftMax layer, enabling the model to perform multi-class classification and accurately distinguish between different diabetic retinopathy categories. This hierarchical structure of individual encoders allows the model to progressively extract informative features from the input data, ultimately enhancing its performance in diabetic retinopathy detection tasks. Below is discussed with the help of the figures how auto-encoders work. The architecture is shown in Figs. 4, 5, 6, 7 and 8. The encoding process is mathematically described in Eq. (1), and the decoding process in Eq. (2) (see Appendix A for detailed equations).

Increasing or decreasing the number of auto-encoders in the proposed approach could have several impacts on the performance and behavior of the model. Adding more auto-encoders would increase the overall complexity of the model, potentially allowing it to capture more intricate features and patterns in the data. Conversely, reducing the number of auto-encoders would simplify the model, which might make it more computationally efficient but could limit its capacity to represent complex relationships in the data. More auto-encoders typically mean longer training times due to the increased number of parameters and computations involved. Conversely, fewer auto-encoders might lead to faster training times but may sacrifice model performance⁴¹.

The detailed steps for training the enhanced stacked auto-encoders for disease are provided in Algorithm 1 (see Appendix B).

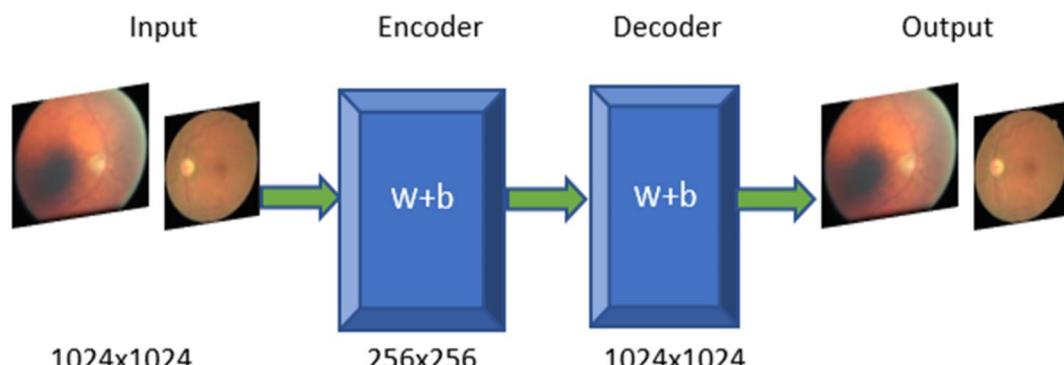


Fig. 4. Encoder 1.

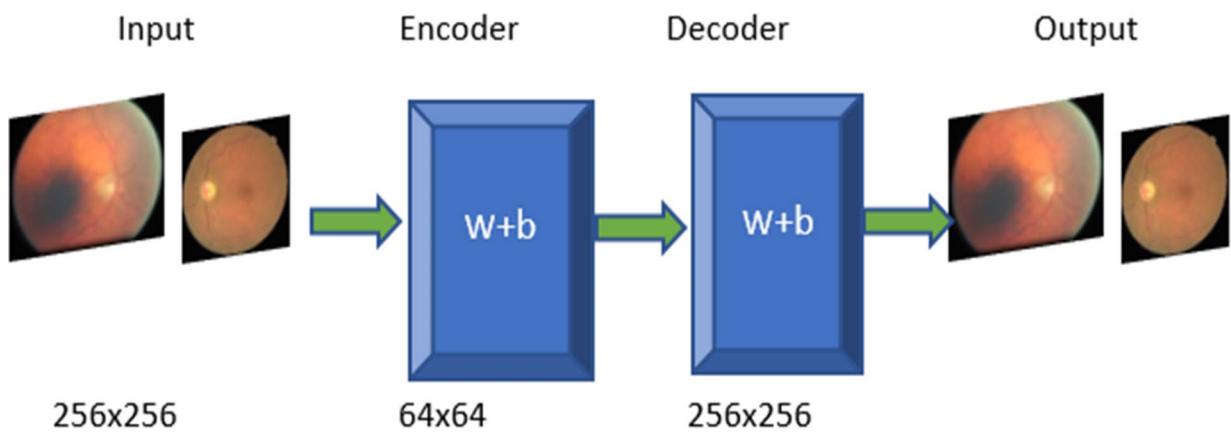


Fig. 5. Encoder 2.

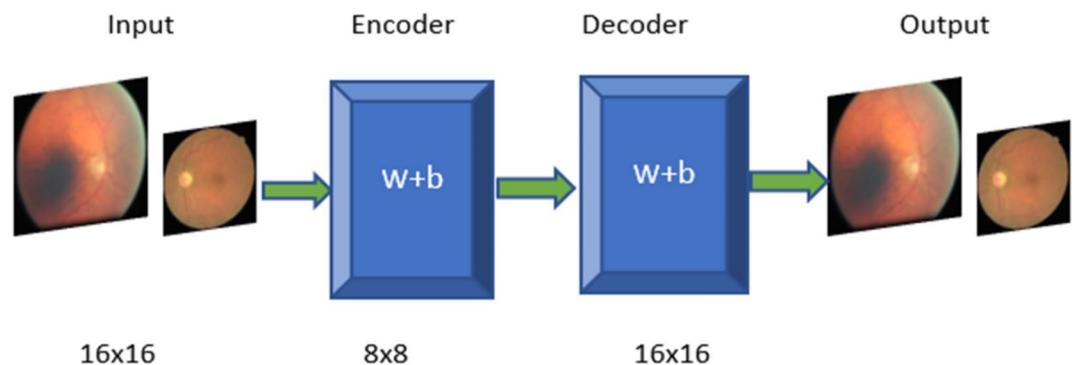


Fig. 6. Encoder 3.

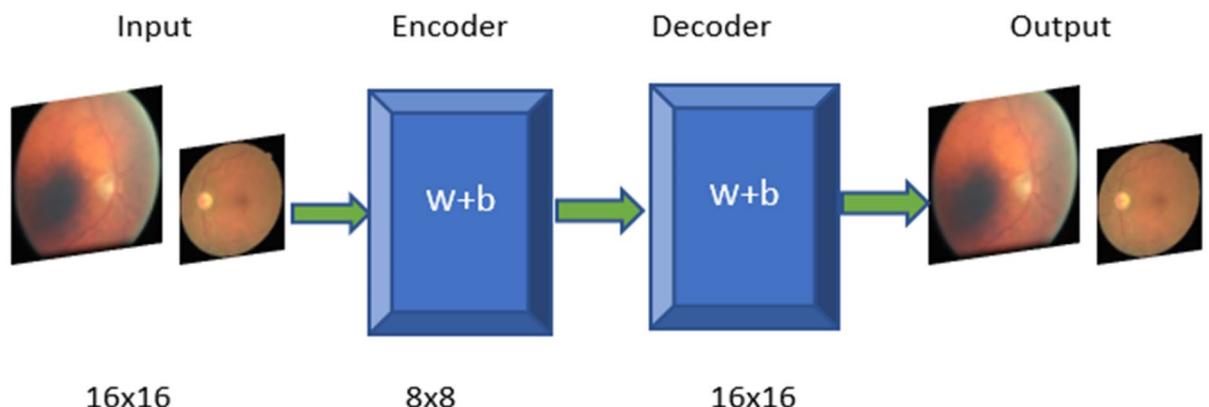


Fig. 7. Encoder 4.

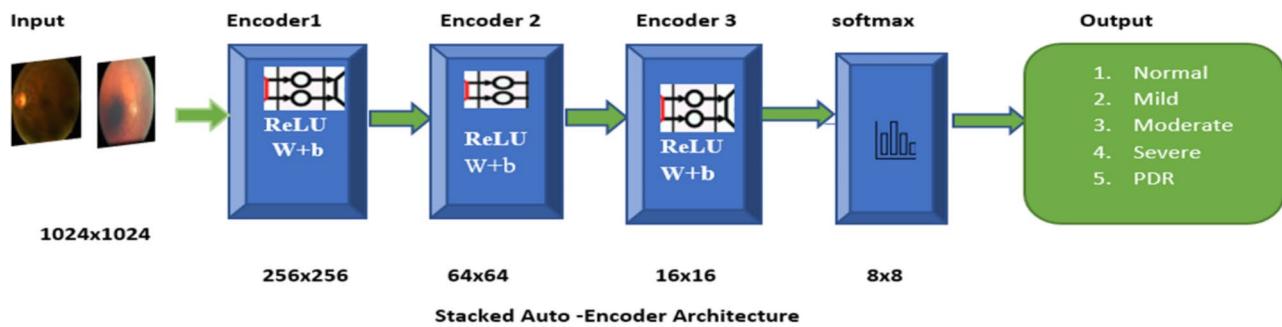


Fig. 8. Encoders are stacked.

Dropout technique

Dropout is a viable approach widely used to minimize overfitting while training neural networks. Overfitting occurs when the model learns to capture noise or irrelevant patterns in the training data, leading to poor generalization to unseen data. Underfitting occurs when the model fails to capture the underlying structure of the data, resulting in high bias and low model complexity. The model may have low training and validation accuracy, indicating that it is unable to learn meaningful features from the data⁴². Good fitting occurs when the model learns to generalize well to unseen data by capturing the underlying patterns in the training data without overfitting or underfitting. The model achieves high training accuracy while maintaining good performance on validation or test data⁴³. The impacts of overfitting, underfitting, and good fitting on stacked auto-encoders are similar to those on other types of deep learning models. Overfitting can lead to overly complex representations, underfitting results in simplistic representations, and good fitting strikes a balance between complexity and generalization, resulting in effective learned representations.

During the training process, dropout transforms the neurons of the hidden layer while conserving the weights of neurons⁴⁴. As the deep neural network contains numerous hidden layers between their input and output layers, they form complex relationships. The noise in training data can lead to overfitting. In this case, the dropout technique can rescue the architecture by arbitrarily dropping out neurons in the hidden layers. This dropout is temporary which means a few neurons are switched off randomly in the network from all the input and output connections. The phenomenon can also be termed a “thinned Network”. Each neuron in the layer has a definite value of p irrespective of other neurons. The maximum range for probability varies between 1 and 0.5. The network with n neurons turned into a 2^n thinned network. This technique is only applicable to training data and not in a testing phase. During testing, an average approximation of units is made to lessen the generalization errors. Adding noise in hidden layers and averaging methods while testing the model make it robust and help to improve errors in hidden layers⁴⁵.

This paper focuses on improving the feature extraction power and accuracy of stacked auto-encoders for the classification of diabetic retinopathy. Figure 9 illustrates the connections between the neurons in the neural network without using dropout approach, in which all neurons are connected. Also, Fig. 10 shows the neural network architecture with using dropout approach.

Some of the latest auto-encoder models that incorporate dropout mechanisms include Variational Auto-encoders (VAEs), Denoising Auto-encoders, and Convolutional Auto-encoders (CAEs). VAEs are a type of generative model that incorporates a probabilistic approach to encoding data. Dropout layers can be included in the encoder and decoder components of VAEs to regularize the model and improve generalization⁴⁶. Denoising auto-encoders are trained to remove noise from input data, and dropout layers can be used as a form of noise injection during training. By randomly dropping out units in the input data or hidden layers, denoising auto-encoders learn to recover the original, noise-free data. CAEs are designed for image data and leverage convolutional layers for encoding and decoding. Dropout layers can be added to the fully connected layers in the encoder and decoder parts of CAEs to prevent overfitting and improve reconstruction accuracy.

While dropout has been utilized in auto-encoder models previously, the proposed approach distinguishes itself by its application within the context of diabetic retinopathy classification. The proposed auto-encoder model have a customized architecture tailored specifically for the task of diabetic retinopathy classification. This architecture could be optimized to handle the characteristics of retinal images and extract relevant features efficiently^{46,47}.

Rectified linear unit (ReLU)

ReLU is a non-linear activation function, most preferable over other activation functions such as sigmoid function, SoftMax function, and hyperbolic functions. Because only a specified number of activated neurons at a time, ReLU is more efficient among all activation functions⁴⁴. The neurons are deactivated when the results of linear mapping are zero {Sharma, 2017 #39}. Mathematically it can be written as:

$$f(x) = \max(0, x) \quad (1)$$

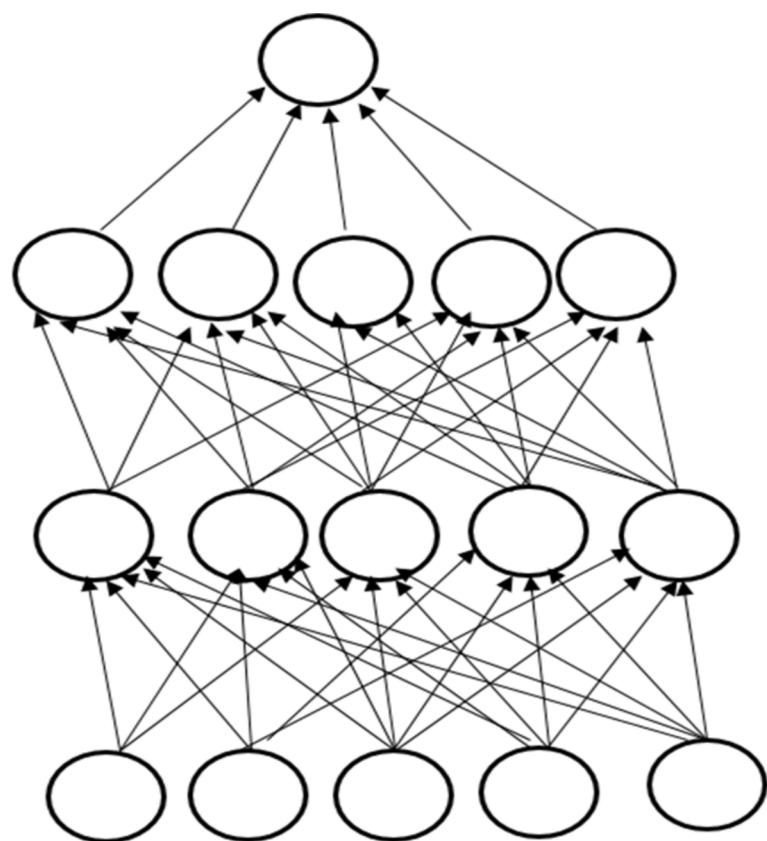


Fig. 9. A neural network without dropout approach, all neurons are connected.

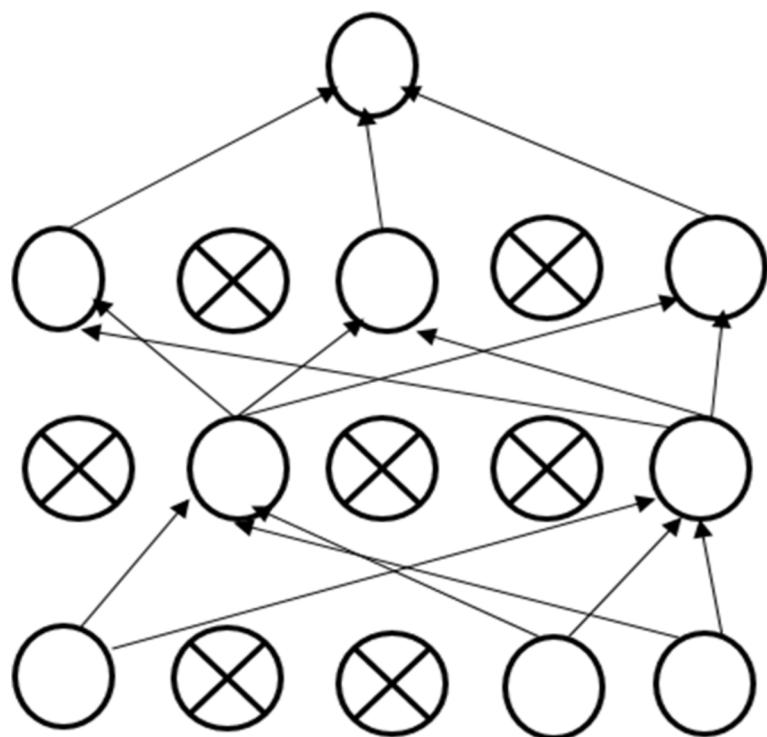


Fig. 10. A neural network with dropout approach, some neurons are disconnected.

Performance measures

Following performance evaluators are measured in this proposed method stacked auto-encoder, and the definition of each is given below:

Accuracy

When we use the term accuracy, we usually imply accuracy. The number of correct predictions divided by the total number of input samples is the ratio².

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

Sensitivity

The fraction of actual positive cases that were projected as positive is known as sensitivity (or true positive). Recall is another synonym for sensitivity. Sensitivity is a measure of the proportion of patients with the condition who were accurately identified as those with the disease. The sensitivity is a measure of how well our model detects True Positives².

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad (3)$$

Precision

The ratio of true positives to all positives is known as precision².

$$\text{Precision} = \frac{TP}{TP + FP} \quad (4)$$

F1-Score

The normal distribution of Precision and Recall is the F1-Score. As a result, this score considers includes false positives and false negatives². Although it is not as simple as accuracy, F1 is frequently more effective than accuracy, especially if the class distribution is unequal. When false positives and false negatives entail equivalent values, accuracy works well. When false positives and false negatives possess equivalent values, accuracy works well. It is best to look for both Precision and Recall unless the value of false positives and false negatives is considerably different.

$$\text{F1-Score} = 2 \times \left(\frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \right) \quad (5)$$

Receiver operating characteristic (ROC)

A receiver operating characteristic (ROC) curve is a graph that shows how well a training set performs across all categorization levels {Bradley, 1997 #40}⁴⁸.

The curve is obtained by using two parameters True Positive and False Positive.

Area under curve (AUC)

As we have defined the ROC, the area obtained under that ROC is called the Area under Curve (AUC). The AUC is a composite measure of efficiency that considers all possible classification levels⁴⁹.

Confusion matrix

The confusion matrix depicts the true positives and true negatives.

Developed framework

The developed framework of diabetic retinopathy classification from dataset images is given in Fig. 11. It consists of several steps, including data preprocessing, data augmentation, training enhanced stacked auto-encoders, and proposed algorithm development. In the following subsection, each step is explained in detail.

Data preprocessing

To carry out this research dataset is taken from the publicly available dataset KAGGLE¹² which contains five classes containing images 35216. As the dataset contains five classes, images in each class are unevenly distributed. The size of each image is 3888 × 2951. In this step, the images of each class are resized and normalized to be suitable for training and evaluating the developed model.

Data augmentation

In the original dataset, the distribution of images across the five diabetic retinopathy (DR) stages is highly imbalanced. For example, the “Normal level 0” class has 25,810 images, while the “Severe level 3” and “PDR level 4” classes have only 873 and 708 images, respectively. This imbalance can lead to poor classification performance.

To address this issue, **data augmentation** is applied to the underrepresented classes, generating synthetic samples that increase their image counts. Data augmentation techniques used include:

- *Random rotation* between 0 and 180 degrees,
- *Random zooming*, Varying the image size to simulate different scales,

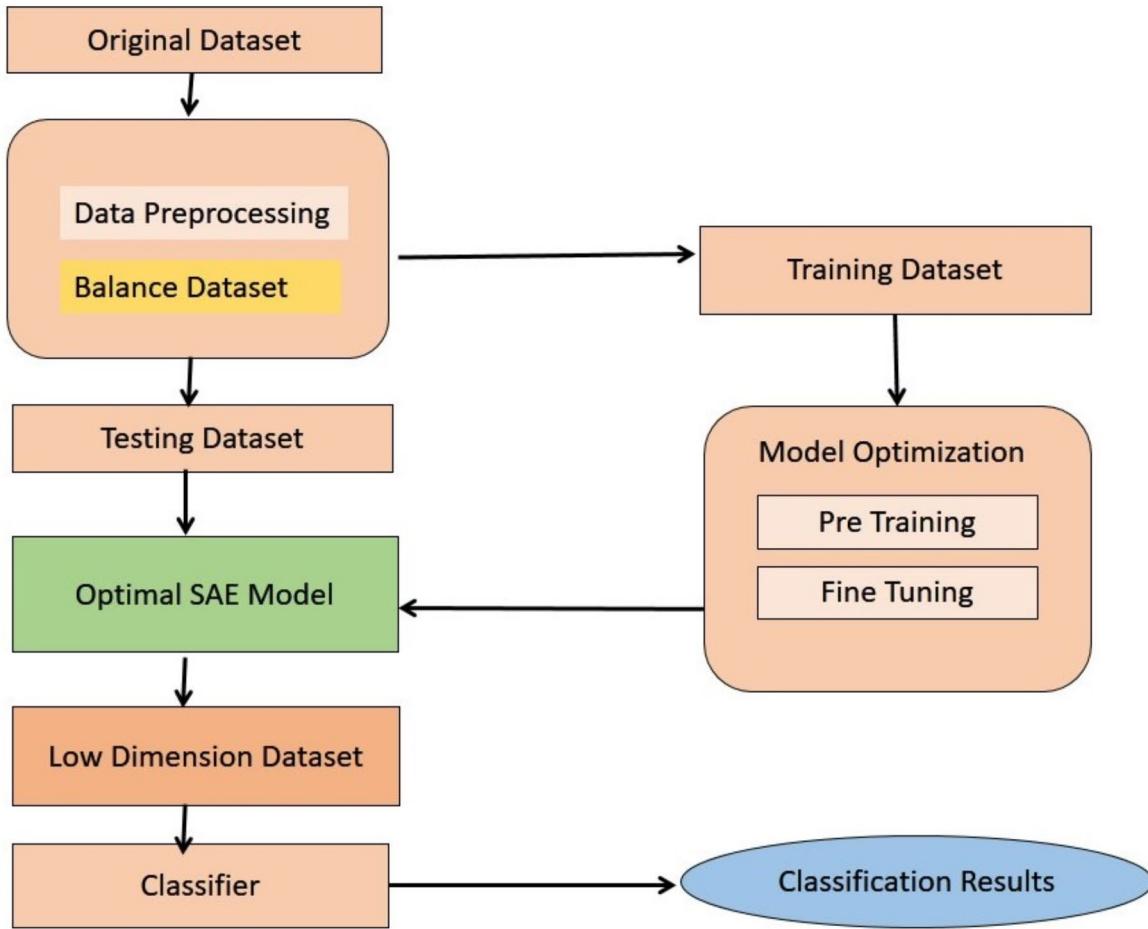


Fig. 11. Developed framework of diabetic retinopathy classification.

- *Random shifting horizontally and vertically,*
- *Flipping images horizontally.*

These transformations create new variations of the existing images, which effectively boost the sample size for each underrepresented class. By applying these techniques, each underrepresented class (e.g., “Severe level 3” and “PDR level 4”) reaches a balanced count. After augmentation and downsampling, each class is set to contain 500 images, ensuring equal representation across all DR stages during model training and preventing the model from becoming biased towards the “Normal level 0” class. Additionally, the variations introduced through augmentation improve the model’s ability to generalize across diverse image conditions, which can enhance classification accuracy for all classes. This balanced dataset ensures the model has an equal opportunity to learn features from each class, reducing bias towards larger classes and improving overall performance, particularly for underrepresented classes.

Training enhanced stacked auto-encoders

To classify the fundus images into five classes, we use stacked auto-encoders with multiple hidden layers to solve this problem. We train the network with multiple hidden layers but only one layer at a time to get effective training results. In the first place by using auto encoders, we train hidden layers one by one in an unsupervised manner. Furthermore, we need a classifier to achieve classification results thereby we train the ReLU layer. All these layers will be joined together forming stacked auto-encoders. Now this whole model needs to be trained but in a supervised manner.

After training the model we compute the results on test images. Test images will be arranged into a matrix to use in the stacked network. Earlier we discussed the two phases: pre-training and fine-tuning. When the confusion matrix is formed after test images, we need to back-propagate the network, which is fine-tuning. It is the retraining of the neural network by training data in a supervised manner.

Proposed algorithm development

The proposed algorithm, as summarized in Algorithm 2, takes retinal fundus images (X, Y), where Y contains five classes: Normal, Mild, Moderate, Severe, and PDR. The output is a classification of images into these five categories after training a stacked auto-encoder model. The main steps of the algorithm include pre-processing, training the first and second auto-encoders to capture key features, training a ReLU layer to refine activation, and finally stacking these layers to form the final network. For further details on the algorithm, please see Appendix C.

Experiments and discussion

Experimental setup

The proposed model, Stacked Auto-Encoder, has been trained on an Intel Core i5 8th Gen-7700 quad-core processor, 16GB RAM, and 64-bit Windows 10 operating system. Anaconda 3 packages used for training and testing purposes include Jupyter Notebook, Keras, and TensorFlow.

Dataset

The proposed model has been trained and tested on the KAGGLE dataset (<https://www.kaggle.com/competition/s/diabetic-retinopathy-detection>, last access: 8 September 2023), which contains 35126 images with dimensions 3888×2951 . This dataset contains five classes, four representing different stages of diabetic retinopathy (mild, moderate, severe, and proliferative) and one representing healthy (normal) images. The distribution of five classes is detailed in Appendix D, Dataset information, Table 1.

Parameters

The training parameters for the stacked auto-encoder are provided in Appendix E, Table 2. These parameters were chosen to optimize the training process for diabetic retinopathy detection and classification, aiming to achieve the best possible performance on the retinal images dataset.

Experimental results

In this paper, results are obtained before enhancing the stacked auto-encoders and after optimizing the model as well. Below are the graphs, which show the accuracy results of a model before optimization.

Before applying data augmentation and dropout techniques, we can see in Figs. 12 and 13 that the accuracy of the model is low; whereas, in Fig. 14, we can see that there is some improvements in the accuracy after applying data augmentation and dropout techniques.

To bring further improvements in model performance, a modified architecture of the deep learning model called enhanced stacked auto-encoder (e-SAE) is built, which overcomes the loopholes of previously applied models as well. The enhancement has been introduced by the dropout mechanism, which is used to reduce the overfitting problem of the standard SAE that leads to poor performance of the model during the testing phase, causing less accuracy. The built model is applied to the KAGGLE dataset. The dataset is split into different ratios i.e., 50:50, 60:40, 70:30, and 75:25 for training and testing.

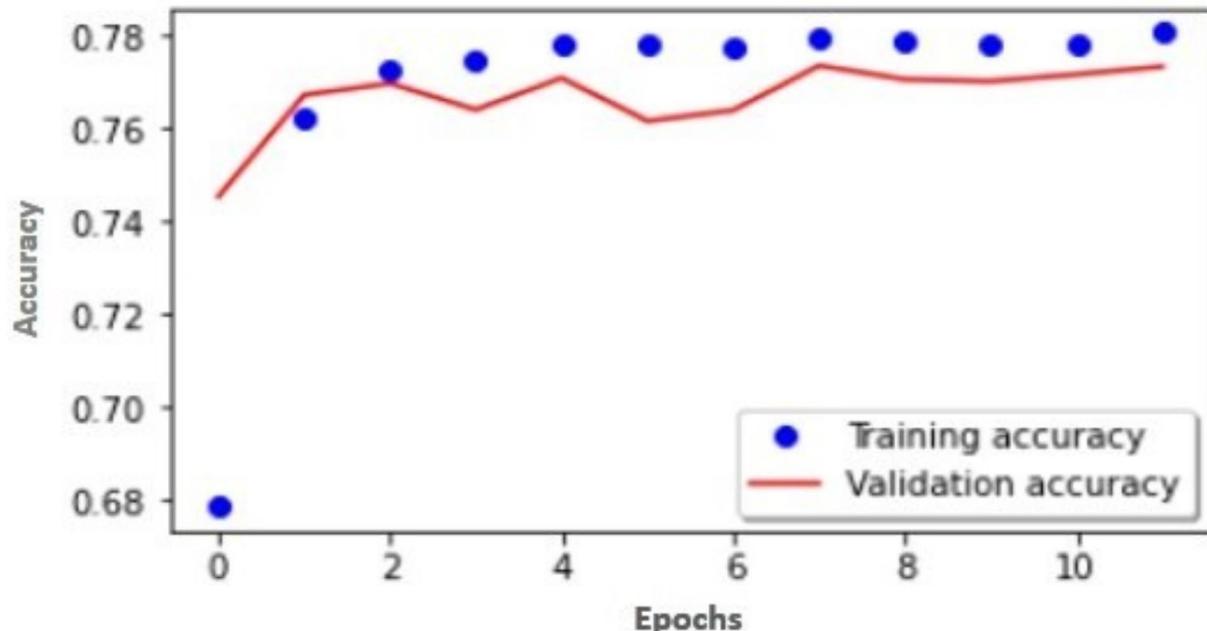


Fig. 12. Accuracy of non-augmented data.

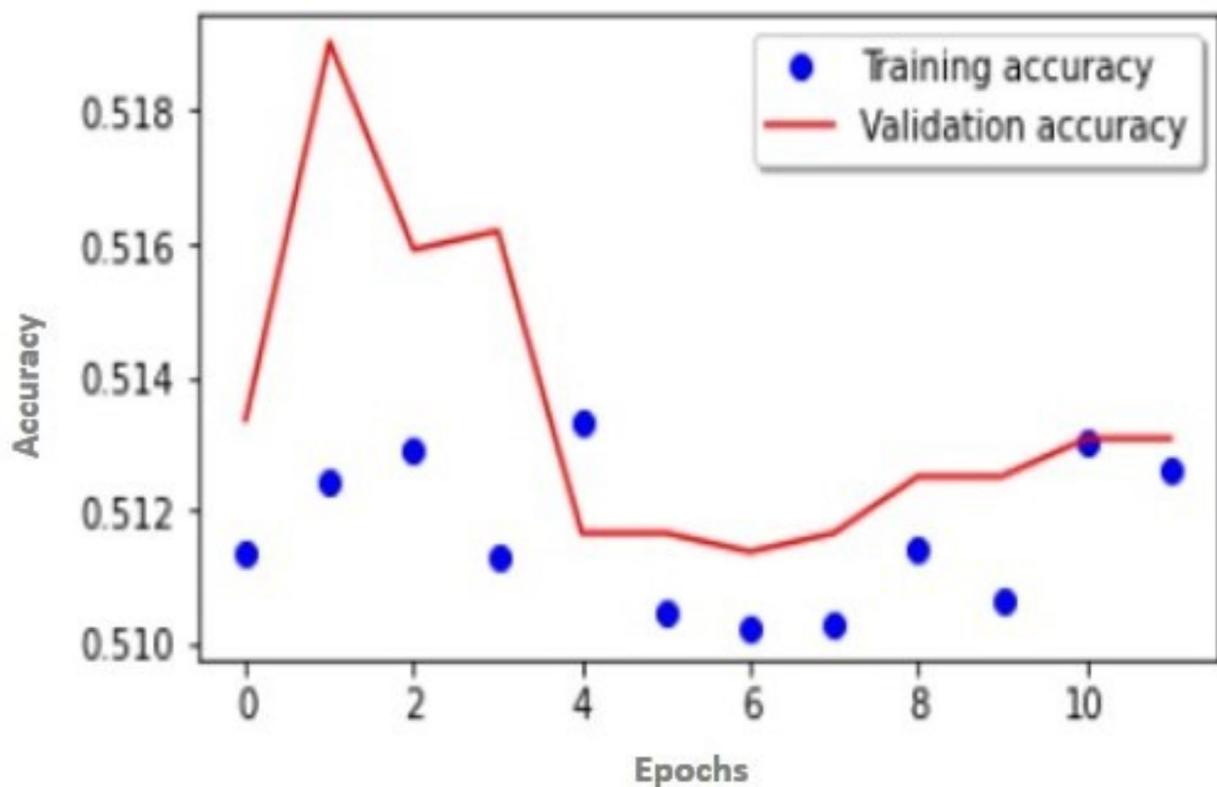


Fig. 13. Accuracy without dropout.

The testing mechanism in the study, which involves splitting the dataset into various training and testing ratios, differs from cross-validation. In the train-test splitting approach, the dataset is divided into distinct subsets for training and testing, allowing for the model's performance evaluation on independent data portions⁴⁴. Conversely, cross-validation partitions the dataset into multiple subsets or folds, with each fold serving as both a training and testing set iteratively.

In mathematical terms, let's denote the dataset as D consisting of n data points. In train-test splitting, the dataset is partitioned into two subsets: the training set, train D_{train} , and the testing set, test D_{test} . Let's say the training set contains m data points, and the testing set contains $n - m$ data points. The model is trained on train D_{train} and evaluated on test D_{test} . This process is typically done once, resulting in a single estimate of the model's performance metrics.

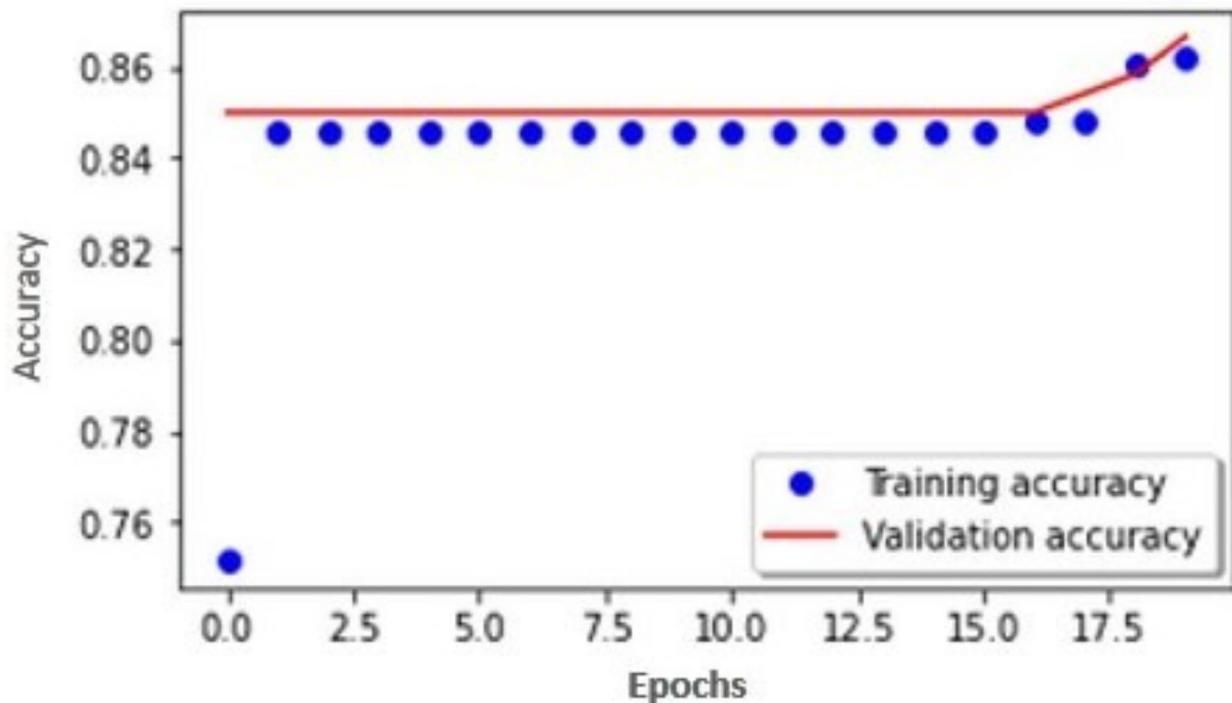
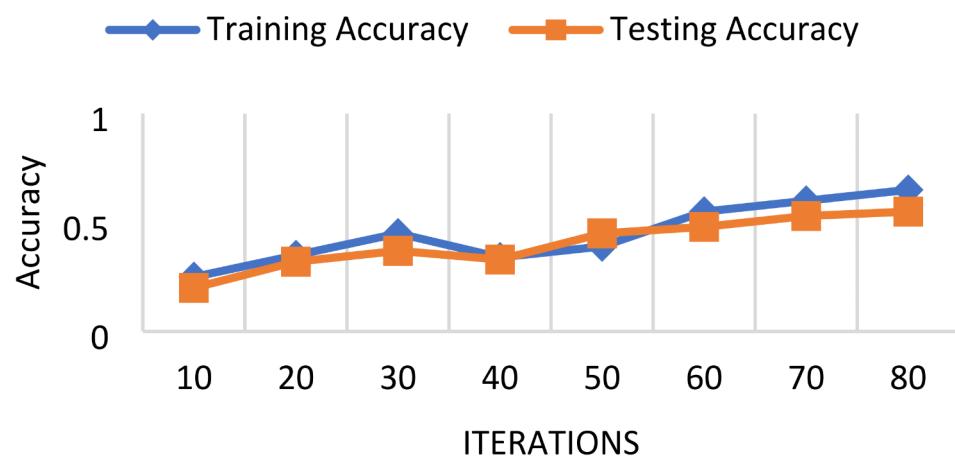
On the other hand, in k -fold cross-validation⁴⁴, the dataset is divided into k equal-sized folds $1, 2, \dots, D1, D2, \dots, Dk$. The model is trained k times, each time using $k - 1$ folds for training and the remaining fold for testing. This ensures that each data point is used for both training and testing, and the process is repeated until each fold has been used as the test set exactly once. The final performance metric is then computed by averaging the metrics obtained from each iteration.

In our study, we selected 80 epochs for training based on empirical results and prior literature, which suggested that this number provided sufficient opportunity for the model to learn the patterns in the data while balancing training time. We monitored the training and validation loss throughout the process to ensure the model was not overfitting. Additionally, early stopping was considered as a regularization technique to prevent overfitting. However, we did not implement it in the final model due to the observed stability of the validation performance over the course of training. We found that the model demonstrated a steady decrease in the validation loss, suggesting that further training up to 80 epochs was beneficial for the learning process.

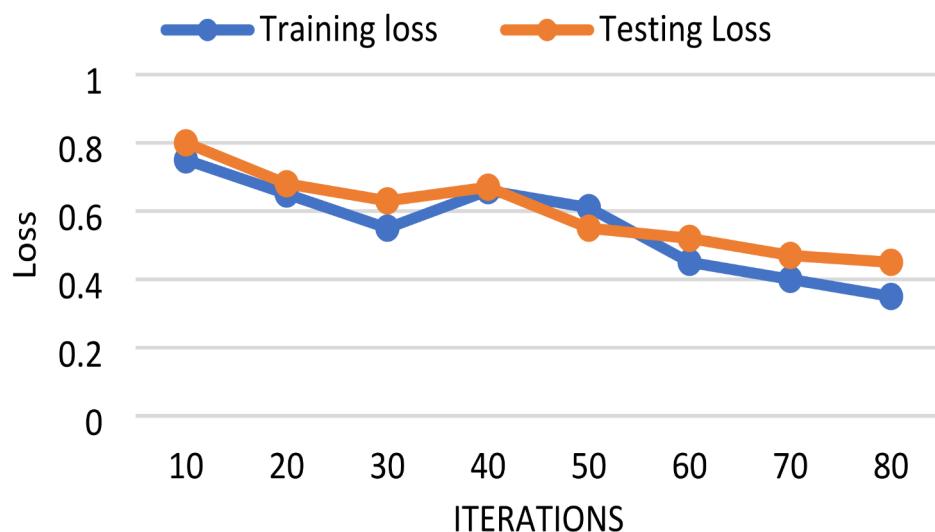
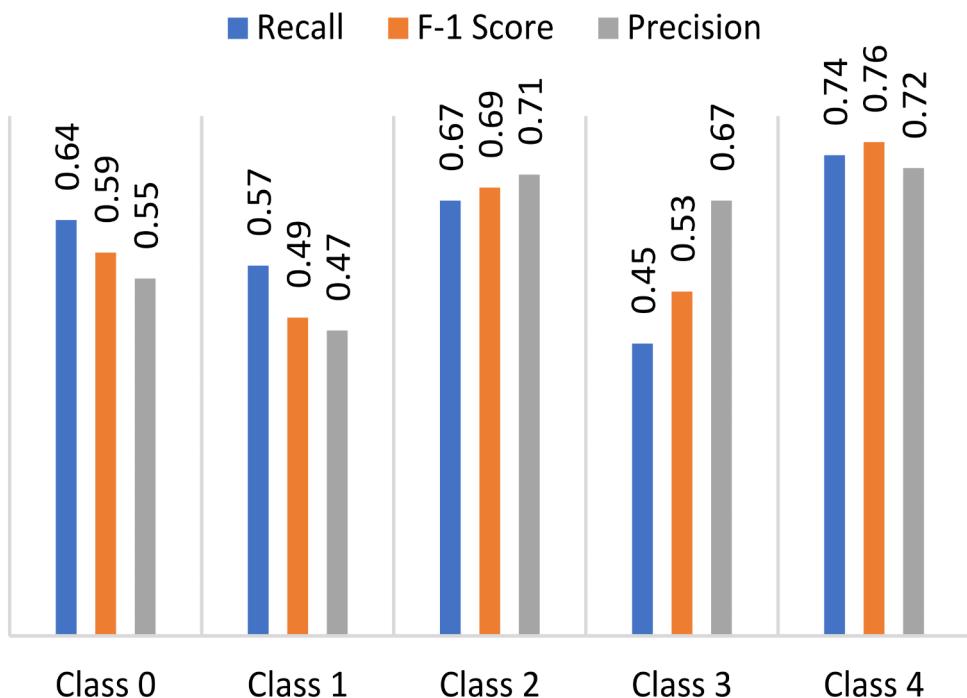
As, earlier it is discussed we have trained and tested the balanced dataset on different ratios. The results obtained in terms of accuracy, loss, and class-wise performance metrics including Recall, F1-Score, and Precision are calculated. Besides that, the Class-wise Receiver Operating Curve is displayed.

Results for 50:50 training/testing split

Figures 15 and 16 demonstrate the accuracy and loss of using the 50:50 splitting ratio of the dataset. Moreover, Figs. 17 and 18 illustrate the performance metrics and receiver operating characteristics (ROC) curves of the classes for testing the same splitting ratio.

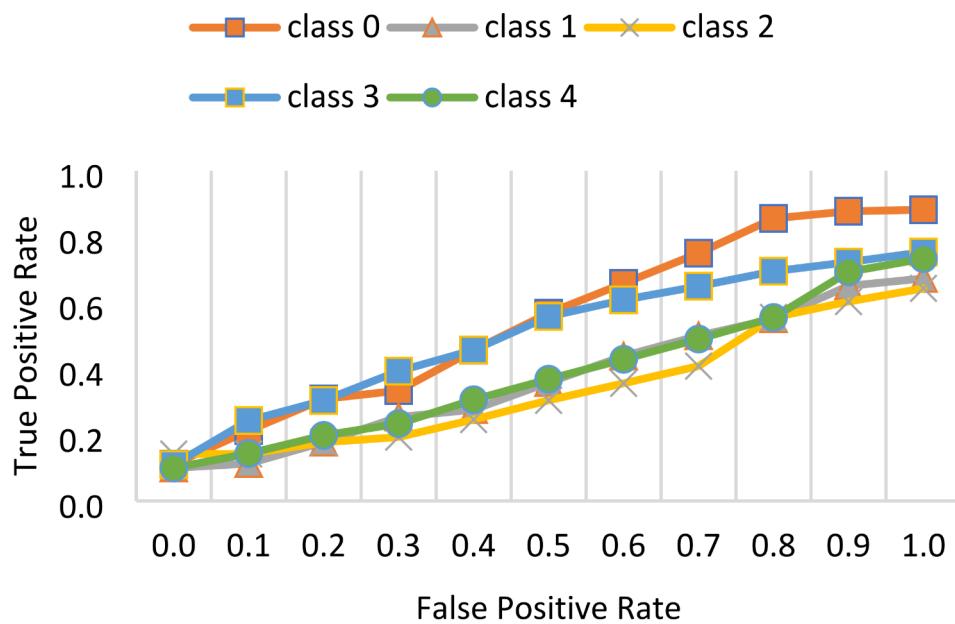
**Fig. 14.** Accuracy of augmented data and dropout.**Fig. 15.** Accuracy at 50:50.

At first, a balanced dataset was split into a ratio of 50:50. The Model has been trained and tested into 80 iterations. As shown in Fig. 15, the gap in the curves of both training and testing accuracy is small, which gives that there is no overfitting in the training phase. The highest accuracy achieved during the training of the model is 65% while the testing accuracy is 55%. The graph shows the positive behavior that deflects the model loss both in training and testing, presented in Fig. 16. Figure 17 gives the performance metrics of the model in terms of Recall, F1-Score, and Precision. The results values of these performance metrics are given in Table 1. From Fig. 18, we can provide a more comprehensive investigation of evaluating the performance of the model.

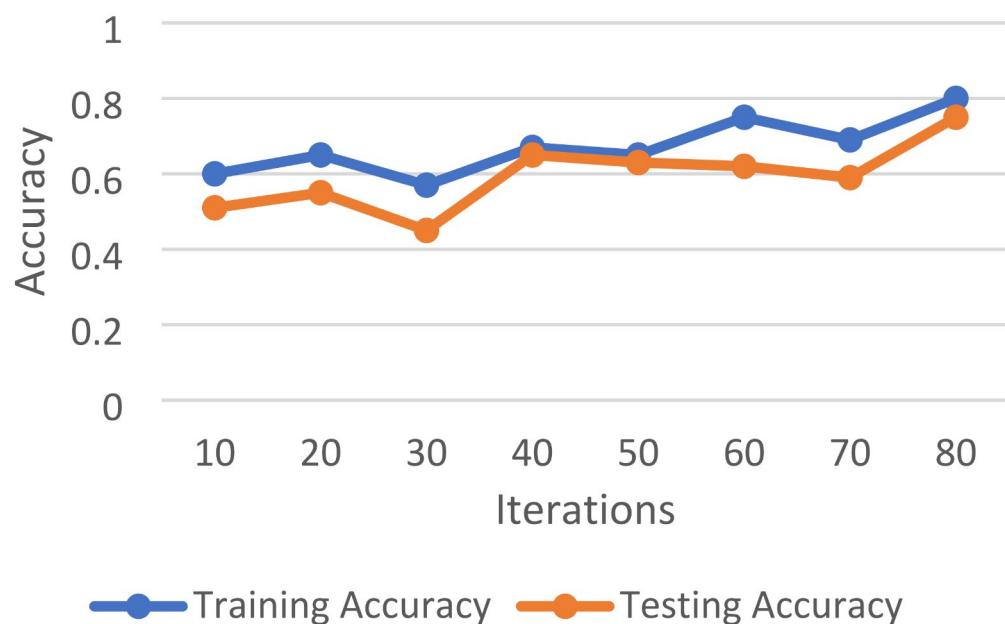
**Fig. 16.** Loss at 50:50.**Fig. 17.** Performance metrics of a balanced dataset at 50:50.

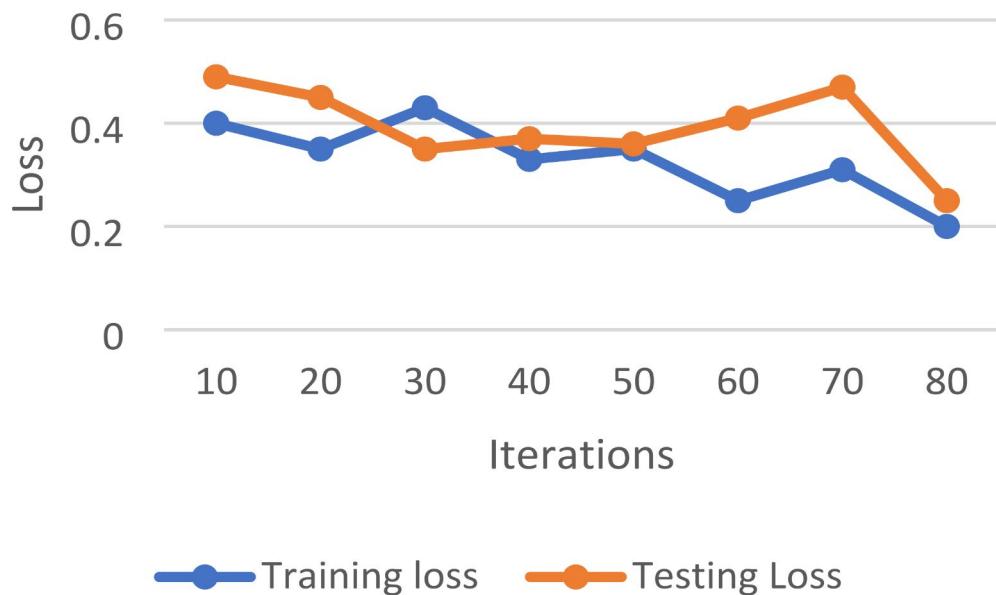
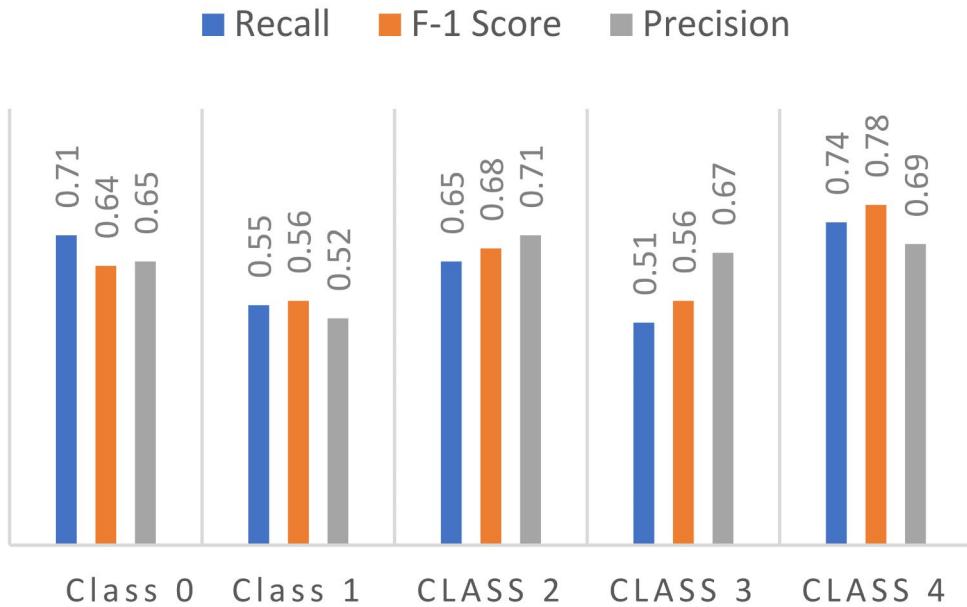
Results for 60:40 training/testing split

The balanced dataset is split into a ratio of 60:40. The model is trained and tested into 80 iterations. Figure 19 shows the model accuracy of both trained and tested models. The highest accuracy achieved during the training of the model is 80% while the testing accuracy is 75%. The graph shows the positive behavior that deflects the model loss both in training and testing, presented in Fig. 20. Figure 21 provides the class-wise performance metrics Recall, F1-Score, and Precision of a model. The results obtained are given in Table 2 below. ROC of all the classes (0–4) have been shown in Fig. 22. Table 2 lists the results values of Recall, F1-Score, and Precision, presenting the performance of the model for classifying the classes of diabetic retinopathy of 60:40 splitting ratio of the dataset.

**Fig. 18.** ROC curves of classes (0–4) at 50:50.

	Class 0	Class 1	Class 2	Class 3	Class 4
Recall	0.64	0.57	0.67	0.45	0.74
F-1 Score	0.59	0.49	0.69	0.53	0.76
Precision	0.55	0.47	0.71	0.67	0.72

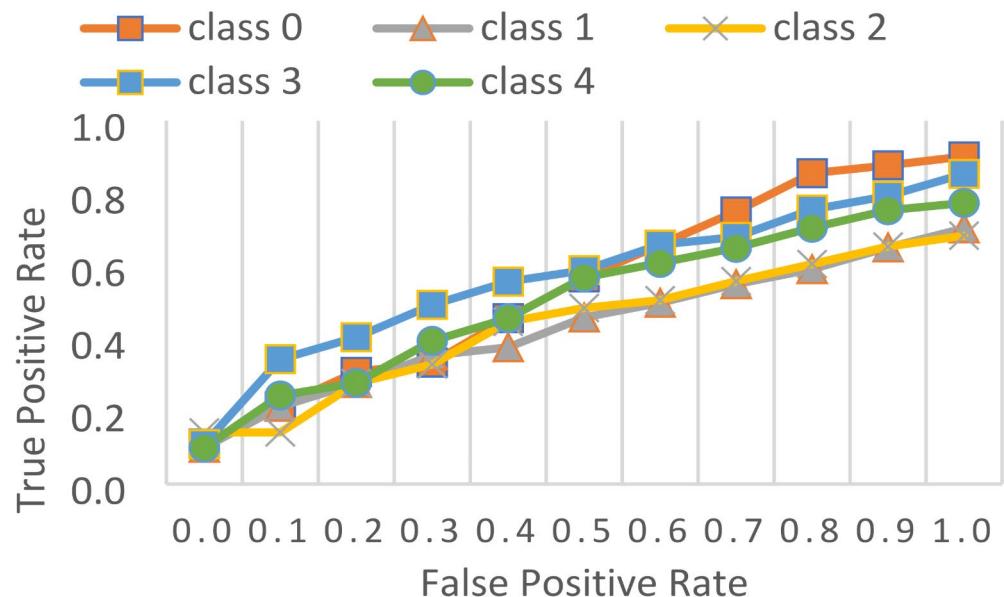
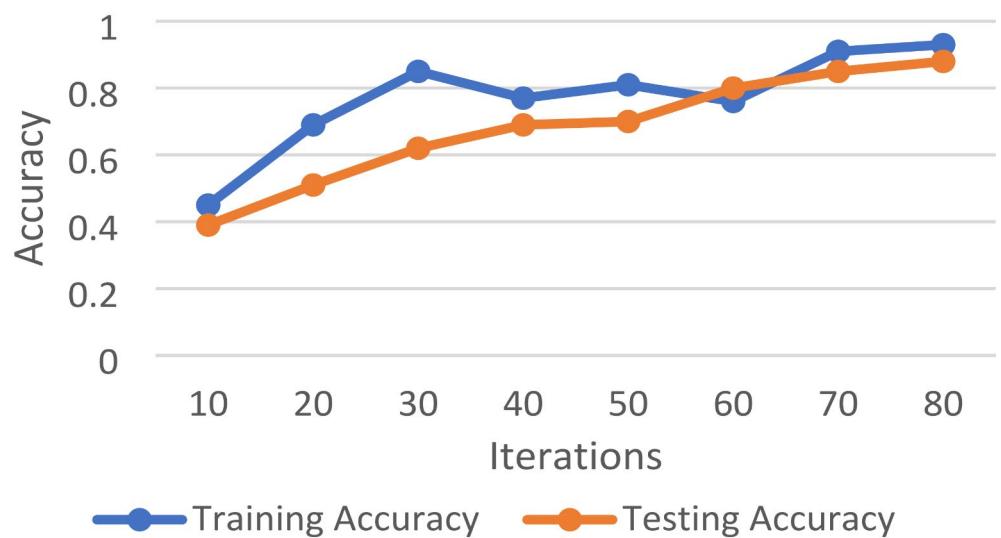
Table 1. Evaluation results of class-wise (0–4) classification for the 50:50 splitting ratio of the dataset.**Fig. 19.** Accuracy at 60:40.

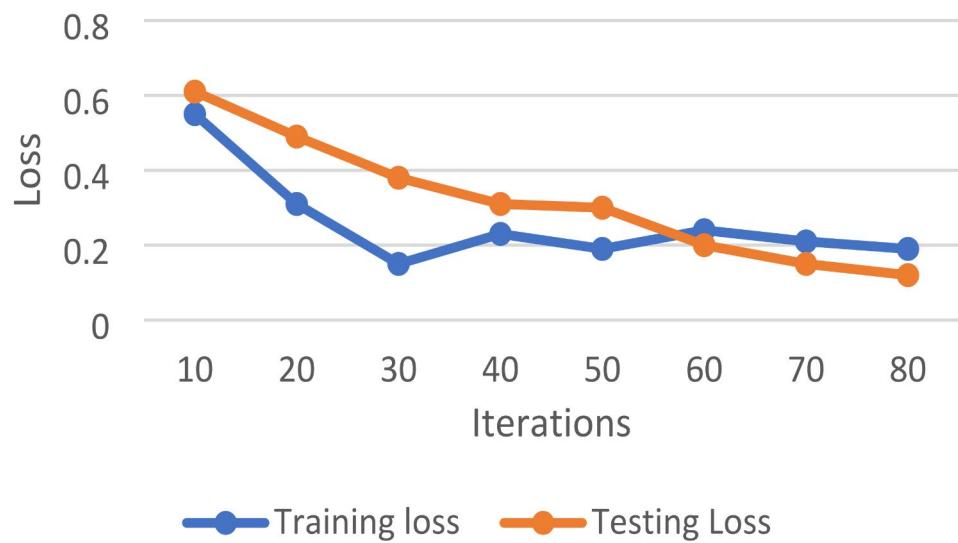
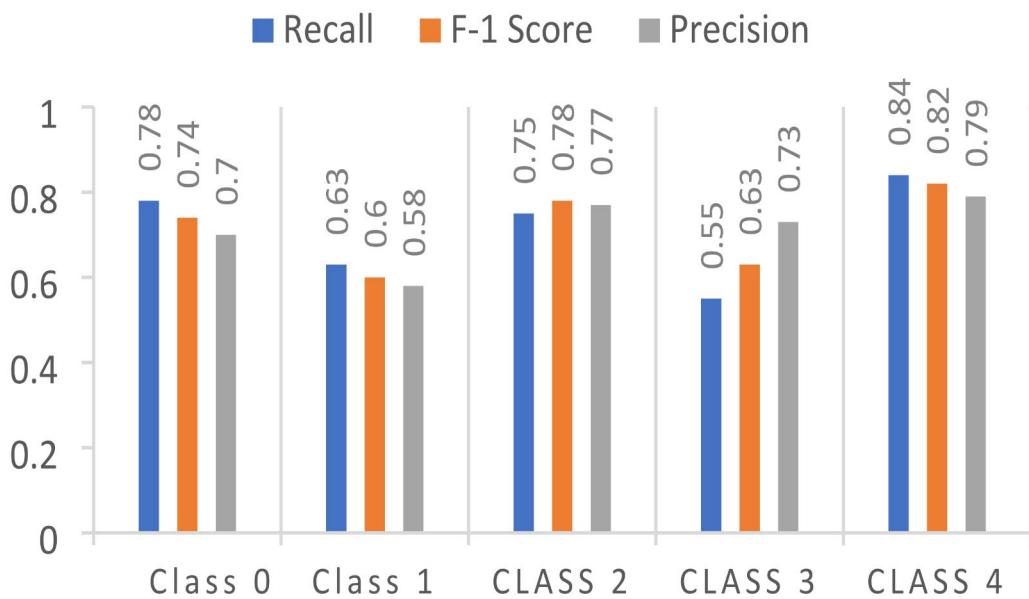
**Fig. 20.** Loss at 60:40.**Fig. 21.** Performance metrics of all classes (0–4) at 60:40.

Results for 70:30 training/testing split

The balanced dataset has been split into a ratio of 70:30. The Model has been trained and tested into 80 iterations. Figure 23. Shows the model accuracy of both trained and tested models. The highest accuracy achieved during the training of the model is 93% while the testing accuracy is 88%. The graph shows the positive behavior that deflects the model loss both in training and testing, presented in Fig. 24. Figure 25 gives the class-wise performance metrics of Recall, F1-Score, and Precision of a model. The ROC curves of all the classes (0–4) are also visualized in Fig. 26. Furthermore, Table 3 presents the results values of Recall, F1-Score, and Precision, presenting the performance of the model for classifying the classes of diabetic retinopathy of 70:30 splitting ratio of the dataset.

	Class 0	Class 1	Class 2	Class 3	Class 4
Recall	0.71	0.55	0.65	0.51	0.74
F-1 Score	0.64	0.56	0.68	0.56	0.78
Precision	0.65	0.52	0.71	0.67	0.69

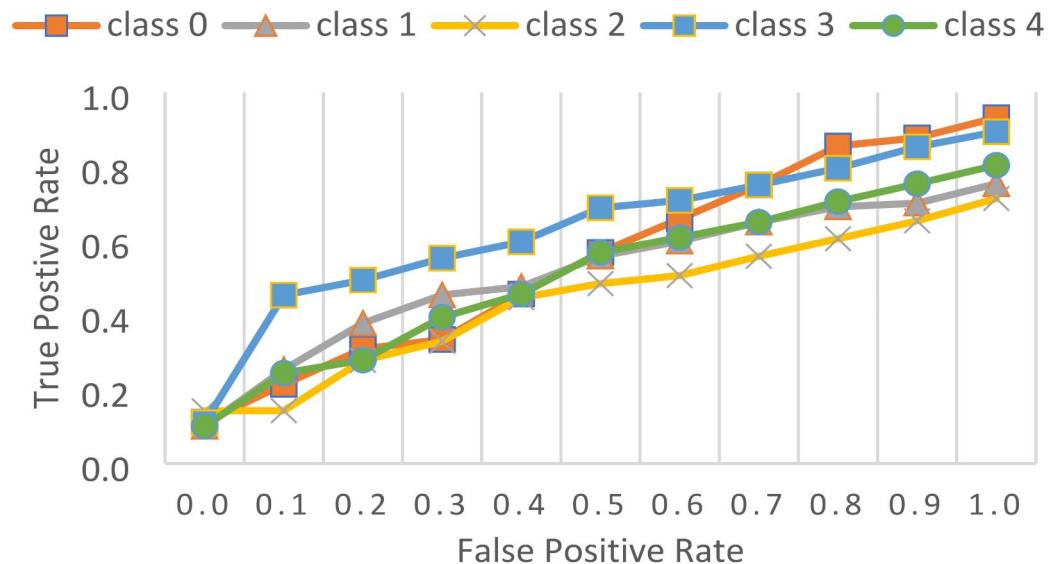
Table 2. Evaluation results of class-wise (0–4) classification for the 60:40 splitting ratio of the dataset.**Fig. 22.** ROC curves of classes (0–4) at 60:40.**Fig. 23.** Accuracy at 70:30.

**Fig. 24.** Loss at 70:30.**Fig. 25.** Performance metrics of a balanced dataset at 70:30.

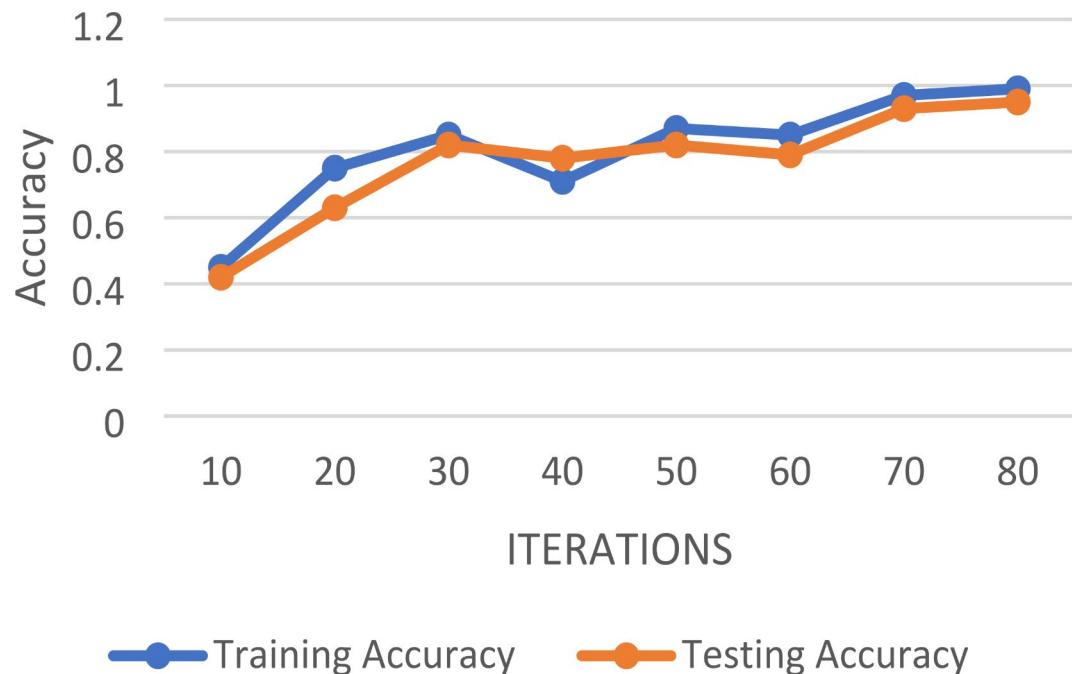
Results for 75:25 training/testing split

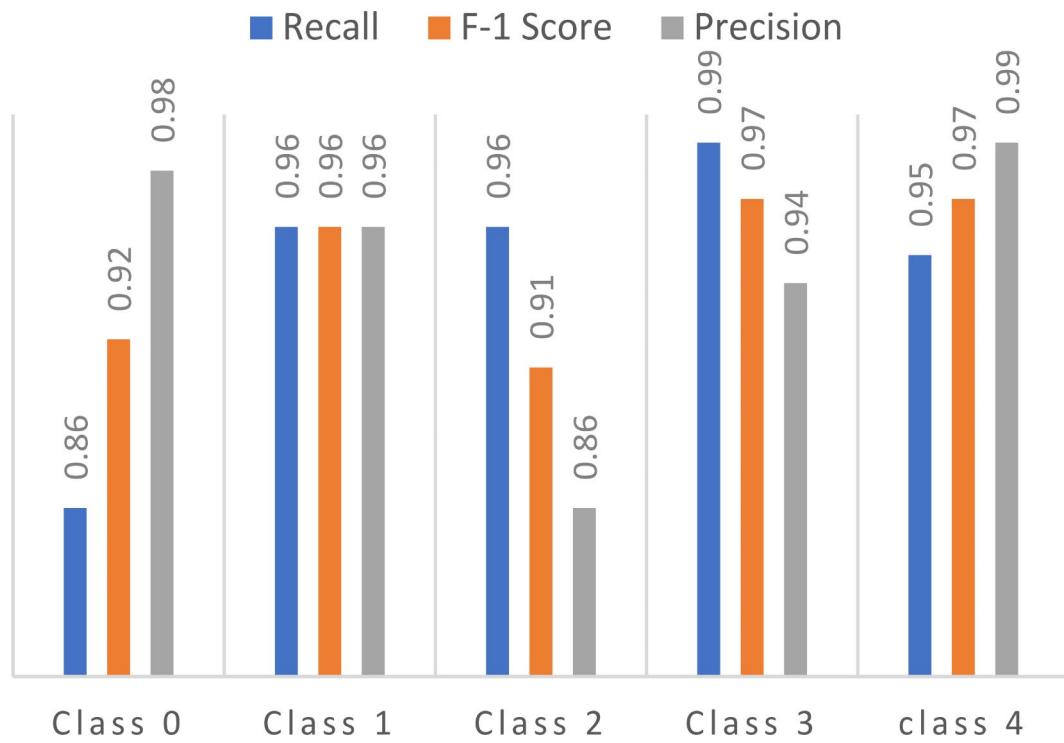
Now, the balanced dataset has been split into a ratio of 75:25. The Model has been trained and tested into 80 iterations. Figure 27 shows the model accuracy of both the trained and tested model. The highest accuracy achieved during the training of the model is 93% while the testing accuracy is 88%. The graph shows the positive behavior that deflects the model loss both in training and testing, presented in Fig. 28. Figure 29 gives the class-wise performance metrics Recall, F1-Score, and Precision of a model. The obtained results of Recall, F1-Score, and Precision are given in Table 4. The ROC curves of all the classes (0–4) are shown in Fig. 30.

It is important to clarify the distinction between testing accuracy and validation accuracy. Testing accuracy refers to the performance of the model evaluated on an unseen dataset that was not used during the training process. This is evaluated on the test subset derived from the train-test split ratios (e.g., 50:50, 60:40, 70:30, and 75:25). This is distinct from validation accuracy, which was monitored during training to fine-tune hyperparameters and avoid overfitting. The reported accuracy values, such as 88% for the 70:30 split, refer to testing accuracy obtained on the independent test set, ensuring valid comparisons with other studies. This

**Fig. 26.** ROC curves of classes (0–4) at 70:30.

	Class 0	Class 1	Class 2	Class 3	Class 4
Recall	0.78	0.63	0.75	0.55	0.84
F-1 Score	0.74	0.60	0.78	0.63	0.82
Precision	0.70	0.58	0.77	0.73	0.79

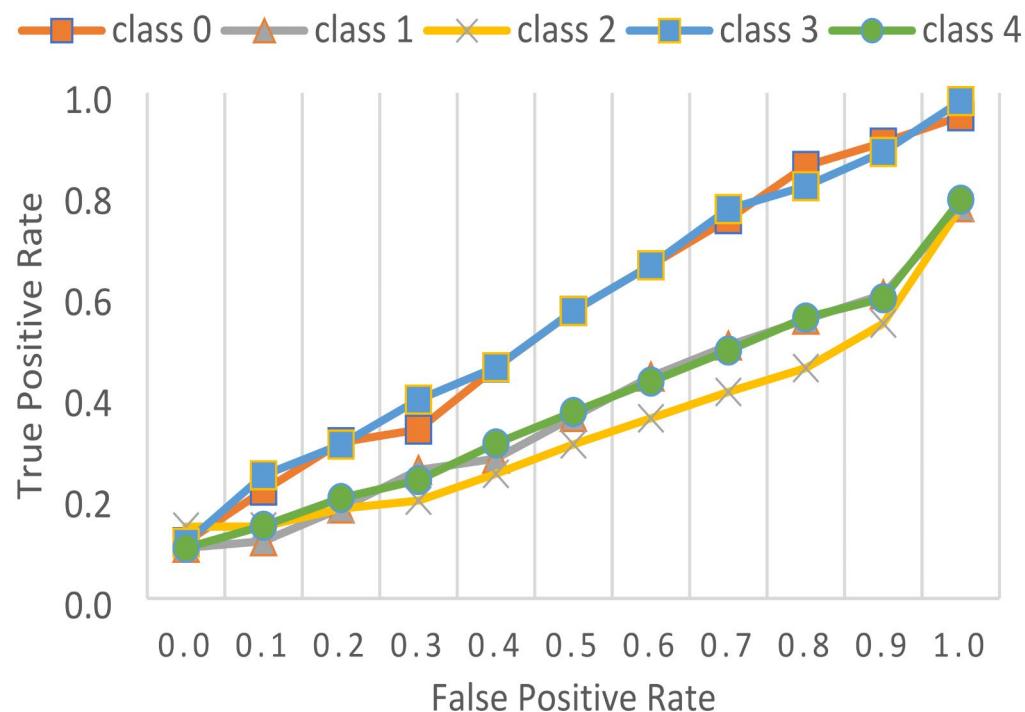
Table 3. Evaluation results of class-wise (0–4) classification for the 70:30 splitting ratio of the dataset.**Fig. 27.** Accuracy at 75:25.

**Fig. 28.** Loss at 75:25.**Fig. 29.** Performance metrics of all classes (0–4) at 75:25.

distinction aligns with standard practices to ensure that reported metrics truly reflect the model's generalization capability.

The evaluation of the model across different training/testing ratios (50:50, 60:40, 70:30, and 75:25) demonstrates its robustness and ability to generalize as the proportion of training data increases. At smaller ratios, such as 50:50, the model shows relatively lower performance, with training accuracy reaching 65% but testing accuracy dropping to 55%. This indicates that with limited training data, the model struggles to generalize well to unseen data, potentially due to overfitting on the small training set. As the training ratio

	Class 0	Class 1	Class 2	Class 3	Class 4
Recall	0.86	0.96	0.96	0.99	0.95
F-1 Score	0.92	0.96	0.91	0.97	0.97
Precision	0.98	0.96	0.86	0.94	0.98

Table 4. Evaluation results of class-wise (0–4) classification for the 75:25 splitting ratio of the dataset.**Fig. 30.** ROC curves of classes (0–4) at 75:25.

increases to 60:40, accuracy improves, with training accuracy reaching 80% and testing accuracy at 75%. With higher training ratios, such as 70:30 and 75:25, the model's performance shows significant improvement, with the 70:30 split achieving training accuracy of 93% and testing accuracy of 88%.

The results indicate that the model's performance improves with larger training datasets, demonstrating its robustness and enhanced generalization ability. The 75:25 split achieves the highest accuracy and recall, but improvements plateau after the 70:30 split, suggesting diminishing returns from further increases in training data. Importantly, the model does not show signs of overfitting, consistently performing well across different splits. The 70:30 ratio appears to strike a good balance, providing sufficient training data while maintaining robust evaluation on the testing set. Overall, the analysis suggests that the model is both effective and generalizable, with no significant issues of underfitting or overfitting across the training/testing ratios.

Validation on an external dataset

To demonstrate the generalizability and robustness of our model, we validated it on an independent external dataset. The Diabetic Retinopathy Dataset from Kaggle⁵⁰ was selected for this purpose. This dataset contains the same five classes as our training dataset: one healthy class and four diabetic retinopathy (DR) stages (mild, moderate, severe, and proliferative). We selected 100 images from each class (500 images in total) for testing. The best model developed during the 70:30 train-test split was used for validation.

Dataset details

- *Classes:* Healthy, Mild, Moderate, Severe, Proliferative DR.
- *Images per Class:* 100 images for each class.
- *Preprocessing:* Consistent preprocessing steps, including resizing, normalization, and augmentation, were applied to ensure the external dataset matched the format of the training dataset.

Model predictions on external dataset

- *Healthy*: 98 correctly classified, 2 misclassified.
- *Mild*: 90 correctly classified, 10 misclassified.
- *Moderate*: 78 correctly classified, 22 misclassified.
- *Severe*: 92 correctly classified, 8 misclassified.
- *Proliferative*: 85 correctly classified, 15 misclassified. The distribution of the predictions and misclassifications demonstrates that the model performs consistently across all classes, with higher accuracy in Healthy and Severe categories, while Moderate and Proliferative DR pose slightly more challenges. This highlights the need for further optimization to handle complex cases more effectively. The predictions made by the model are visualized in the Fig. 31.

Results

The model's performance on the external dataset was evaluated using standard metrics such as accuracy, precision, and recall. The results are summarized below:

- *Accuracy*: The model achieved an overall accuracy of 79.5%, correctly predicting 443 out of 500 test images.
- *Precision*: The precision of the model across all classes was 88.6%, reflecting its ability to minimize false positives and provide reliable predictions.
- *Recall*: The recall was also 88.6%, indicating the model's effectiveness in correctly identifying true positives for all classes. These results highlight the robustness and generalizability of the model when applied to an unseen dataset. Notably, the high precision and recall demonstrate that the model effectively balances minimizing false alarms and identifying actual cases of diabetic retinopathy. The consistent performance across multiple classes validates the model's potential for real-world applications in diabetic retinopathy detection. To contextualize our model's performance, we conducted a comparative analysis with recent prominent studies, as presented in Table 5 below.

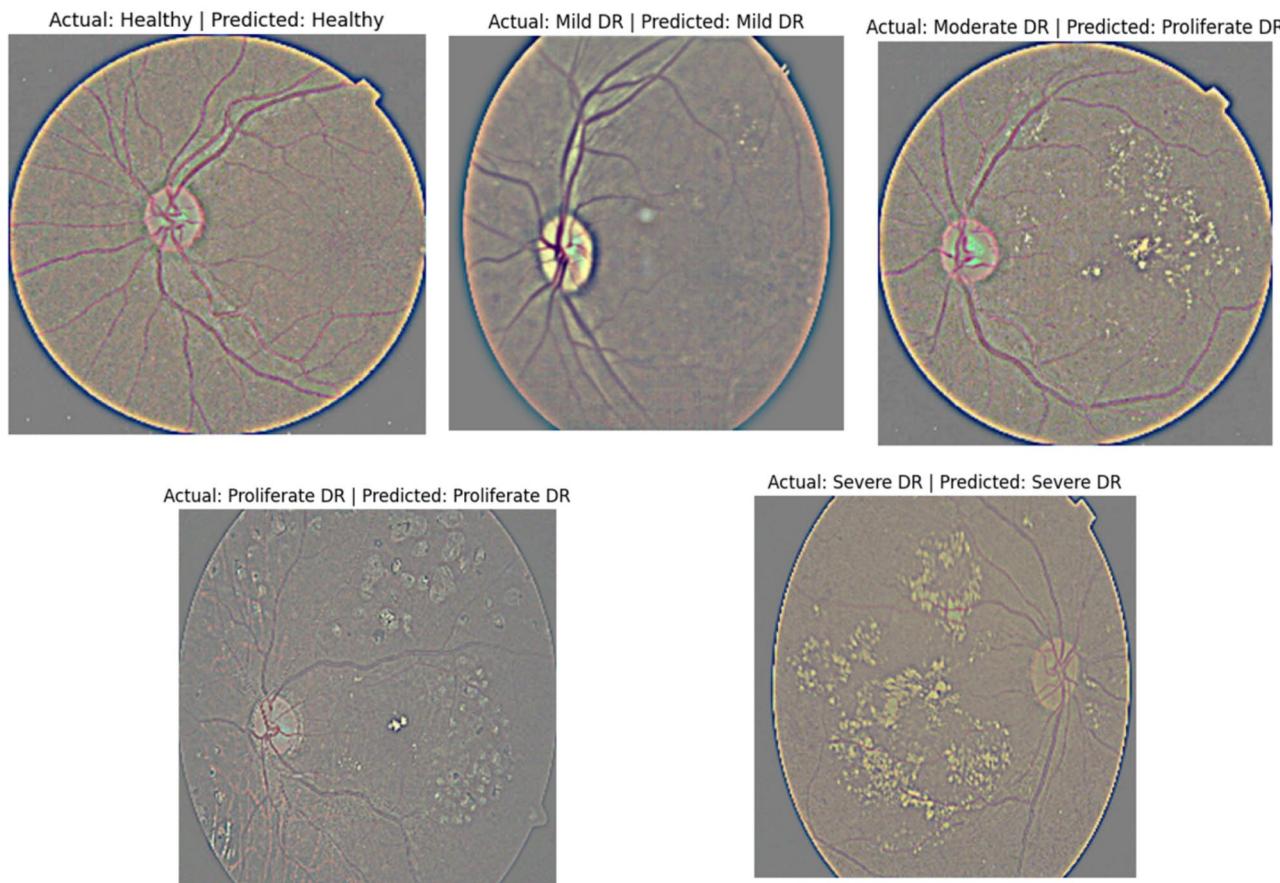
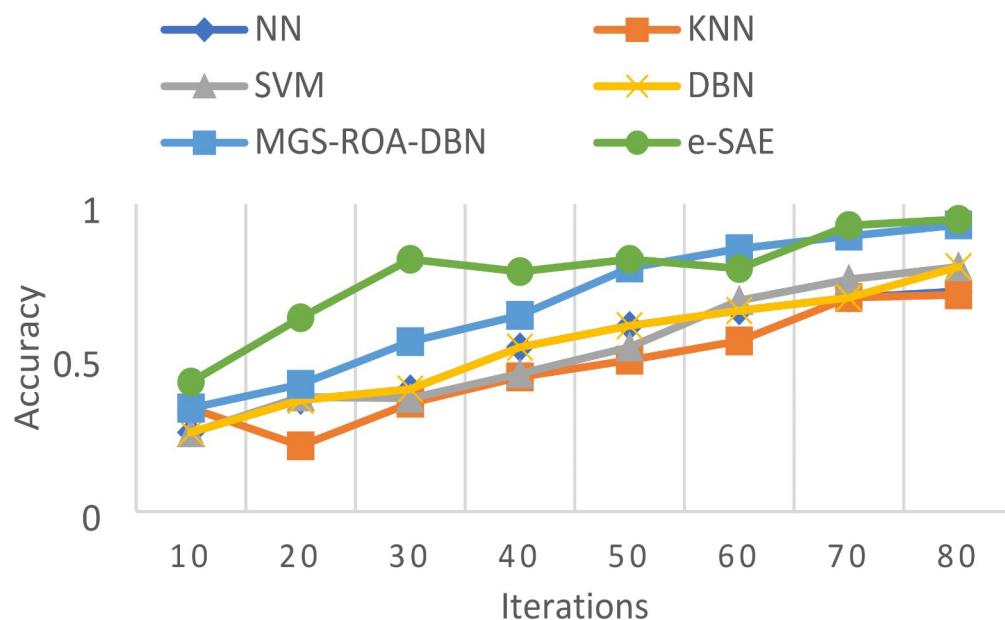


Fig. 31. Predictions Made by the Model on the External Dataset for Each Class (Healthy, Mild, Moderate, Severe, Proliferative DR).

Study	Dataset	Methodology	Analysis type	Performance	
Qummar et al. ²	KAGGLE	Deep CNN models Ensemble(ResNet50, InceptionV3, Xception, Dense121, densenet169)	Normal Mild Moderate Severe Proliferative DR	Accuracy = 80.8% Recall = 51.5% Specificity = 86.72% Precision = 63.85% F1 score = 53.74%	
Luo et al. ¹³	MESSIDOR DRIVE DIARETDBI	Self-Supervised Fuzzy Clustering Network	NO DR DR	Accuracy MESSIDOR: 87.6% DRIVE: 81.7% DIARETDBI: 84.7%	
Zhou et al. ¹⁴	Retinopathy Online Challenge	Sparse Principal component analysis	Micro Aneurysm Non- Micro Aneurysm	Sensitivity against state of art methods: 0.135 ± 0.012	
Cao et al. ¹⁵	DIARETDBI	1-Random Forest with 25 trees 2-Single Hidden layer Neural Networks with 10–100 neurons 3-SVM	Micro Aneurysm Non-Micro aneurysm	Accuracy SVM: 98.5% NN: 97.9% RF: 97.2%	F-measure SVM: 92.6%, NN: 89.4% RF: 88.8%
Qiao et al. ⁴	data port	Deep convolution neural network	Micro aneurysm Non-Micro aneurysm	Sensitivity: 96.5% Specificity: 92.4%	
Adal et al. ¹⁷	Rotterdam Eye Hospital	Retinal change detection using Laplacian of Gaussian.	Retinal Lesions No retinal Lesions	Sensitivity = 80% using SVM	
Zhou et al. ¹⁸	DIARETDBI e-optha EX	Supervised multi-variable classification algorithm	Exudates Non- Exudates	AUC DiaretDBI = 0.970 AUC e-Optha EX = 0.9655	
Dai et al. ²⁰	DIARETDBI	ALEXNET CNN	Micro aneurysm	Recall = 0.878 Precision = 0.997 Accuracy = 0.961 F1 score = 0.934	
Verma et al. ¹⁶	STARE	Random Forest Algorithm	Normal Moderate Non-proliferative	Normal accuracy = 90% Moderate and Non-proliferative accuracy = 87.5%	
Adly et al. ⁵¹	EyePACS	Binary tree-based multiclass classifier of CNN	Normal Mild Moderate Severe PDR	Accuracy = 83%	
Meshram et al. ⁵²	KAGGLE	CNN architecture and EfficientNet architecture	0- No DR, 1- Low, 2- Medium, 3- High, 4- Proliferative	EfficientNet Accuracy = 81.81%	CNN Accuracy = 71.41%
Aswathi et al. ⁵³	Messidor dataset	VGG19 InceptionV3 MobileNet ResNet NASNet	Normal Mild Moderate severe	Model	Accuracy
				VGG19 InceptionV3 MobileNet ResNet NASNet	48.83% 48.75% 51.25% 50% 50.62%
Butt et al. ⁵⁴	APOTOS KAGGLE	Hybrid Convolutional Neural Network	No DR Mild Moderate Severe PDR	Accuracy = 89.29%	
Al-Antary et al. ⁵⁵	APOTOS Kaggle, EyePacs Kaggle	Deep learning model (MSA-Net)	No DR Mild Moderate Severe PDR	APOTOS Kaggle,	EyePacs Kaggle
				Accuracy = 84.6%	Accuracy = 87.8%
Dutta et al. ⁵⁶	Kaggle	AlexNet GoogLeNet VGG16 VGG-19 Xception InceptionV3 Finetuned VGG-19 architecture	No DR Mild Moderate Severe PDR	Models	Accuracy
				AlexNet	40.7%
				GoogLeNet	41.68%
				VGG16	63%
				VGG-19	65%
				Xception	70.1%
				InceptionV3	75%
				Finetuned VGG-19 architecture	77%
Menaouer et al. ⁵⁷	Kaggle	Hybrid Models(CNN, VGG16, VGG19)	Normal Mild Moderate Severe Proliferative DR	Accuracy = 90.60% Recall = 95% Precision = 94.66% F1 score = 94%	
Shankar et al. ⁵⁸	MESSIDOR dataset	Synergic deep learning model	Normal Mild Moderate Severe Proliferative DR	Accuracy = 99% Sensitivity = 98% Specificity = 99%	
Continued					

Study	Dataset	Methodology	Analysis type	Performance	
Mohanty et al. ⁵⁹	Kaggle	Hybrid network(VGG16, XGBoost Classifier), DenseNet 121	Normal	Model	Accuracy
			Mild	Hybrid Network	79.5%
			Moderate	DenseNet 121	97.3%
			Severe		
Qureshi et al. ⁶⁰	EyePACS	Active deep learning CNN Architecture	Proliferative DR	Accuracy = 98% Sensitivity = 92.2% Specificity = 95.1% F-measure = 93%	
Proposed	Dataset 1 (Kaggle) Dataset 2 (Kaggle ⁵⁰)	Enhanced Stacked auto-encoders	Normal	Dataset 1	External Dataset 2
			Mild	Accuracy = 88%	
			Moderate	Accuracy = 79.5%	
			Severe		
			Proliferative DR		

Table 5. Comparative analysis of proposed work compared with some recent related works.**Fig. 32.** Accuracy comparison with existing models.

Comparative analysis

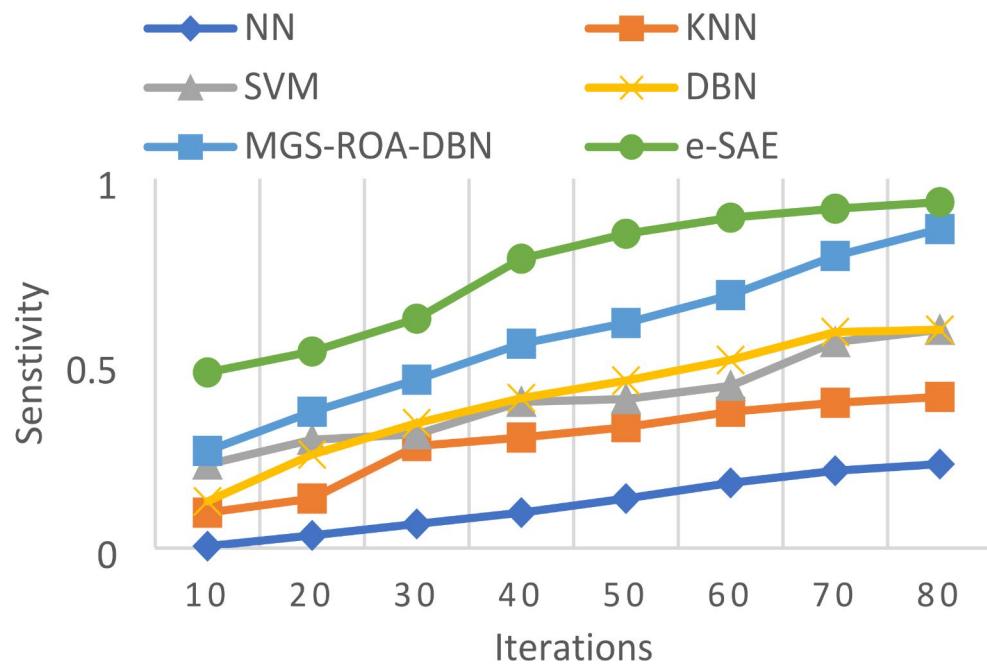
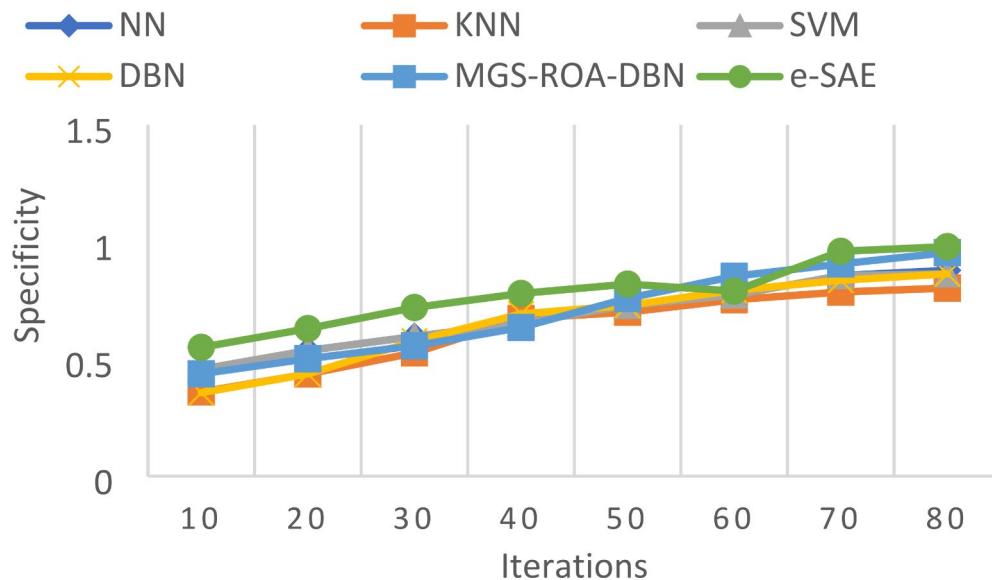
In this subsection, a comparison of the proposed model is given with some existing models such as Neural Networks, K-Nearest Neighbor, Support Vector Machines, Deep Belief Neural Networks, and MGS-ROA-DBN¹². Table 5 summarizes the performance, the dataset, the methodology, and the classified class labels of some recent related works compared with the proposed work. In Figs. 32, 33, 34, 35 and 36, we visualize the comparison results of accuracy, sensitivity, specificity, precision, and F1-Score of the model against the existing models, respectively. As shown in all of these mentioned figures, the e-SAE model achieved the highest results painted with the green color curve at the different iterations.

Time complexity

Training a stacked auto-encoder involves multiple forward and backward passes through the network for each training iteration. During the forward pass, the input data is propagated through the network layer by layer, involving matrix multiplications and activation function computations. The time complexity of the forward pass depends on the size of the input data and the number of neurons in each layer. Let's denote the total number of neurons in all layers as N, and the time complexity of processing a single neuron as O(1). Therefore, the overall time complexity of the forward pass is approximately O(N).

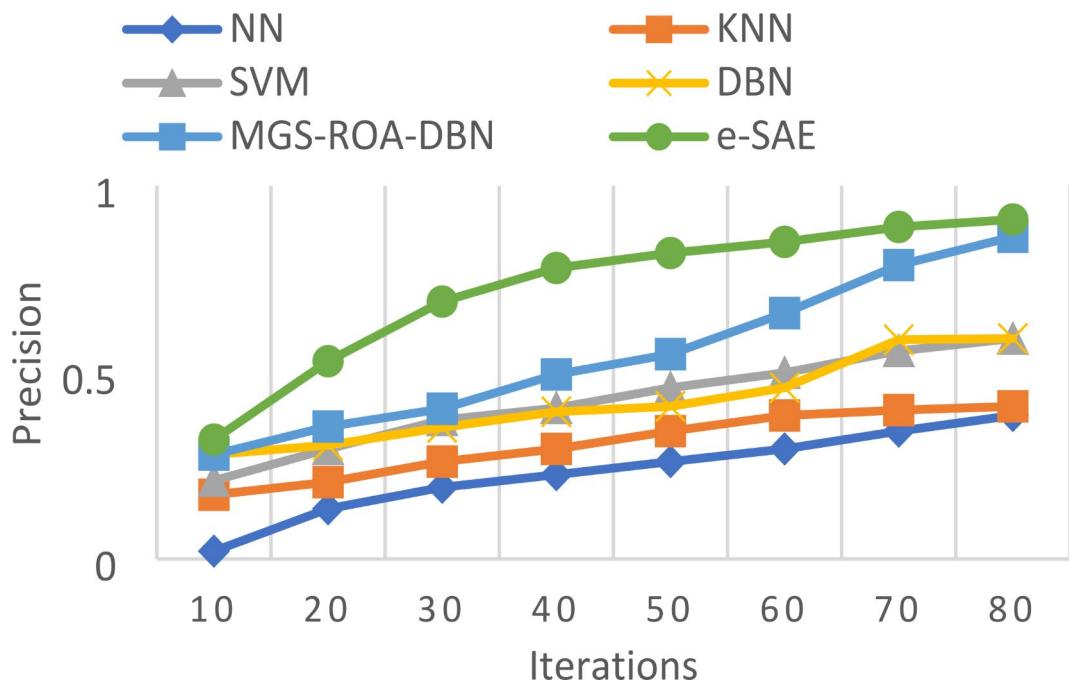
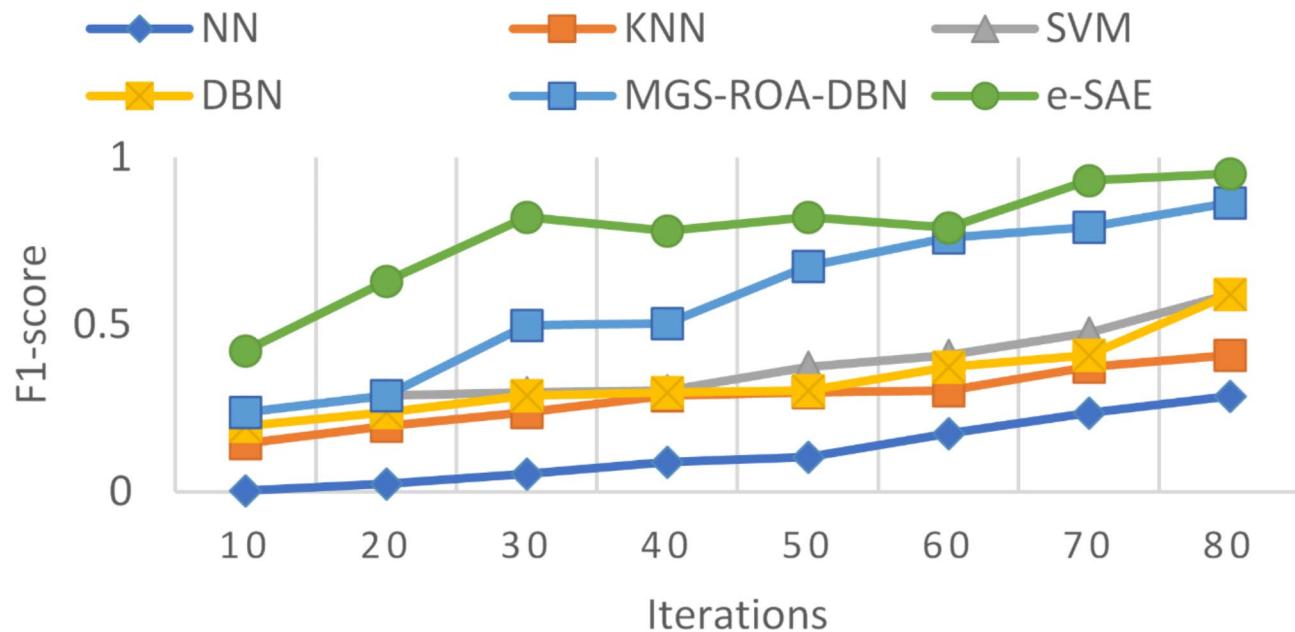
During the backward pass, the gradients of the loss function with respect to the parameters of each layer are computed through backpropagation. Similar to the forward pass, the time complexity of the backward pass is approximately O(N).

Since each auto-encoder in the stack is trained separately, the time complexity of training each auto-encoder depends on the number of layers and the size of the input data. Let's denote the number of layers in each auto-

**Fig. 33.** Sensitivity comparison with existing models.**Fig. 34.** Specificity comparison with existing models.

encoder as L and the size of the input data as M. Therefore, the time complexity of training each auto-encoder is approximately $O(L^*M)$.

The auto-encoders are trained in a step-by-step manner, the overall training time of the stacked auto-encoder is the sum of the training time for each individual auto-encoder. Considering these factors, the overall time complexity of training the proposed stacked auto-encoder can be expressed as approximately $O(N + L^*M)$, where N is the total number of neurons in all layers of the network, L is the number of layers in each auto-encoder, and M is the size of the input data.

**Fig. 35.** Precision comparison with existing models.**Fig. 36.** F1-Score comparison with existing models.

Conclusions

Diabetes is a very common disease among masses irrespective of age group due to excessive intake of sugar in the diet. It has adverse effects on many other organs including kidneys, heart, gums, feet, nerves, and eyes. Our research focuses on the eye-related disease called Diabetic Retinopathy, in which the retina of the eye is damaged due to diabetes. It may be because of blurred vision at early stages but as the disease progresses it may lead to permanent vision loss. The disease is classified into five stages: normal, mild, moderate, severe, and proliferative diabetic retinopathy.

Manual methods have been practiced for so long to detect diabetic retinopathy but there are greater chances of human error. Many computer-aided systems have been developed to overcome human error and get accurate results. Machine learning algorithms are widely used in the detection and classification of diabetic retinopathy,

but deep learning algorithms have attracted research attention due to their structure resembling the human brain. The proposed stacked auto-encoders are used to detect and classify the diabetic retinopathy stages. CNN has been replaced by a stacked auto-encoder because it has been frequently used in the past decade. It depends upon multiple layers which slow down the performance of the system while our proposed technique is more reliable than CNN in terms of time complexity, error minimization, and noise reduction. In previous studies, only 3 stages of the disease have been studied while our system will work on 5 stages of Diabetic Retinopathy. The results show that the proposed algorithm gives more accurate results for the detection of stages of disease than other deep learning algorithms. This will help to early diagnose the stage of the disease and prevent the patient from blindness if treated in time. Results obtained in this proposed model show improved accuracy than other state-of-the-art methods. While the class-wise performance metrics were also evaluated and it is contemplated that class 4 outperforms among all the classes in accuracy, recall, F1-Score, and precision.

In the future, we will try to implement this model on other datasets for evaluation. Besides that, we may use other deep-learning models to make the detection of Diabetic retinopathy more accurate. Moreover, the proposed technique may be used to detect other diseases such as heart disease.

Data availability

The authors confirm that the data supporting the findings of this study are available within the article.

Received: 22 September 2024; Accepted: 6 January 2025

Published online: 20 January 2025

References

- Shankar, K. et al. Automated detection and classification of fundus diabetic retinopathy images using synergic deep learning model. *Pattern Recogn. Lett.* **133**, 210–216 (2020).
- Dutta, S., Manideep, B., Basha, S. M., Caytiles, R. D. & Iyengar, N. Classification of diabetic retinopathy images by using deep learning models. *Int. J. Grid Distrib. Comput.* **11**, 89–106 (2018).
- Qiao, L., Zhu, Y. & Zhou, H. Diabetic retinopathy detection using prognosis of microaneurysm and early diagnosis system for non-proliferative diabetic retinopathy based on deep learning algorithms. *IEEE Access* **8**, 104292–104302 (2020).
- Jagan Mohan, N., Murugan, R., Goel, T., Mirjalili, S. & Roy, P. A novel four-step feature selection technique for diabetic retinopathy grading. *Phys. Eng. Sci. Medicine* **44**, 1351–1366 (2021).
- Yun, W. L. et al. Identification of different stages of diabetic retinopathy using retinal optical images. *Inf. Sci.* **178**, 106–121 (2008).
- Kaveeshwar, S. A. & Cornwall, J. The current state of diabetes mellitus in India. *Australas. Medical J.* **7**, 45 (2014).
- Saeed, F., Hussain, M. & Aboalsamh, H. A. Automatic diabetic retinopathy diagnosis using adaptive fine-tuned convolutional neural network. *IEEE Access* **9**, 41344–41359 (2021).
- Luo, Y., Pan, J., Fan, S., Du, Z. & Zhang, G. Retinal image classification by self-supervised fuzzy clustering network. *IEEE Access* **8**, 92352–92362 (2020).
- Shivappriya, S. & Harikumar, R. Performance analysis of deep neural network and stacked auto-encoder for image classification. *Comput. Intell. Sustain. Syst.: Intell. Sustain. Comput.*, 1–16 (2019).
- Singh, S. K. & Goyal, A. A stack auto-encoders based deep neural network approach for cervical cell classification in pap-smear images. *Recent Adv. Comput. Sci. Commun. (Formerly Recent Patents Comput. Sci.)* **14**, 62–70 (2021).
- Khan, Z. et al. Diabetic retinopathy detection using VGG-NIN a deep learning architecture. *IEEE Access* **9**, 61408–61416 (2021).
- @kaggle. Diabetic Retinopathy Detection. [online] Available at: <https://www.kaggle.com/competitions/diabetic-retinopathy-detection/data> [Accessed 8 Sep 2023].
- Zhou, W., Wu, C., Chen, D., Yi, Y. & Du, W. Automatic microaneurysm detection using the sparse principal component analysis-based unsupervised classification method. *IEEE Access* **5**, 2563–2572 (2017).
- Cao, W., Czarnek, N., Shan, J. & Li, L. Microaneurysm detection using principal component analysis and machine learning methods. *IEEE Trans. Nanobiosci.* **17**, 191–198 (2018).
- Adal, K. M. et al. An automated system for the detection and classification of retinal changes due to red lesions in longitudinal fundus images. *IEEE Trans. Biomedical Eng.* **65**, 1382–1390 (2017).
- Qummar, S. et al. A deep learning ensemble approach for diabetic retinopathy detection. *IEEE Access* **7**, 150530–150539 (2019).
- Raman, V., Then, P. & Sumari, P. Proposed retinal abnormality detection and classification approach: Computer aided detection for diabetic retinopathy by machine learning approaches. In *Proceedings of the 2016 8th IEEE International conference on communication software and networks (ICCSN)*, pp. 636–641 (2016).
- Rakshitha, T., Devaraj, D. & Kumar, S. P. Comparative study of imaging transforms on diabetic retinopathy images. In *Proceedings of the 2016 IEEE international conference on recent trends in electronics, information & communication technology (RTEICT)*, pp. 118–122 (2016).
- Asiri, N., Hussain, M., Al Adel, F. & Alzaidi, N. Deep learning based computer-aided diagnosis systems for diabetic retinopathy: A survey. *Artif. Intell. Medicine* **99**, 101701 (2019).
- Dai, L. et al. Clinical report guided retinal microaneurysm detection with multi-sieving deep learning. *IEEE Trans. Medical Imaging* **37**, 1149–1161 (2018).
- Ghosh, R., Ghosh, K. & Maitra, S. Automatic detection and classification of diabetic retinopathy stages using CNN. In *Proceedings of the 2017 4th international conference on signal processing and integrated networks (SPIN)*, pp. 550–554 (2017).
- Jain, A., Jalui, A., Jasani, J., Lahoti, Y. & Karani, R. Deep learning for detection and severity classification of diabetic retinopathy. In *Proceedings of the 2019 1st international conference on innovations in information and communication technology (ICIICT)*, pp. 1–6 (2019).
- Sungheetha, A. Design an early detection and classification for diabetic retinopathy by deep feature extraction based convolution neural network. *J. Trends Comput. Sci. Smart Technol.* **3**, 81–94 (2021).
- Hemanth, D. J., Deperioglu, O. & Kose, U. An enhanced diabetic retinopathy detection and classification approach using deep convolutional neural network. *Neural Comput. Appl.* **32**, 707–721 (2020).
- Amin, J., Sharif, M., Rehman, A., Raza, M. & Mufti, M. R. Diabetic retinopathy detection and classification using hybrid feature set. *Microscopy Res. Tech.* **81**, 990–996 (2018).
- Khan, S. H., Abbas, Z. & Rizvi, S. D. Classification of diabetic retinopathy images based on customised CNN architecture. In *Proceedings of the 2019 Amity international conference on artificial intelligence (AICAI)*, pp. 244–248 (2019).
- Nguyen, Q. H., Muthuraman, R., Singh, L., Sen, G., Tran, A. C., Nguyen, B. P. & Chua, M. Diabetic retinopathy detection using deep learning. In *Proceedings of the 4th international conference on machine learning and soft computing*, pp. 103–107 (2020).
- Mishra, S., Hanchate, S. & Saquib, Z. Diabetic retinopathy detection using deep learning. In *2020 International conference on smart technologies in computing, electrical and electronics (ICSTCEE)*, pp. 515–520. (IEEE, 2020).

29. Gangwar, A. K. & Ravi, V. Diabetic retinopathy detection using transfer learning and deep learning. In *Evolution in computational intelligence: frontiers in intelligent computing: Theory and applications (FICTA 2020)*, Vol. 1 (pp. 679–689). (Springer Singapore, 2021).
30. Khan, Z. et al. Diabetic retinopathy detection using VGG-NIN a deep learning architecture. *IEEE Access* **9**, 61408–61416 (2021).
31. Mushtaq, G. & Siddiqui, F. Detection of diabetic retinopathy using deep learning methodology. *IOP Conf. Ser.: Mater. Sci. Eng.* **1070**(1), 012049 (2021).
32. Abdelmaksoud, E., El-Sappagh, S., Barakat, S., Abuhmed, T. & Elmogy, M. Automatic diabetic retinopathy grading system based on detecting multiple retinal lesions. *IEEE Access* **9**, 15939–15960 (2021).
33. Yamashita, R., Nishio, M., Do, R. K. G. & Togashi, K. Convolutional neural networks: An overview and application in radiology. *Insights Imaging* **9**, 611–629 (2018).
34. Mohammadian, S., Karsaz, A. & Roshan, Y. M. Comparative study of fine-tuning of pre-trained convolutional neural networks for diabetic retinopathy screening. In *Proceedings of the 2017 24th national and 2nd international Iranian conference on biomedical engineering (ICBME)*, pp. 1–6 (2017).
35. Yan, B. & Han, G. Effective feature extraction via stacked sparse auto-encoder to improve intrusion detection system. *IEEE Access* **6**, 41238–41248 (2018).
36. Nayak, D. R., Dash, R., Majhi, B., Pachori, R. B. & Zhang, Y. A deep stacked random vector functional link network auto-encoder for diagnosis of brain abnormalities and breast cancer. *Biomedical Signal Process. Control* **58**, 101860 (2020).
37. Nagasubramanian, G., Sankayya, M., Al-Turjman, F. & Tsaramirisis, G. Parkinson data analysis and prediction system using multi-variant stacked auto encoder. *IEEE Access* **8**, 127004–127013 (2020).
38. Kadam, V. J., Jadhav, S. M. & Vijayakumar, K. Breast cancer diagnosis using feature ensemble learning based on stacked sparse auto-encoders and softmax regression. *J. Medical Syst.* **43**, 263 (2019).
39. Khamparia, A. et al. KDSAE: Chronic kidney disease classification with multimedia data learning using deep stacked auto-encoder network. *Multimedia Tools Appl.* **79**, 35425–35440 (2020).
40. Qi, M., Cui, S., Chang, X., Xu, Y., Meng, H., Wang, Y. & Yin, T. Multi-region nonuniform brightness correction algorithm based on L-channel gamma transform. *Security Commun. Netw.* **2022**, (2022).
41. Lu, S., Yang, J., Yang, B., Li, X., Yin, Z., Yin, L. & Zheng, W. Surgical instrument posture estimation and tracking based on LSTM. *ICT Express* **2024**.
42. Jia, Y., Chen, G., & Chi, H. Retinal fundus image super-resolution based on generative adversarial network guided with vascular structure prior. *Sci. Rep.* **14**(1), 22786. <https://doi.org/10.1038/s41598-024-74186-x> (2024).
43. Zhu, Y. et al. Deep learning-based predictive identification of neural stem cell differentiation. *Nat. Commun.* **12**, 2614 (2021).
44. Liu, G., Bao, H. & Han, B. A stacked auto-encoder-based deep neural network for achieving gearbox fault diagnosis. *Math. Problems Eng.* **2018**, 1–10 (2018).
45. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. & Salakhutdinov, R. Dropout: A simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **15**, 1929–1958 (2014).
46. Mandal, S., Bhuiya, S. & Lugez, E. Using attention-based convolutional auto-encoders for catheter path reconstruction in ultrasound images.
47. Liu, N. et al. An eyelid parameters auto-measuring method based on 3D scanning. *Displays* **69**, 102063 (2021).
48. Pepe, M. S. Receiver operating characteristic methodology. *J. Am. Stat. Assoc.* **95**, 308–311 (2000).
49. Hanley, J. A. & McNeil, B. J. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* **143**, 29–36 (1982).
50. @kaggle. Diabetic Retinopathy Dataset. Available at: <https://www.kaggle.com/datasets/sachinkumar413/diabetic-retinopathy-dataset> [Accessed 10 Nov 2024].
51. Adly, M. M., Ghoneim, A. S., & Youssif, A. A. (2019). On the grading of diabetic retinopathies using a binary-tree-based multiclass classifier of cnns. *Int. J. Comput. Sci. Inf. Security (IJCSIS)*, **17**(1).
52. Meshram, A., Dembla, D. & Anooja, A. Development and analysis of deep learning model based on multiclass classification of retinal image for early detection of diabetic retinopathy. *ASEAN Eng. J.* **13**(3), 89–97 (2023).
53. Aswathi, T., Swapna, T. R. & Padmavathi, S. Transfer learning approach for grading of diabetic retinopathy. *J. Phys.: Conf. Ser.* **1767**(1), 012033 (2021).
54. Butt, M. M., Iskandar, D. A., Abdelhamid, S. E., Latif, G. & Alghazo, R. Diabetic retinopathy detection from fundus images of the eye using hybrid deep learning features. *Diagnostics* **12**(7), 1607 (2022).
55. Al-Antary, M. T. & Arafa, Y. Multi-scale attention network for diabetic retinopathy classification. *IEEE Access* **9**, 54190–54200 (2021).
56. Dutta, A., Agarwal, P., Mittal, A. & Khandelwal, S. Detecting grades of diabetic retinopathy by extraction of retinal lesions using digital fundus images. *Res. Biomed. Eng.* **37**, 641–656 (2021).
57. Menaouer, B., Dermane, Z., El Houda Kebir, N. & Matta, N. Diabetic retinopathy classification using hybrid deep learning approach. *SN Comput. Sci.* **3**(5), 357 (2022).
58. Shankar, K. et al. Automated detection and classification of fundus diabetic retinopathy images using synergic deep learning model. *Pattern Recogn. Lett.* **133**, 210–216 (2020).
59. Mohanty, C. et al. Using deep learning architectures for detection and classification of diabetic retinopathy. *Sensors* **23**(12), 5726 (2023).
60. Qureshi, I., Ma, J. & Abbas, Q. Diabetic retinopathy detection and stage classification in eye fundus images using active deep learning. *Multimedia Tools Appl.* **80**(8), 11691–11721 (2021).

Acknowledgments

This work was supported in part by the National Research Foundation of Korea (NRF) grant 2022R1G1A1003531, 2022R1A4A3018824, and Institute of Information and Communications Technology Planning and Evaluation (IITP) grant RS-2020-II201741, RS-2022-00155885, RS-2024-00423071 all funded by the Korea government (MSIT).

Author contributions

Conceptualization, S.A. and F.W.; methodology, S.A.; software, I.K.; validation, S.A., F.W. and S.A.; formal analysis, A.A.; investigation, J.K.; resources, S.A.; data curation, K.U.; writing—original draft preparation, S.A.; writing—review and editing, J.K. and Y.L.; visualization, S.A.; supervision, S.A.; project administration, Y.L. and J.K.; funding acquisition, Y.L. All authors have read and agreed to the published version of the manuscript.”

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-85752-2>.

Correspondence and requests for materials should be addressed to Y.L.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025