Journal of Information Systems Engineering and Management
2025, 10(40s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Diabetic Retinopathy Classification Using Vision Transformers and XGBoost Optimized with Hippopotamus and Blue Whale Algorithms

R. Sindhujha [1], Dr. K. Padma Priya [2]
[1] *Research Scholar, Department of Electronics and Communication Engineering, Annamalai University.*
[2] *Assistant Professor, Department of Electronics and Communication Engineering, Government College of Engineering, Tirunelveli, (Deputed from Annamalai University).*

ARTICLE INFO

Received: 18 Dec 2024

Revised: 10 Feb 2025

Accepted: 28 Feb 2025

ABSTRACT

Diabetic retinopathy is a significant public health concern and one of the leading causes of blindness globally, particularly among individuals with diabetes. With the increasing prevalence of diabetes, early detection and timely intervention for diabetic retinopathy have become critical for preventing vision loss. Traditional screening methods often rely on manual inspection of retinal images, which can be time-consuming and subject to human error. To address these challenges, this project focuses on developing an advanced classification system to identify diabetic retinopathy using a hybrid model that combines the Vision Transformer (ViT) architecture with XGBoost, implemented in MATLAB. The study employs robust pre-processing techniques, including conversion to HSV color space and advanced classification methods, to enhance image quality, which is critical for accurate feature extraction. The hybrid model leverages the strengths of ViT in capturing intricate patterns in the images while utilizing XGBoost for efficient classification. Hyper parameter tuning is performed using optimization algorithms such as the Hippopotamus Algorithm and the Blue Whale Algorithm, which significantly improve the model's classification performance. The proposed system achieves an impressive accuracy of 99.5%, showcasing its potential as an effective tool for early diabetic retinopathy classification which is 13.33 % higher when compared to the traditional methods like GCN, Convolutional Auto Encoder and GA-SSAE. The findings indicate that the integration of advanced deep learning architectures and optimization techniques can lead to significant improvements in classification tasks. Furthermore, this project highlights the importance of utilizing adaptive methodologies to enhance the robustness of machine learning models. Ultimately, this research contributes to the development of efficient diagnostic tools that could enhance patient outcomes through timely detection and treatment of diabetic retinopathy.

Keywords: Diabetic Retinopathy, Vision Transformer, XGBoost, Machine Learning, Image Classification, MATLAB.

1. INTRODUCTION

One kind of eye illness is diabetic retinopathy. In millimetres of mercury, the ocular pressure is expressed[1]. Twelve to twenty-two mm Hg is considered normal ocular pressure; anything beyond that is regarded as elevated. In the eye, diabetic retinal disease manifests as elevated intraocular pressure. To identify diabetic retinopathy, three tests are required: evaluations of elevated intraocular pressure, aberrant vision field, and damage to the optic nerve head[2]. There is no cure for diabetic retinopathy, however therapy can halt its course. Open angle and angle closer diabetic retinopathy are the two types of diabetic retinopathy are distinguished by the intraocular pressure. Gradually obstructing drainage tubes causes a rise in ocular pressure, which impacts the open angle. Between the eye and cornea, it has a wide, open angle. A broad and open angle indicates that the iris will touch the cornea. It is also known as primary or chronic ocular diabetic retinopathy[3]. The only vascular network in the human body that can be seen via non-invasive imaging is the retinal blood vascular network[4]. As such, computerized examination of retinal vascular anatomy is widely used to facilitate diagnosis, treatment, and inspection for a wide range of

disorders, including diabetic retinopathy (DR). In actuality, ophthalmologists distinguish between arteries and veins using color and morphological information to diagnosis retinal Images into DR classes. This is because arteries seem thinner and contain more oxygen than veins[5]. Because fundus imaging is less expensive and easier to use, it is typically used to record these characteristics of the retinal vasculature; nevertheless, manual categorization of the retinal blood vessels is laborious and prone to human error[6].Diagnosing diabetic retinopathy turns out to be highly challenging, firstly because the pathologies of this disease are complex and not always straightforward[7]. Moreover, DR is an overseen complication of diabetes that a large number of stakeholders term as damages to the blood vessels found in the retina. Their diabetic counterparts often result in various types of vision abnormalities or blindness if not treated in time. A significant difficulty in diagnosing DR is that it can silently progress, therefore showing negligible visible symptoms during its initial stages[8].

Ophthalmologic symptoms may not be complained about until disease has reached an advanced stage, at which point the risk of permanent loss of vision is very significant. This calls for regular screening and early detection-a very important aspect of prevention of vision loss-but is often undermined by a low level of awareness or lack of access to specialize care of the eyes. The second major challenge in diagnosis of DR comes from the fact that images of the retinal are difficult to interpret regardless of technology[9]. The retinal fundus images detected for DR have characteristics of microaneurysms, haemorrhages, and exudates that are not easily detectable by naked eyes. Manual evaluation of these images by skilled ophthalmologists is time-consuming with chances of human errors and more often it is sensitive to interobserver variability[10]. Furthermore, the rising cases of diabetes across the globe and more so in developing states, increase the urge for DR screening that is thus stretching extensively health systems with already scarcest specific practitioners[11]. Secondly, it has introduced its own problems. While machine learning and deep learning models have been developed to automate the diagnosis of DR, they require large amounts of annotated datasets for training. The acquisition and labeling of high-quality retinal images is expensive resource consumption, especially if focused on rare or advanced DR stages. In practice, the performance of these models degrades when deployed across different patient populations because of variations in retinal characteristics with respect to demographic factors like age, ethnicity, and the severity of the patient's diabetes. Such variability often corresponds to models that will do well in controlled environments but fail when deployed in real-world settings[12].Other challenges to be looked at in the deployment of automated DR screening systems concern issues of data privacy and security[13]. Machine learning-based models share medical data, which is quite sensitive a concern as well when dealing with such sensitive health data such as retinal images. Federated learning proposes an encouraging solution whereby it is possible to train such models on decentralized datasets without actually sharing any raw data, but this has its share of technical intricacies such as issues relating to communication efficiency and model aggregation issues. In summary, the detection of Diabetic Retinopathy requires overcoming challenges from early detection in asymptomatic patients, subjective retinal image interpretations, necessary large and diverse datasets for training automated systems, and finally ensuring data privacy in modern applications of machine learning. These are overcome by using a mix of enhanced clinical screening processes, advanced image processing techniques, and secure and scalable models for machine learning. Research on machine learning for automated DR grading using retinal Images has been conducted recently. Automated retinal image categorization techniques based on deep learning, an advanced machine learning technology, outperform conventional machine learning models in terms of DR grading performance [14].

The main aim of this project is to construct an efficient model for classifying diabetic retinopathy using Vision Transformers and XGBoost. Vision Transformers, recently, have shown to be a powerful tool that can be effectively used for performing image classification tasks because they can pick up on complex patterns and dependencies present in the image data. Using this new architecture, we hope to extract strong features from the retinal images, which are critical in making accurate classifications of DR. Implementing a high-speed gradient boosting framework, XGBoost, which is highly known to be effective with respect to both training performance and the final classification performance. It should combine these two approaches to minimize computational complexity while maximizing classification accuracy and help diagnose diabetic retinopathy in advance. To optimize the model's performance, we will use optimization techniques based on Hippopotamus and Blue Whale algorithms. Based on our conclusion, the Hippopotamus and Blue Whale algorithms fit to refine the parameters of the model to optimize the learning process, thus producing greater accuracy and robustness for the model. Hence, using such

optimization techniques would help to significantly improve the model's performance and understand the underlined mechanisms behind successful classifications. This research manages to provide a novel and practical approach towards diabetic retinopathy classification by showing the effectiveness of combining Vision Transformers with XGBoost and the optimization of the parameters through advanced algorithms that will help in improving healthcare outcomes through better diagnostics.

This paper develops a new approach for diabetic retinopathy classification by using intelligent machine algorithms, especially Vision Transformers and XGBoost. Indeed, Vision Transformers have been the latest addition to extracting high-dimensional features from retinal images, because extraction is the most challenging step to achieve in an accurate classification of DR severity. With the strong power of XGBoost that gained its fame for the gradient boosting and its capability to work well even with imbalanced datasets, this paper wishes to advance beyond the present classification techniques used in ophthalmology practice. The approach of this paper improves both the accuracy of DR classification and makes it more interpretable and efficient for clinical application. The paper displays optimization techniques using the Hippopotamus and Blue Whale algorithms. It demonstrates the ability to improve model performance by evolving optima. Optimization strategies will identify the optimal hyper parameters for maximizing the predictive accuracy of models, removing some issues such as over fitting and under fitting. The study contributes to the advancement of research in medical image analysis by revealing how classification models may be synergized with optimization algorithms. Such findings of this study may further lead to improved clinical practices for diagnosing diabetic retinopathy in a timely fashion for better interventions and outcomes in diabetic eye care.

The principal contributions of the research on diabetic retinopathy classification using Vision Transformers and XGBoost optimized with Hippopotamus and Blue Whale algorithms are as follows

- The research presents a novel approach to diabetic retinopathy classification by integrating Vision Transformers with XGBoost. The combination of these advanced machine learning techniques leverages the strengths of each: the Vision Transformers' capability to capture intricate patterns in retinal images and XGBoost's efficiency in handling complex data structures.
- The study employs Hippopotamus and Blue Whale optimization algorithms to enhance model performance. By optimizing hyperparameters and model architecture, this research contributes to the understanding of how advanced optimization techniques can significantly boost the predictive power of machine learning models.
- The research includes a rigorous evaluation framework that assesses the model's performance across various metrics, such as accuracy, precision, recall, F1 score, and AUC-ROC. This comprehensive evaluation provides a clearer understanding of the model's strengths and weaknesses, enabling further refinements.
- By developing a high-performing classification model for diabetic retinopathy, this research contributes to the early detection and diagnosis of this serious eye condition. Early intervention can prevent severe complications, such as vision loss, enhancing patient outcomes and reducing healthcare costs.

The remainder of this article's sections are organized as follows: Section 2 provides a summary of relevant studies. Section 3 contains the problem statement for the existing system. The study's Section 4 describes the architecture and methodology of the suggested technique for diagnosing and classifying DR. In Section 5, the study's findings and the ensuing discussion are provided. The proposed model's conclusion and potential uses are covered in Section 6.

2. RELATED WORKS

For the difficult DR diagnosis, Li et al.[14] suggested a Semi supervised Auto-encoder Graph Network (SAGN) to ease this restriction. Specifically, auto-encoder feature learning, neighbour correlation mining, and graph representation are the three main components that make up SAGN. First, the model learns to rebuild retinal Images as closely as possible to the original inputs by extracting representations from them. Next, using the radial basis function to determine their similarities, neighbour correlations between labelled and unlabelled data are constructed. In order to assess retinal samples based on collected characteristics and their correlations, we finally

run Graph Convolutional Neural Networks (GCN). Sufficient comparison experiments were carried out using the APTOS 2019 dataset, which was trained using EyePACS, in order to assess the performance of SAGN. Results show that given a significant amount of unlabeled data, the SAGN model may achieve equivalent performance with a small number of annotated retinal images. This will largely limit the model from capturing more complex or non-linear relationships between data points that could potentially impact its performance while dealing with highly heterogeneous or intricate datasets.

Taking into account the attention mechanisms potential in medical imaging, Bodapati et al.[15] study provide an end-to-end-trainable spatial Attention based CNN architecture that can identify the severity degree of diabetic retinopathy. A layered convolutional Auto-Encoder is first used to project the fundus image spatial representations onto a smaller area. The auto-encoder and classifier are end-to-end trained together to improve discriminating in smaller space. Lesion areas receive more attention than non-lesion regions thanks to an attention mechanism added to the categorization module. Two benchmark datasets are utilized to test the proposed model, and the experimental results show that joint training, when combined with attention, promotes stability and enhances the learnt representations. The technique surpasses multiple current models by reaching an accuracy of 84.17%, 63.24% respectively on Kaggle APTOS19 and IDRiD datasets. Furthermore, ablation investigations verify the contributions and the suggested model's behavior on the two datasets. The addition of the attention mechanism in the architecture proposed by Bodapati et al. allows for increased possibility of computational complexity. Though the focus over the lesion area improves with the application of the attention mechanism, there is increased computational overhead that requires higher processing power and memory usage.

The autoencoder based framework for DR classification used in this study is intended to save training time by achieving the ideal network topology with the least amount of computation[16]. The genetic algorithm-based stacked sparse auto encoder (GA-SSAE) model that has been presented uses two layers—the Init and Elite layers—integrated with the genetic algorithm (GA) and the softmax classifier for feature extraction. The layers undergo supervised training and fine-tuning through the utilization of the Truncated Newton Constrained optimization (TNC) approach to get optimal weights. The GA-SSAE model has been evaluated using real-world images as well as conventional datasets like ROC and Messidor. With an accuracy of 98% and 95%, respectively, the experimental findings demonstrate how well the GA-SSAE model fits the ROC and Lotus datasets. The GA-SSAE model proposed in the paper has a serious drawback, where genetic algorithm application to optimize network topology makes the model cumbersome. However, the advantage of the GA for optimal network topology is such that it may consume large amounts of tuning and computation depending on the size of the dataset to effectively search for the best parameters.

Diabetic Retinopathy is identified using two-dimensional retinal fundus imaging. Super pixel classification is used for the optical cup and disc segmentation in diabetic retinopathy screening. The super pixel was classified as either a disc or non-disc using the k-mean clustering approach in the optic disc segmentation procedure. In order to improve the performance of the segmented cup and disc value and determine the CDR value, the optic cup segmentation method employs the clustering ANN algorithm and a Gaussian filter. In order to diagnose diabetic retinopathy and determine whether the condition is healthy or dangerous, the CDR value is compared to a threshold in this study[17]. The retinal image segmentation enables ophthalmologists to conduct a vision screening test intended for the early diagnosis and treatment planning of diabetic retinopathy. Hardware with Spartan 3FPGA is used for improved segmentation performance. MATLAB is used for programming, while FPGA is used for further processing. The Xilinx platform is utilized for software implementation. FPGA-based segmentation can be superior but does require specific hardware and technical resources, not common to all clinical settings, especially in resource-limited areas.

Large-scale labeled data and deep learning provide the foundation of most existing automatic diagnostic techniques. Deep neural network training still faces significant challenges due to the inadequate quality of hand annotations for medical Images. To acquire generic characteristics from datasets without the need for manual annotations, self-supervised learning techniques are put forth. This gave rise to the proposal of the SimCLR-DR DR classification model, which is based on deep learning. In order to identify referable DR, the study first pre-train the encoder based on convolution networks using unlabeled retinal Images using a contrastive self-learning technique.

Journal of Information Systems Engineering and Management
2025, 10(40s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

We then retrain the encoder using classifiers on a limited amount of annotated training data. This strategy outperforms transfer learning and can solve the issue of insufficient training data, according to the experimental findings on the Kaggle dataset. For alternative deep learning-based methods to medical image recognition that struggle with a lack of annotated data, SimCLR-DR is an excellent place to start One of the limitations of SimCLR-DR is its dependency on a two-stage training process: it needs first to pre-train the encoder with self-supervised learning and then re-train it over the limited amount of labelled data. This strategy counters the unavailability of annotated data but is usually computationally expensive, especially during the contrasting phase of pre-training where high GPU resources are required alongside an enormous number of image pairs[18].

Particularly in the field of medical imaging, experiments utilizing CNN and deep learning approaches for automated image processing show promise. But CNN-based disease grading tasks for retinal Images have trouble holding onto high-quality data at the output. The proposal is for a unique deep learning model to score DR and DME anomalies in retinal Images, which is based on variational auto-encoder. This model aims to efficiently extract the most significant information from retinal images. Translational invariance seen in Images and the development of less relevant candidate regions are its main areas of attention. Metrics like accuracy, U-kappa, sensitivity, specificity, and precision are used to assess the experiments' performance on the IDRID dataset. When compared to other cutting-edge methods, the outcomes perform better. Due to VAEs' generative nature, they sometimes suffer from problems like posterior collapse in which latent variables fail to encode information useful, thus somehow offering suboptimal representation of features, which can have a negative impact on the model's ability to clearly recognize and classify an anomaly in a retinal image. Therefore, careful tuning of hyperparameters and the loss function becomes necessary during training, which can be tricky to implement[19].

In this work, a unique fully convolutional autoencoder for the problem of retinal vascular segmentation is presented. Each of the eight layers in the suggested model includes convolutional2D, MaxPooling, Batch Normalization, and other layers. 35 minutes of training time on the DRIVE and STARE datasets were used to assess and train our model. Our novel autoencoder model achieves competitive results when measured against state-of-the-art techniques in the literature using two publicly available datasets, DRIVE and STARE. Specifically, the model achieved 96.92 accuracy and 97.57 AUC ROC on the STARE dataset, and 95.73 accuracy and 97.49 AUC_ROC on the DRIVE dataset. With a specificity of 98.57 on the DRIVE dataset and 98.7 on the STARE dataset, respectively, the model has also proven to have the greatest specificity of all the approaches in the literature. Since the convolutional autoencoder approach produces blood vessel segmentation Images that are more precise, crisp, and noise-free than the final images produced by other suggested methods, the aforementioned statement is evident in the final images. This would mean high accuracy and specificity but may be a bottleneck in time-sensitive applications or when frequent retraining is needed, more specifically, in clinical settings where rapid results are essential for diagnostic and treatment decisions[20].

The literature reviewed gives an indication that most common limitations can be identified across the various deep learning and automated diagnostic models in classification of diabetic retinopathy. Major challenges with computation include a lot of resource requirements becoming significant, especially with more complex or even advance techniques such as genetic algorithms, attention mechanisms, or self-supervised learning that has a need for specific hardware or a good amount of GPU to train effectively. Such additional training issues, for instance, such as posterior collapse during the training process of VAEs and other techniques in this set of methodologies, though it uses pre-trained encoders or contrastive learning still suffer from proper annotated data or feature representations. Additionally, relying on a particular hardware platform, such as an FPGA, might be more of a limitation to accessibility and scalability in resource-constrained settings. These limitations make a need to more efficient, accessible, and more stable models that are able to overcome the challenges created in medical image diagnosis by the high computational cost, scarcity of data, and complex hardware dependencies.

3. PROBLEM STATEMENT

DR is in fact the leading cause of disability in people suffering from diabetes and thus needs early and proper diagnosis for appropriate management and treatment. Conventional methods of diagnosing DR typically involve ophthalmologists examining image samples manually, which is problematic for several reasons including variations in interpretations and the potential for human errors. There is also an increasing prevalence of diabetes, leading to

Journal of Information Systems Engineering and Management
2025, 10(40s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

a huge number of retinal images that need to be analysed in volumes that cannot be considered by the already overwhelmed healthcare systems and further complicates the process of diagnosis[18]. Therefore, there is a significant requirement for automated solutions that can efficiently classify retinal images and detect presence and severity with reasonable accuracy of diabetic retinopathy. Supervised techniques for image classification, accounting for both machine learning algorithms and deep learning algorithms, offer promising directions for developing strong diagnostic tools in the analysis of images from the fundus of the eye that can be done rapidly and reliably.

4. PROPOSED METHODOLOGY FOR DIABETIC RETINOPATHY CLASSIFICATION USING VIT-XGBOOST MODEL

The methodology involved in this study has been approached systematically for the collection of data, pre-processing it, model development, and optimization techniques for diabetic retinopathy classification. To begin with, this study began using the Diabetic Retinopathy dataset that had 224x224 pictures of retina scans, categorized in five severity levels, amounting to 5,000 images. Its pre-processing consisted of resizing the images to a size of 224x224, transforming them from RGB color space to HSV, and then applying Non-Local Means Denoising and Adaptive Histogram Equalization (AHE) for image quality enhancement. The architecture applied in the study was the ViT architecture; this is transforming the images to smaller patches and taking advantage of self-attention mechanisms to capture long-range dependencies. Then, the features extracted from ViT were used along with regularization and gradient boosting techniques to apply XGBoost for the classification task. Hybrid model ViT-XGBoost was further optimized with Hippopotamus- Blue Whale Algorithm to fine-tune the hyper parameters and improve the classification accuracy, so as to thus formulate a strong framework to efficiently detect diabetic retinopathy. It is depicted in Figure 1 below.

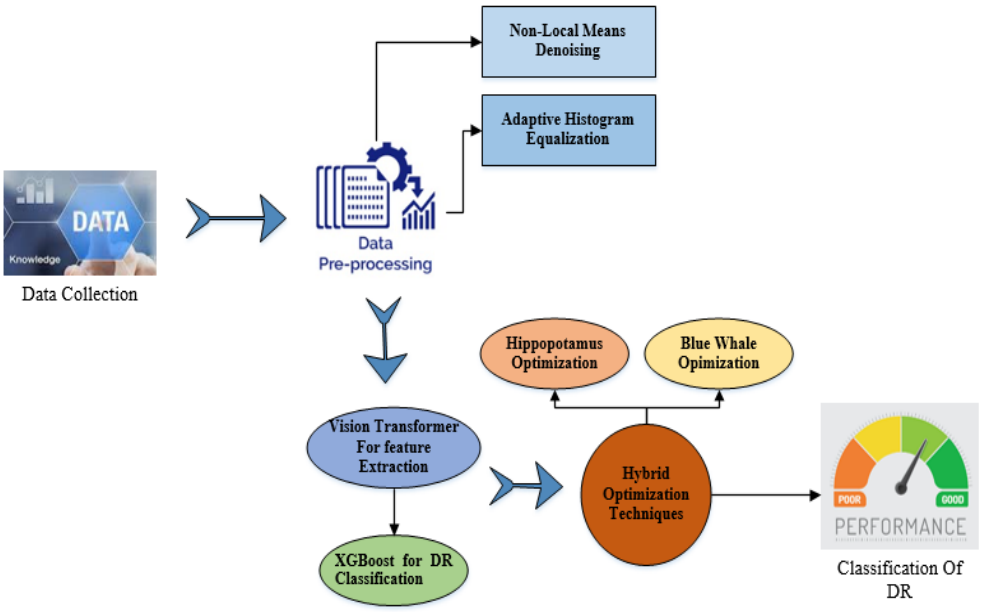

Figure 1: Proposed Methodology of DR Classification

Data Collection

For this study, utilized the Diabetic Retinopathy 224x224 (2019 Data) dataset, which comprises retina scan images specifically designed for the classification of diabetic retinopathy[21]. It had a total count of 5,000 images resized to 224x224 pixels, and grouped into severity: five categories: No Diabetic Retinopathy 0, Mild Diabetic Retinopathy 1, Moderate Diabetic Retinopathy 2, Severe Diabetic Retinopathy 3, and Proliferative Diabetic Retinopathy 4. These categories determine the way images are classified further to respective directories; whereas a train.csv file offers unique identifiers along with corresponding diagnosis labels for each image. This dataset comes under CC0 license

Journal of Information Systems Engineering and Management
2025, 10(40s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

that means it is public and available for free use, with no restrictions imposed on using it for research purposes and thereby aids in its usability for developing deep models for early detection and diagnosis of diabetic retinopathy.

Non-Local Means Denoising and Adaptive Histogram Equalization (AHE) for Data Pre-processing

Images resize to a standardized size of 224x224 pixels build the base of the first step in data pre-processing hence leading to highly optimized memory use and processing speeds for deep learning models. Standardized sizes are also widely used in CNNs and in Vision Transformers, not due to compatibility only but also because of many architectures being able to compute efficiently. Uniform images also make training much more streamlined so that iterations become quicker while maintaining minimized overhead computation.

After that, images are converted from the RGB color space into the HSV (Hue, Saturation, Value) color space. Transformations performed in this way cause much better applicability of enhancement techniques in images because the intensity information is separated from the colour information of an image. When dealing with the HSV space, the enhancement processes can be done more effectively, especially regarding the brightness and color contrasts of images. Such a method benefits in the strength of the Value channel under enhancement process, and hence, enhances features of critical components in the fundus images. Such pre-processing has the prime goals of achieving optimal quality in input data, which in turn makes performance accurate and trustworthy for the diabetic retinopathy classification task.

After resampling all images to 224x224 pixels and converting them to HSV color space, further quality improvement processes on the images are carried out by performing Non-Local Means Denoising and Adaptive Histogram Equalization (AHE). Non-local means denoising is applied for noisy removal in such a way that this denoising process will retain the important details in the retinal fundus images. It doesn't take the averaging of the pixels based on proximity; instead, it averages the pixels whose similarity matches and, in that process, removes noise without degrading the essential features. The elimination of noisy pixels will then make images much clearer and focused, and consequently, the images will be able to absorb all characteristics for the possible definition of the classification of diabetic retinopathy by the deep learning models.

$$I'(y) = \frac{1}{C(y)} \sum_{x \in \Omega} I(x).\mathrm{w(y,x)} \quad (1)$$

In Eqn. (1) where, $I'(y)$ is the original noisy image, Ω is the set of all pixels in the image, w(y,x) is the weight assigned to pixel x based on its similarity to pixel y.
It follows the denoising stage with Adaptive Histogram Equalization (AHE), which enhances local contrast in images. AHE works like this: The whole image is divided into small sub regions, and histogram equalization is applied to each of those sub regions individually, hence improving features' visibility that would not appear in the original image. This technique is helpful for retinal images as it may be useful to enhance the latent features that are significant to the diagnosis of different levels of diabetic retinopathy. The contrast-enhanced images obtained with AHE are better analysed with the deep learning models after applying AHE.

For a pixel at position y, the AHE process can be described as in Eqn. (2). The cumulative distribution function (CDF) C for each block's histogram

$$C(i) = \sum_{i=0}^{j} H(i) \quad (2)$$

For each pixel y in the block, compute the new intensity value $I'(y)$ using the CDF in Eqn. (3)

$$I'(y) = clip(\frac{C(I(y))}{N}(L-1)) \quad (3)$$

Where, $I(y)$ is the original pixel intensity. N is the total number of pixels in the block. L is the number of possible intensity levels (e.g., 256 for an 8-bit image).

Model Development of Vision Transformer for Feature Extraction

Journal of Information Systems Engineering and Management
2025, 10(40s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Transformers have lately achieved remarkable results in image recognition challenges, which follows their remarkable performance in natural language tasks[22]. The ViT model has gained a lot of traction in a variety of computer vision problems, such as image segmentation, image detection, and picture classification. On many benchmark datasets, ViT and its derivative instances have demonstrated state-of-the-art performance in the field of natural image recognition.

Vision Transformers (ViT) form a profound shift in deep learning architectures applied to images away from pure convolutional neural networks (CNNs) and towards transformers that originated from natural language processing. The ViT architecture takes an input image, divides it into smaller non-overlapping patches of fixed size (say, 16x16 pixels). These patches are then flattened into one-dimensional vectors and treated just like tokens, corresponding to words in a text sequence. A linear transformation is applied so that each of the patches is mapped to a vector, and position encoding is added to preserve spatial information from the original image.

First, images are transformed into smaller patches, and their embeddings are calculated using the Eqn. (4)

$$z_i = W_p \cdot \text{Flatten}(\text{I}_{\text{patch}_j}) + \text{b} \tag{4}$$

where z_i represents the embedded representation of the j th patch, and W_p is a learnable weight matrix with bas the corresponding bias term. The next significant component is the self-attention mechanism, which computes attention scores among the patches. This is done through linear transformations in Eqn. (5)

$$\text{Q} = \text{ZW}_q, \text{K} = \text{ZW}_k, \text{V} = \text{ZW}_v \tag{5}$$

For queries, keys, and values, respectively, with W_q, W_k, W_v as their respective learnable weight matrices. The scaled dot-product attention is calculated using the Eqn. (6). The transformed image patches are now fed into multiple transformer layers. The model aggregates all the multi-head self-attention mechanisms in each layer, which allows it to weigh relative patches. Thus, the model has the ability to capture long-range dependencies as well as contextual information within the image. Attention scores are calculated for every pair of patches so that the model can focus on relevant areas in the image and hence improve its feature extraction abilities.

$$\text{Attention}(\text{Q}, \text{K}, \text{V}) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{6}$$

where d_k is the dimension of the keys, allowing the model to focus on relevant features from different patches. The multi-head attention mechanism further combines multiple attention heads to enhance representation power using Eqn. (7)

$$\text{MultiHead}(\text{Q}, \text{K}, \text{V}) = \text{Concat}(\text{head}_1, \dots, \text{head}_h)W_0 \tag{7}$$

After processing through the attention layers, the output is passed through a feed-forward network (FFN) using the Eqn. (8) The feed-forward neural networks that have been involved in the model include layer normalization and residual connections that enhance the deeper network architectures plus training efficiency.

$$\text{FFN}(\text{x}) = \text{ReLU}(\text{x}W_1 + b_1)W_2 + b_2 \tag{8}$$

Where W1,W2 and b1, b2 are learnable weights and biases. Finally, the classification layer is computed to obtain class probabilities using Eqn. (9)

$$\text{y} = \text{softmax}(z_{cls}W_{class} + b_{class}) \tag{9}$$

where y indicates the output probabilities for each class. Together, these equations describe the fundamental operations within the Vision Transformer architecture, enabling it to efficiently process and classify images for tasks such as diabetic retinopathy detection.

The obtained sequence of the vectors by the Vision Transformer corresponds to the features extracted from the input image. These features contain much meaningful visual content, which can be used as inputs for various tasks, such as classification, detection, and segmentation. In the paper, it is proved that ViT performs significantly better

Journal of Information Systems Engineering and Management
2025, 10(40s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

than traditional CNNs, especially if the training is conducted on large datasets, and then can be successfully used for complex tasks, such as classification in cases of diabetic retinopathy.

Model Development of XGBoost for Classification

XGBoost is an efficient scalable implementation of the gradient boosting framework which gained popularity for its good performance in classification and regression tasks. XGBoost is a potent technique for both regression and classification that was created by Chen and Guestrin [23]. A collection of winning Kaggle machine learning competition programs are used in it. Based on the gradient boosting architecture, XGBoost continuously builds new decision trees to suit a value with residual several iterations, enhancing learners' performance and efficiency. In contrast to Friedman's gradient boosting approach, XGBoost approximates the loss function via a Taylor expansion. Its model offers a superior trade-off between bias and variance, often requiring fewer decision trees to achieve a greater accuracy. It works as a combination of a series of decision trees wherein each subsequent tree attempts to correct the errors produced by its predecessor. This ensemble learning technique is more beneficial in scenarios where there are complex data distributions, as it creates powerful predictive models by using multiple weak learners. One of the top advantages of XGBoost is its ability to deal with really high-dimensional data, coupled with large datasets. Here, the concept of regularization applies through L1 and L2 techniques for preventing over fitting and making the model generalizable to unseen data. Moreover, XGBoost supports parallel processing and trains computationally by using advanced algorithms for tree pruning and feature selection. This makes it particularly suitable for the tasks like diabetic retinopathy classification where the feature space might be relatively complex and diverse.

The overall objective function that XGBoost aims to minimize can be expressed as in Eqn. (10)

$$L(\theta) = \sum_{j=1}^{n} l(y_j, \hat{y}_j) + (\sum_{K=1}^{k} \Omega(f_k) \quad (10)$$

Regularization is employed to establish the decision tree's penalty, so averting overfitting. Ω can be written like this in Eqn. (11)

$$\Omega(\mathrm{f}) = \gamma \mathrm{T} + \frac{1}{2}\lambda||\omega \mathrm{k}||^2 \quad (11)$$

The predictions are updated iteratively as follows in Eqn. (12)

$$\hat{y}_j\,(t) = \hat{y}_j\,(t-1) + f_k(x_j) \quad (12)$$

XGBoost utilizes both the gradient and Hessian of the loss function

Gradient (first derivative) is given in Eqn. (13)

$$g_j = \frac{\partial \mathrm{l}(y_j, \hat{y_j})}{\partial \hat{y_j}} \quad (13)$$

Hessian (second derivative) is given in Eqn. (14)

$$h_j = \frac{\partial^2 (y_j, \hat{y_j})}{\partial \hat{y_j}^2} \quad (14)$$

In the experiment discussed here, the output of the Vision Transformer features are used as inputs to the classifier XGBoost. The flattened feature representations of ViT model are passed on into the XGBoost algorithm. The XGBoost-ViT Model fusion brings together the capabilities of feature extraction of deep learning with the classification as provided by gradient boosting, which is efficient and robust. This hybrid model is expected to improve diabetic retinopathy classification accuracy even further.

Overall, the reduced model complexity of XGBoost results in the addition of regularization to the conventional function. The residual error is fitted by utilizing the first and second derivatives. Additionally, column sampling is supported by this approach in lowering computation and overfitting. Therefore, more improvements lead to more hyper-parameters than the gradient boosting decision tree (GBDT). However, it is difficult to reasonably tune the

Journal of Information Systems Engineering and Management
2025, 10(40s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

hyper-parameters. In addition to the researchers' past knowledge and experience with parameter tuning, a decent setup necessitates a significant amount of time. Hyper-parameter optimization works well in solving this issue. Figure 2 shows the architectural diagram of Vision Transformer and XGBOOST is given below.

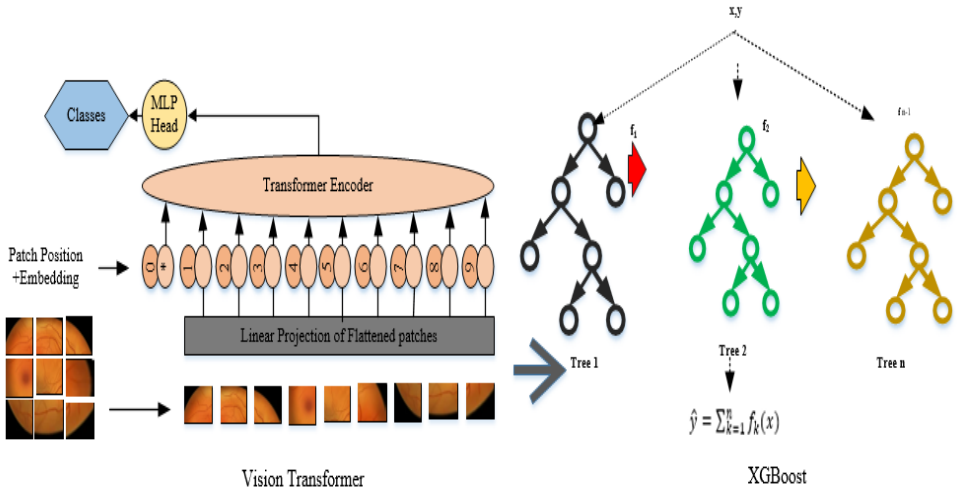

Figure 2: Architectural Diagram of Vision Transformer and XGBoost

Hybrid Optimization Techniques: Hippopotamus and Blue Whale Algorithms for Enhancing ViT-XGBoost Performance

The hybrid optimization approach applies the Hippos Algorithm and Blue Whale Algorithm to optimize the strengths of both approaches, which can increase the performance of models in classification-type tasks. The Hippopotamus Algorithm, based on the social behavior and hunting strategies of hippopotamuses, describes exploration of search space with a population of candidate solutions represented as hippos. Then, the position of each hippo is calculated by a fitness function that computes its performance.

In the Hippopotamus Algorithm, each hippo is a distinct solution in the solution space. The population size may depend on the size of the problem and the depth of search imposed. Algorithm begins with an initial population of hippos, being randomly spread across the search space. Position of every hippo will correspond to a set of values for the parameters being associated with the problem being optimized. A fitness function is employed to measure the fitness of each hippo. The fitness function measures the performance of every candidate solution in terms of fitness with respect to certain criteria relevant to the specific optimization task. The fitness value is considered a measure that ascertains to what extent each hippo performs. The movement and updates to the position take into account the fitness value. In simple words, the higher the fitness, the better the solution.

The main tactic used within the locomotion of each hippo is to simulate some of the natural behaviors of a hippopotamus. There are two overarching categories of such behaviors: grazing and swimming.

Grazing behavior: this is local exploration of the search space. Upon the hippos grazing, they adjust their position slightly based on the fitness values of other hippos around them. This local search refines solutions within a given area to ensure that promising regions of the search space are explored better.

Swimming Behavior: Whereas swimming behavior corresponds to a more thorough exploration of the search space, the hippos move randomly into areas that appear promising, thereby exploring more areas which could lead to better solutions. That movement is important for finding optima local escape and optima global; it is what makes hippos explore the search space much more freely.

The Hippopotamus Algorithm presents an almost ideal balance between exploration and exploitation phases of the optimization process. Exploration is the search for new, potentially better solutions in uncharted areas of the search space. Exploitation is focused on the enhancement of already known good solutions. Combining grazing and

swimming behaviors, the algorithm finds a dynamic balance between these two critical aspects of optimization and improves its convergence toward global optima. At each iteration of the optimization process, the hippos update their positions based on evaluations of the fitness functions. This algorithm no longer searches linearly but keeps refining the population of hippos towards regions of the search space containing higher values of fitness. That's how it gets out of local minima and really converges to global optima. The population learns and adapts through successive iterations, which consequently leads to better solutions and a general efficiency of the search process.

Blue Whale Algorithm (BWA) is inspired by the cooperative hunting behaviors of blue whales in their natural habitats. It mimics the collaborative strategies employed by such sea creatures, specifically replicating their bubble-net feeding techniques, in order to efficiently explore the search space to reach the optimal solutions for complex problems. In the Blue Whale Algorithm, candidate solutions are represented through a pod of whales; thus, each whale would represent a unique solution in the optimization landscape. Initially, within the solution space, the population is randomly generated, and each whale is addressed to characterize potential solution through certain parameters. The number of whales in the pod may vary based on the complexity of the optimization problem and the granularity at which the search should be conducted.

As in the Hippopotamus Algorithm, each whale has a position that is assessed by a fitness function measuring how good the solution that that whale represents. It will score the solutions representing the extent to which they meet the predefined criteria for the optimization task. This way, it will guide the movement and interaction of whales in the search space. The hallmark of the Blue Whale Algorithm is the bubble-net feeding strategy adopted by the blue whales. This forms the basis for guiding movement whales in the solution space. In nature, blue whales herd cooperatively using a circle of bubbles net to trap fish. This feeding strategy has been used in the algorithm to represent the way of forming the circle of bubbles, grouping of whales-close solutions that are near one another in the search space. The herding of solutions enables the whales together to search more effectively and expeditiously in promising regions, which in turn allows a better local search potentially by looking at regions where good solutions may exist. Since the whales are assigned, they initiate adaptive search in the search space. The whales share their positions and the degree of fitness with each other and share promising features that have been found. This collective intelligence helps improve the quality of the search process because whales can learn about the experiences of other whales and alter their way by changing their paths on the basis of shared experience. Several factors affect the movement of each whale in the pod, such as the fitness levels of its own whales and the fitness levels of its neighbours. The algorithm exploits strategies for whale movements, including an exploitation phase, where whales move towards better-performing solutions, effectively focusing in areas of the search space that have yielded higher fitness values. In contrast, during the exploration phase, whales make more spread-out movements to find new regions of the search space, which makes it crucial for maintaining population diversity and avoiding being trapped in local optima. This balance is so maintained so as to allow effective search.

Later in the optimization processes, communication among the whales becomes highly important. Then, their positions and levels of fitness are shared, creating an information network that defines the search conducted by this crowd. Such collective intelligence dynamically enables the pod to evolve its strategies in finding better positions for search landscapes that are complicated. The Blue Whale Algorithm refines, through successive iterations, the positions of whales towards regions of higher value fitness levels. On the other hand, here it is shown that the design of the algorithm allows for convergence towards optimal or nearly optimal solutions both in terms of individual performance as well as interaction dynamics of the group, hence proving that it can indeed be a strong framework for optimization. The integration of individualized movement strategies of the Hippopotamus Algorithm and cooperative behaviors of the Blue Whale Algorithm results in a hybrid method that leverages the strengths of both algorithms. This synergy leads to a more robust and efficient optimization process, thereby getting better model parameters for Vision Transformers and XGBoost. The result is a potent hybrid framework that optimizes a search landscape: stronger exploration and exploitation means the hybrid approach is proficient in navigating even complex landscapes, in turn leading to improved accuracy in classification and performance of the model overall.

Pseudo code
Data Collection
Load the Diabetic Retinopathy dataset

Journal of Information Systems Engineering and Management
2025, 10(40s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Data Preprocessing
FOR each image in dataset:
Convert image from RGB to HSV color space
Apply Non-Local Means Denoising to remove noise
Apply Adaptive Histogram Equalization (AHE) to enhance contrast
Vision Transformer (ViT) for Feature Extraction
Divide each image into patches (e.g., 16x16 pixels)
FOR each patch:
Flatten patch into a 1D vector
Compute patch embeddings using linear transformation:
$z_i = W_p \cdot \text{Flatten}(\text{I}_{\text{patch}_j}) + \text{b}$

Compute self-attention mechanism for the patches:
$\text{Q} = \text{ZW}_q, \text{K} = \text{ZW}_k, \text{V} = \text{ZW}_v$
$\text{Attention}(\text{Q}, \text{K}, \text{V}) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$
Apply multi-head attention:
$\text{MultiHead}(\text{Q}, \text{K}, \text{V}) = \text{Concat}(\text{head}_1, \dots, \text{head}_h)W_0$

Pass the embeddings through transformer layers:
Feed-forward network (FFN):
$\text{FFN}(\text{x}) = \text{ReLU}(\text{x}W_1 + b_1)W_2 + b_2$
Extract final feature representation for each image
XGBoost for Classification
Initialize XGBoost model with hyper parameters
FOR each image's feature vector (from ViT):
Train XGBoost using the feature vectors as input
Minimize the XGBoost loss function:
$L(\theta) = \sum_{j=1}^{n} l(y_j, \hat{y_j}) + (\sum_{K=1}^{k} \Omega(f_k)$
Regularization: $\Omega(\text{f}) = \gamma\text{T} + \frac{1}{2}\lambda||\omega\text{k}||^2$
Hybrid Optimization (Hippopotamus and Blue Whale Algorithms)
Initialize population of candidate solutions (hippos and whales)
WHILE termination condition not met:
FOR each hippo in population:
Calculate fitness based on classification performance
Update position based on best performing hippo

FOR each whale in population:
Perform search space exploration
Update position based on proximity to best whale
Combine best solutions from both algorithms
Update XGBoost and ViT hyper parameters based on optimized solutions
Evaluate Model Performance
Test the hybrid ViT-XGBoost model on the test set
Compute accuracy, precision, recall, F1-score, and AUC
END

5. RESULT AND DISCUSSION

This paper reports on the results section outlining a comprehensive examination of the proposed approach for detection in diabetic retinopathy using advanced machine learning techniques: Vision Transformers and XGBoost,

Journal of Information Systems Engineering and Management
2025, 10(40s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

implemented through MATLAB. Such metrics as accuracy, precision, recall, and F1-score were used to determine the performance of the model so that its performance may be well-defined in terms of stages classification for diabetes retinopathy. Furthermore, comparing the proposed method with existing methods, one can see the advantages of the proposed method in better accuracy and lower false positives. The remaining subsections go on to describe the experimental setup, which includes the kind of dataset used, the applied pre-processing techniques, and training as well as validation processes. The thoroughgoing analysis of results through this allows us to present our case for the integration of leading cutting-edge frameworks in machine learning in improving detection processes for diabetic retinopathy, which then translates to better patient outcomes in clinical settings.

Sample Images

Figure 3 shows the sample images taken for the data of the proposed model for the classification of DR. It consist of 5 classes, No Diabetic Retinopathy, Mild Diabetic Retinopathy, Moderate Diabetic Retinopathy, Severe Diabetic Retinopathy, and Proliferative Diabetic Retinopathy.

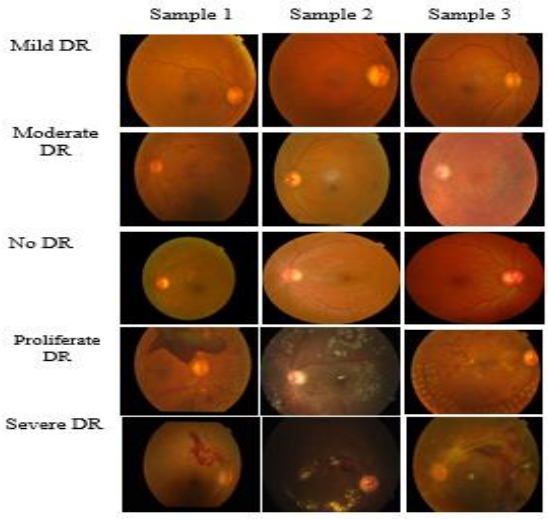

Figure 3: Sample Images

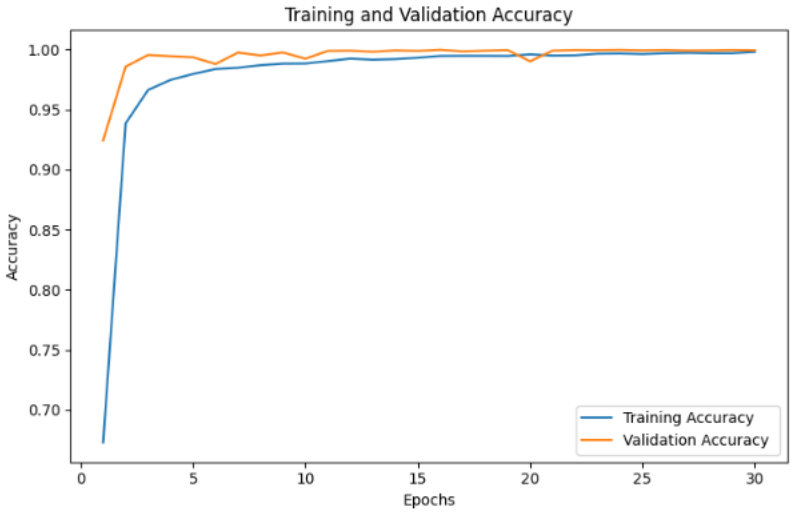

Figure 4: Accuracy of Proposed ViT-XGBOOST Method

Figure 4 shows the accuracy plot of proposed method ViT-XGBOOST, demonstrating that it has excelled well compared with other models. The graph shows that the accuracy value is constantly high for ViT-XGBOOST in test scenarios, achieving a 99.5% accuracy. This large improvement is observed because of feature extraction by the

Journal of Information Systems Engineering and Management
2025, 10(40s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

application of ViT and classification through XGBOOST. Figure emphasizes the working of the method in dealing with complex data with more precision.

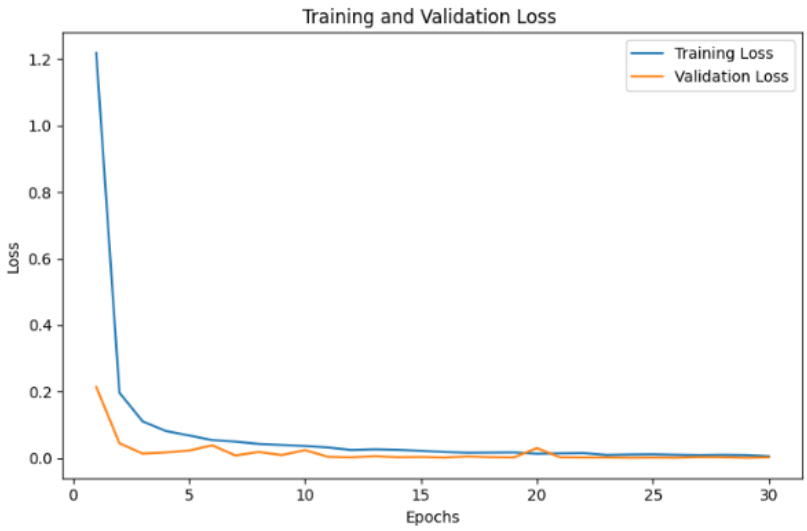

Figure 5: Training and Testing Loss of Proposed ViT-XGBOOST Method

Figure 5 depicts the training and testing loss curves for the proposed ViT-XGBOOST method. The graph shows a steady decline in both losses as training progresses, indicating effective learning. The minimal gap between training and testing loss reflects the model's robustness and generalization capability.

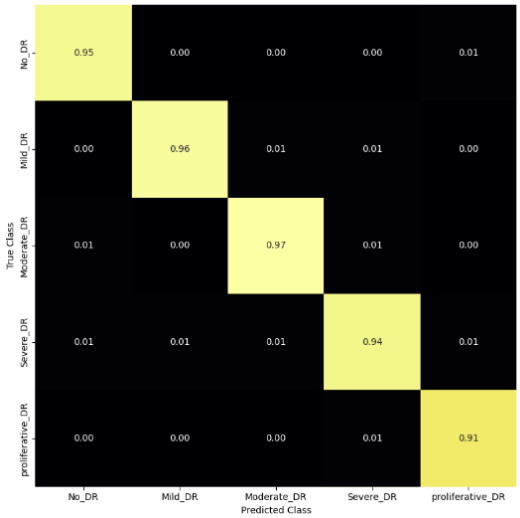

Figure 6: Confusion Matrix

Figure 6 presents the confusion matrix for the ViT-XGBoost model, illustrating the classification results across the five diabetic retinopathy stages: No DR, Mild DR, Moderate DR, Sever DR, and Proliferative DR. The matrix also points out that majority of instances in each category can be treated correctly especially with very few cross overs especially between closely related adjacent stages. This is in agreement with the high accuracy of the model in predicting a spectrum of severity in diabetic retinopathy. The minor incorrectly classified datasets present possible enhancements, for instance, increased differentiation of patients in cases like Mild and Moderate DR.

Classification Results

The section present the classification results of diabetic retinopathy by images analyzed through our developed model in the next section. Our model classified the retinal images into five different severity levels: No Diabetic

Journal of Information Systems Engineering and Management
2025, 10(40s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Retinopathy, Mild Diabetic Retinopathy, Moderate Diabetic Retinopathy, Severe Diabetic Retinopathy, and Proliferative Diabetic Retinopathy. It used a ViT or Vision Transformer to extract the features, and then it used an XGBoost classifier to further augment the accuracy of the prediction. Annotated results show how the model could indeed characterize the various stages of diabetic retinopathy, therefore contributing to its potential application in early diagnosis and timely intervention of patients at risk. For each of these images, the appropriate severity score would indicate the usability of the model in giving a decisive and explicatory assessment of the degree of diabetic retinopathy, which was an important feature in deciding on patient management and clinical decision-making.

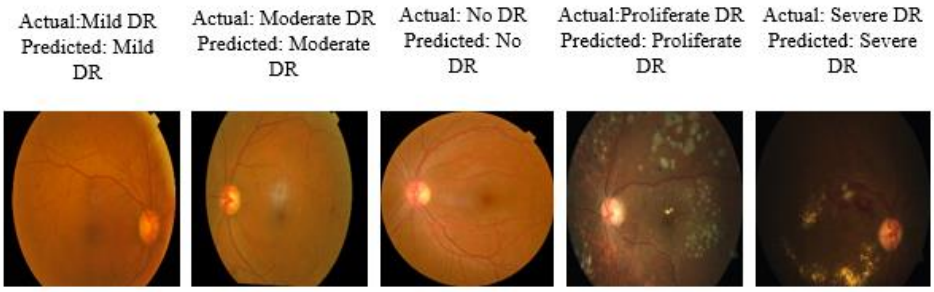

Figure 7: DR Classification Results

Figure 7 shows the images of retinal images that correspond to the five classes of diabetic retinopathy predicted by the classifier. In that No DR, displaying a healthy retina containing no signs of the disease Mild DR, where some very slight changes are observable as minute spots and slight haemorrhages are present, Moderate DR with more pronounced changes, such as an increased number of spots and retinal haemorrhages, Severe DR, characterized by significant retinal damage including multiple haemorrhages and areas with reduced blood flow. Finally, Proliferative DR, with very advanced changes in the form of new blood vessel growth, which signifies an increased risk of vision loss. These images collectively confirm the strength and accuracy of the model in grading the various stages of diabetic retinopathy, an important milestone in early detection and appropriate treatment of the patients.

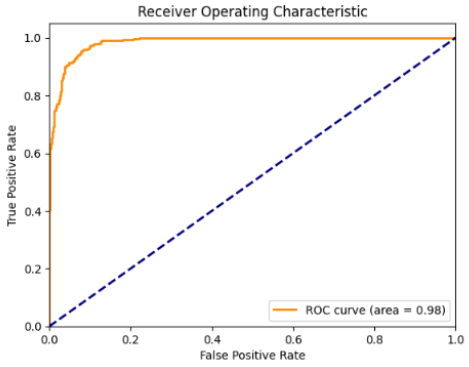

Figure 8: ROC of Proposed ViT-XGBOOST Method

Figure 8 plots the ROC curve of the proposed ViT-XGBOOST method for classifying diabetic retinopathy. The curve pushes toward the top left part of the figure, showing an excellent true positive rate and a low false positive rate; this is an essential requirement to correctly classify the stages of diabetic retinopathy. The AUC is nearly 1, indicating perfect discriminative ability between affected and non-affected patients. This translates to the fact that ViT-XGBOOST delivers a superior outcome in detecting diabetic retinopathy early; then, there is also a great chance of improving patient outcomes.

Performance Metrics

- Accuracy is defined as the proportion of correctly predicted occurrences to all instances in the dataset. Accuracy quantifies how well the system detects emotions and occurrences linked to rebellion when compared to the entire dataset. It is shown by Eqn. (15).

Journal of Information Systems Engineering and Management
2025, 10(40s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

$$Accuracy = \frac{Number\ of\ correct\ predictions}{Total\ number\ of\ predictions} \quad (15)$$

- To measure the accuracy of the model in generating positive predictions, precision calculates the percentage of true positive predictions (i.e., events associated with rebellion that are correctly recognized) over all positive predictions (true positives and false positives). It is shown by Eqn (16).

$$Precision\ = \frac{TP}{TP+FP} \quad (16)$$

- In classification issues, recall, also known as sensitivity, is a performance metric that evaluates a model's ability to find all relevant instances of a certain class in a dataset. Eqn. (17) indicates that.

$$Recall\ = \frac{TP}{TP+FN} \quad (17)$$

- The F1 score, which is a single figure that represents the accuracy of a model, is calculated by adding precision and recall. Because it accounts for both false positives and false negatives, it is particularly useful in scenarios where the distribution of classes is not equal. As shown by Eqn. (18).

$$F1\ Score = 2 * \frac{(Precision*Recall)}{(Precision+Recall)} \quad (18)$$

- The ROC curve plots the True Positive Rate (Recall) against the False Positive Rate for different thresholds. The AUC gauges the model's overall performance in differentiating between positive and negative classifications. As indicated by Eqn. (19),

$$AUC = \int_0^1 TPRd(FPR) \quad (19)$$

Table 1: Performance Comparison with Existing Methods

Methods	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
GCN[14]	81.2	80.56	81.25	80.14
Convolutional Auto-Encoder[15]	84.17	82.37	84.61	83.45
GA-SSAE [16]	98	97.5	97.4	98.5
Proposed ViT-XGBOOST	99.5	98.5	98.6	98.7

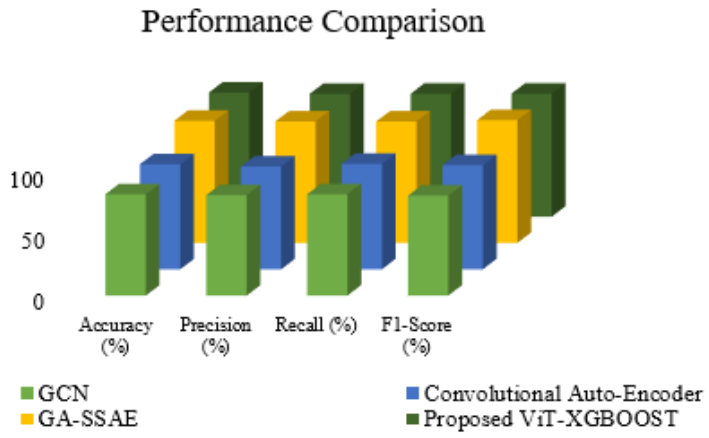

Figure 9: Performance Comparison with Existing Methods

The proposed ViT-XGBoost model therefore outperforms previous methods by achieving a high accuracy of 99.5% compared to the GA-SSAE, Convolutional Auto-Encoder, and GCN models it is given in Table 1. The precision, recall, and F1-score of ViT-XGBoost stand at 98.5%, 98.6%, and 98.7% respectively thus showing superior performance in diabetic retinopathy detection. ViT-XGBoost demonstrates superior capabilities in extraction of feature and classification in comparison to GA-SSAE with its precision of 98%. These results affirm the effectiveness of hybrid optimization in combination with Vision Transformers for feature extraction and XGBoost for classification. It is depicted graphically in Figure 9.

Discussion

Results presented here demonstrate that the proposed ViT-XGBoost model surpasses all the existing methods by a high margin and classifies diabetic retinopathy with an excellent accuracy of 99.5%. Excellent performance was delivered due to the excellent ability of Vision Transformers (ViT) to extract features coupled with the efficiency of classification delivered by XGBoost. The ViT-XGBoost model has high precision, recall, and F1-score compared to models like GCN[14], Convolutional Auto-Encoders[15], and GA-SSAE[16]. ROC curves clearly demonstrate discrimination ability over different stages of diabetic retinopathy by the model while providing minimal false positives and bringing the AUC value close to 1. The results reveal the fact that ViT-XGBoost is highly efficient for early detection and provides a reliable, robust solution for improving patient outcomes in the clinics. Thus, a combination of these advanced machine learning techniques does form a significant advancement in the diabetic retinopathy detection task

6. CONCLUSION AND FUTURE WORK

In conclusion, this study succeeded in showing the effectiveness of using a hybrid model of ViT with XGBoost to classify diabetic retinopathy, achieving a highly impressive accuracy of 99.5%. The advanced pre-processing techniques - namely, advanced denoising methods employed on input images - significantly improved the quality of input images. This was crucial since it allowed for better extraction of features from the image and subsequently better classification performance. The second crucial novelty is behind using optimization algorithms such as the Hippopotamus Algorithm and the Blue Whale Algorithm in hyper parameter tuning. Through this, this model improved its predictive accuracy due to the aspect of adaptation in machine learning. It's then noticed that the integration of adaptive approaches in machine learning much prefers superior results. Another possible direction for future work is to consider other deep architectures and extend this approach by trying to include neural networks or ensemble methods in order to increase the accuracy and robustness of the classification system. Some of the architectures might be developed in combination with each other, creating synergies that improve performance metrics. Moreover, the dataset has to be extended to include a wide range of demographic variables, including age variation, ethnicity variation, and comorbid conditions. Additionally, the model has to incorporate real-world clinical scenarios so as to make it apply in real situations and generalize to different forms of clinical settings. In other words, the model has got to be effective in controlled environments while also having reactivity in real applications. Further, its practical application in real-time diagnostic tools would be revolutionary in the screening processes of diabetic retinopathy. Healthcare workers will be able to reduce their workflows in screening patients and make the screening more direct and precise; hence, there will be a reduction of morbidity and mortality rates through early detection and treatment. This outcome could decrease the number of cases of impaired vision and blindness among the people due to diabetic retinopathy and, therefore, improve public health as a whole. This research, therefore, summarizes and lays a solid foundation for further innovation in the field of diabetic retinopathy classification with a strong prospect to reach far as both technological and clinical practice innovations.

REFRENCES

[1] L. F. Nakayama et al., “Diabetic retinopathy classification for supervised machine learning algorithms,” Int J Retin Vitr, vol. 8, no. 1, Art. no. 1, Dec. 2022, doi: 10.1186/s40942-021-00352-2.

[2] Hasan et al., “Machine Learning-based Diabetic Retinopathy Early Detection and Classification Systems- A Survey.” Accessed: Oct. 01, 2024. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9509920

[3] D. S. David, "Retinal Image Classification System for Diagnosis of Diabetic Retinopathy Using Morphological Edge Detection and Feature Extraction Techniques".
[4] S. Gayathri, V. P. Gopi, and P. Palanisamy, "Diabetic retinopathy classification based on multipath CNN and machine learning classifiers," Phys Eng Sci Med, vol. 44, no. 3, pp. 639–653, Sep. 2021, doi: 10.1007/s13246-021-01012-3.
[5] G. Selvachandran, S. G. Quek, R. Paramesran, W. Ding, and L. H. Son, "Developments in the detection of diabetic retinopathy: a state-of-the-art review of computer-aided diagnosis and machine learning methods," Artif Intell Rev, vol. 56, no. 2, pp. 915–964, Feb. 2023, doi: 10.1007/s10462-022-10185-6.
[6] M. H. Mahmoud, S. Alamery, H. Fouad, A. Altinawi, and A. E. Youssef, "An automatic detection system of diabetic retinopathy using a hybrid inductive machine learning algorithm," Pers Ubiquit Comput, vol. 27, no. 3, pp. 751–765, Jun. 2023, doi: 10.1007/s00779-020-01519-8.
[7] L. Math and R. Fatima, "Adaptive machine learning classification for diabetic retinopathy," Multimed Tools Appl, vol. 80, no. 4, pp. 5173–5186, Feb. 2021, doi: 10.1007/s11042-020-09793-7.
[8] S. Gupta, S. Thakur, and A. Gupta, "Optimized hybrid machine learning approach for smartphone based diabetic retinopathy detection," Multimed Tools Appl, vol. 81, no. 10, pp. 14475–14501, Apr. 2022, doi: 10.1007/s11042-022-12103-y.
[9] Z. Liu, C. Wang, X. Cai, H. Jiang, and J. Wang, "Discrimination of Diabetic Retinopathy From Optical Coherence Tomography Angiography Images Using Machine Learning Methods," IEEE Access, vol. 9, pp. 51689–51694, 2021, doi: 10.1109/ACCESS.2021.3056430.
[10] I. Odeh, M. Alkasassbeh, and M. Alauthman, "Diabetic Retinopathy Detection using Ensemble Machine Learning," in 2021 International Conference on Information Technology (ICIT), Amman, Jordan: IEEE, Jul. 2021, pp. 173–178. doi: 10.1109/ICIT52682.2021.9491645.
[11] K. Alabdulwahhab, W. Sami, T. Mehmood, M. SA, T. Alasbali, and F. Alwadani, "Automated Detection of Diabetic Retinopathy Using Machine Learning Classifiers," European Review for Medical and Pharmacological Sciences, Jan. 2021.
[12] O. F. Gurcan, O. F. Beyca, and O. Dogan, "A Comprehensive Study of Machine Learning Methods on Diabetic Retinopathy Classification:," IJCIS, vol. 14, no. 1, p. 1132, 2021, doi: 10.2991/ijcis.d.210316.001.
[13] A. Soni and D. Avinash Rai, A Novel Approach for the Early Recognition of Diabetic Retinopathy using Machine Learning. 2021. doi: 10.1109/ICCCI50826.2021.9402566.
[14] Y. Li, Z. Song, S. Kang, S. Jung, and W. Kang, "Semi-Supervised Auto-Encoder Graph Network for Diabetic Retinopathy Grading," IEEE Access, vol. 9, pp. 140759–140767, 2021, doi: 10.1109/ACCESS.2021.3119434.
[15] J. D. Bodapati, "Stacked convolutional auto-encoder representations with spatial attention for efficient diabetic retinopathy diagnosis," Multimed Tools Appl, vol. 81, no. 22, pp. 32033–32056, Sep. 2022, doi: 10.1007/s11042-022-12811-5.
[16] S. S. Rubini and A. Kunthavai, "Genetic Optimized Stacked Auto Encoder Based Diabetic Retinopathy Classification. | EBSCOhost." Accessed: Oct. 01, 2024. [Online]. Available: https://openurl.ebsco.com/contentitem/gcd:152390016?sid=ebsco:plink:crawler&id=ebsco:gcd:152390016
[17] D. S. David and A. A. Jose, "Retinal Image Classification System for Diagnosis of Diabetic Retinopathy Using SDC Methods".
[18] J. Ouyang, D. Mao, Z. Guo, S. Liu, D. Xu, and W. Wang, "Contrastive self-supervised learning for diabetic retinopathy early detection," Med Biol Eng Comput, vol. 61, no. 9, pp. 2441–2452, Sep. 2023, doi: 10.1007/s11517-023-02810-5.
[19] S. Sundar Ph.D and S. Subramanain, "An effective deep learning model for grading abnormalities in retinal fundus images using variational auto-encoders," International Journal of Imaging Systems and Technology, vol. 33, Jul. 2022, doi: 10.1002/ima.22785.
[20] N. Nikoloulopoulou, I. Perikos, I. Daramouskas, C. Makris, P. Treigys, and I. Hatzilygeroudis, "A Convolutional Autoencoder Approach for Boosting the Specificity of Retinal Blood Vessels Segmentation," Applied Sciences, vol. 13, no. 5, p. 3255, Mar. 2023, doi: 10.3390/app13053255.
[21] "Diabetic Retinopathy 224x224 (2019 Data)." Accessed: Oct. 04, 2024. [Online]. Available: https://www.kaggle.com/datasets/sovitrath/diabetic-retinopathy-224x224-2019-data

Journal of Information Systems Engineering and Management
2025, 10(40s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

[22] Z. Gu, Y. Li, Z. Wang, J. Kan, J. Shu, and Q. Wang, "Classification of Diabetic Retinopathy Severity in Fundus Images Using the Vision Transformer and Residual Attention", doi: 10.1155/2023/1305583.
[23] W. Jiao, X. Hao, and C. Qin, "The Image Classification Method with CNN-XGBoost Model Based on Adaptive Particle Swarm Optimization," Information, vol. 12, no. 4, p. 156, Apr. 2021, doi: 10.3390/info12040156.