

Continuous Eye Disease Severity Evaluation System using Siamese Neural Networks

Muyideen Abdulraheem¹, Idowu D. Oladipo¹, Sunday Adeola Ajagbe ^{2,✉},
Ghaniyyat B. Balogun¹, Mulikat Bola Akanbi¹, and Nissi O. Emma-Adamah³

¹University of Ilorin, Ilorin, Nigeria

{muyideen, odidowu, balogun.gb}@unilorin.edu.ng, nissiea76@gmail.com

²First Technical University, Ibadan, Nigeria

sunday.ajagbe@tech-u.edu.ng

³Kwara State Polytechnic, Ilorin, Nigeria

akanbiforu@gmail.com

Abstract

Evaluating the severity of eye diseases using medical images is a very essential and routine task performed in medical diagnosis and treatment. Current grading systems which are largely based on discrete classification are unreliable and do not reflect the entire spectrum of eye disease severity. The unreliability of discrete classification systems for eye diseases is clear, as classification is subjective and done based on the personal opinion of various medical experts, which may vary. In a bid to solve these issues, this study proposes a system for determining the severity of eye diseases on a continuous range using a twin-convoluted neural network approach known as Siamese Neural Networks. This system is demonstrated in the domain of diabetic retinopathy. Samples of retinal fundus images from an eye clinic in India are taken as test cases to evaluate the performance of a Siamese Triplet network which attempts to find the distance between their image embedding. The outputs of the Siamese network when a reference image is juxtaposed with a collection of images with distant severity categories (negative images), as well as when two reference images are compared to each other, are found to have a positive correlation (95%) with originally assigned severity classes. Hence, these outputs indicate a continuous range of the severity and change in eye diseases.

Keywords: Eye Diseases · Siamese · Neural Networks · Convoluted Neural Network · Classification

Received: 23 February 2023 · Accepted: 15 March 2023 · Published: 30 March 2023.

1 Introduction

Eye diseases, if left untreated, can result in partial or complete loss of eyesight. Early identification of many eye illnesses can help avoid visual loss. The assessment of the gravity of eye diseases and the change over time in medical images is a common and crucial activity. In diabetic retinopathy, for example, as a result of retinal blood vessel injury, vascular alterations might be a key indicator of disease progression and responsiveness to treatment. The study examines three eye diseases that threaten eye vision, diabetic retinopathy, cataracts, and diabetic macular edema. Diabetic Retinopathy (DR) is one of the most common eye disease impediments arising from diabetes and it is primarily

responsible for blindness in many patients with diabetes. It is caused by injury to the retinal blood vessels. The retina is responsible for the detection of light and transmission of a signal to the brain. These impulses are decoded by the brain for it to see the objects around it. Early identification of retinal abnormalities can prevent diabetes, in addition to being critical in avoiding the advancement of DR diabetic retinopathy, as well as other problems that can result in visual loss. Systems built on the identification of particular abnormal patterns or lesions are now used to automatically screen diabetic retinopathy using fundus imaging [1] for instance, microaneurysms (tiny saccular capillary dilations that look like dark red circular dots with sharp edges on the retina backdrop), exudates (which are lipid and protein deposits in the retina that cause brilliant yellowish-white lesions with conspicuous and uneven borders.), cotton wool spots (fluffy white spots on the retina arising from axoplasmic material accumulations inside the nerve fiber layer) and hemorrhages (a large amount of blood has accumulated in the retina.) [2, 3].

Figure 1 shows the Retinal fundus image of an eye while Figure 2 illustrates the retinal fundus image cross-section. Using medical images to assess the severity of eye diseases and their course over time is a task routinely carried out by medical practitioners [4]. For a variety of eye diseases, the process of the disease may have a wide range of unending severity, subject to modification with time. However, the severity of these diseases is often divided into ordinal categories (for example normal, moderate, and severe), and changes within these categories may not generally be understood. This is further complicated by the wide range of domain specialists' interpretations of classes, which might lead to modifications in clinical therapy [5, 6]. As a result, rating severity and assessing change continuously has the potential to help better monitor disease progression, improve patient treatment, and improve medical research.

Deep Learning algorithms are gaining a lot of traction in the medical imaging sector these days as a way to tackle a variety of issues. In radiology, an example is detecting diseases or anomalies in X-ray pictures and classifying them into various kinds or severity levels [7, 8]. The majority of published work has been concentrated on predicting discrete labels including separating patient images into distinct categories. However, in the course of assessing disease severity and change by predicting absolute class labels and comparing them, granularity past the absolute class labels is not attainable [6].

Siamese neural networks, created to check the legitimacy of card (mainly credit card) signatures, are one method for analyzing the similarity between two images [9]. These networks consist of twin neural networks each able to learn an input vector's hidden representation. In addition, the networks share the same configuration, parameters, and weights and the output may be regarded as the semantic resemblance between the extrapolated representations of both input vectors [10]. The Euclidean distance of the networks' last two layers may be determined, which measures the distance between the

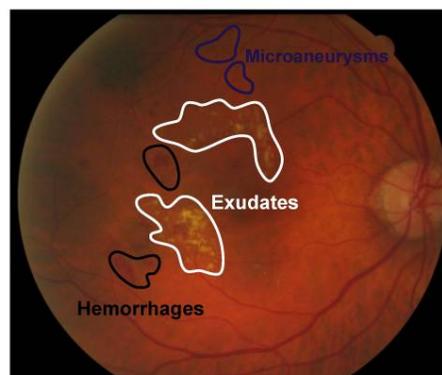


Figure 1: Retinal fundus image

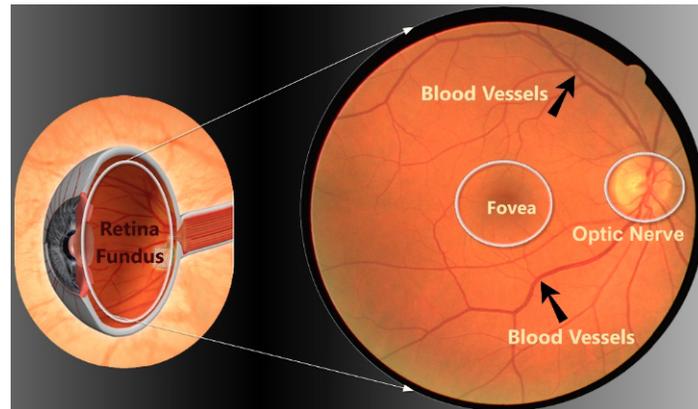


Figure 2: Retinal fundus image cross-section

two images in terms of the imaging characteristics being trained on, like illness indicators. Therefore, the task of evaluating the gravity of eye diseases on a continuous scale using images may be recast as a problem centered on metric learning. Disease severity may therefore be reduced to the median of the Euclidean distances obtained when an image is compared pairwise to a set of "normal" pictures [4]. However, this study proposes comparing an image-of-interest to both images which show the presence of the disease as well as "normal" images to better capture similarity and dissimilarity, crucial factors to consider when attempting to determine disease severity.

The routine task of evaluating eye disease severity and change over time is currently done using grading systems that employ discrete severity categories. However, these grading systems are unreliable as they are subject to domain experts' varying interpretations and opinions on severity category criteria for the diseases. The underlying continuous spectrum of severity is likewise not reflected by these distinct classifications.

In addition, since patient visits are not always regular or guaranteed, the availability of multiple images per patient taken at different time points is unlikely, this poses a challenge to evaluating the severity of eye diseases on a continuous scale as well as predicting symptom trajectories from data taken at single time points. This study realizes these challenges and proposes a system that generates a quantitative severity score that adequately represents the underlying continuous nature of eye disease severity.

This work aims to propose a system a continuous eye disease severity evaluation system using Siamese Neural Networks. The following are the objectives of the study:

- Apply a sequence of image pre-processing techniques to retinal fundus images.
- Train a U-Net for retinal vessel segmentation.
- Utilize triplet mining for selecting ideal triplet images for training
- Design a Triplet Siamese neural network for learning image embeddings.
- Evaluate results by calculating the correlation between the final output of the Siamese neural networks and disease severity grade.

2 Related Work

In a bid to develop efficient systems for evaluating and predicting eye disease severity, a plethora of studies have been conducted and a lot of varying techniques researched in multiple domains. Ting

et al., [11] reviewed articles centered on eye conditions like glaucoma and diabetic retinopathy, and macular diseases associated with age. In their study, they compiled a list of publications published between 2016 and 2018 and summarized them. They compiled a list of studies that employed fundus and optical coherence tomography pictures, as well as TL methodologies in their research. Mookiah et al., [12] reviewed computer-aided diabetic retinopathy systems which are largely based on lesions and retinal vascular alterations. Their review discusses the advantages of automated DR detection systems over manual ones.

Previous works were centered around feature extraction and image processing methods for detecting DR using a variety of computer-aided systems. Acharya et al., [13] combined morphology-based image processing techniques and support vector machine techniques for automated DR detection. Training the model in this study was made based on the presence or absence of four salient features, blood vessels, microaneurysms, exudates, and hemorrhages observed from fundus images. The study's results present the use of a Support Vector Machine (SVM) classifier which indicated an accuracy of 85 percent in predicting the unknown DR class with all four features proving significant. Similarly, Anant et al., [14] offered a strategy for detecting texture and wavelet characteristics using image processing and data mining. The results were produced for images from the DIARETDB1 standard database and assessed using sensitivity, specificity, and accuracy metrics. The method yielded a model achieving an accuracy of 97.75%.

Colomer et al., [2] proposed a method for DR detection based on the combination of morphological (granulometry) and texture descriptors from retinal images. Their suggested method eliminates the need for the segmentation of eye lesions prior to the categorization stage. Their proposed methodology creates the basis for the exploration of higher-level computer-aided diagnostic algorithms for automated DR detection and classification. Similarly, Mushtaq et al., [15] apply a Densely Connected Convolutional Network for the early detection of DR. The proposed methodology provides a robust system for automated DR detection accomplished through a series of steps including data collection, preprocessing, augmentation, and modeling. The study employs preprocessing techniques such as data augmentation to address the class imbalance in the dataset, thereby outperforming a number of machine learning classifiers such as Decision Tree (DT), K-Nearest Neighbor (KNN), Support Vector Machine (SVM) [16].

Most of the research done regarding DR severity evaluation and prediction has been based on distinct DR severity classes such as non-proliferative, severe, moderate, mild, and normal. Leeza et al., [17] proposed an enhanced automated technique for evaluating the gravity of diabetic retinopathy that does not need any pre- or post-processing stages. Pathological explicit picture representation is incorporated into a learning outline with this method. Areas of interest are identified to compute images' descriptive features of the retina using a speeded-up robust features technique and a histogram of directed gradients to generate the dictionary of visual characteristics. The five severity levels of diabetic retinopathy include normal, mild, moderate, and severe non-proliferative diabetic retinopathy, as well as proliferative diabetic retinopathy, as determined by a radial basis kernel support vector machine and a neural network. In comparison to the stated state-of-the-art approaches, the suggested system achieved better results with 95.92 percent sensitivity and 98.90 percent specificity. A mechanism for classifying colored fundus images at various stages of non-proliferative diabetic retinopathy (NPDR) was suggested by Akram et al., [18].

The proposed system consists of a new hybrid classifier for retinal lesion detection, preprocessing, extraction of patient lesions, formulation of feature sets, and classification. Preprocessing involves removing background pixels and isolating the blood vessels and optic discs from the retina images. This study combined a Gaussian Mixture Model-based expansion of the m-Medoids-based modeling technique to form an ensemble classifier targeted at achieving higher classification accuracy. Prakash [19] proposed a severity analysis method for retinal images showing the presence of DR, involving several phases; pre-processing phase, segmentation phase, feature extraction phase, and classification phase. In the proposed method, after pre-processing and before analyzing the retinal images, the images are classified into normal and abnormal images using a Neural Network classifier based on

.....

the mean, variance, entropy, area, diameter, and number of areas obtained from the segmentation optic disk of the retinal images, as well as the mean, variance, entropy, area, diameter, and number of regions acquired from the segmented blood vessels of the images. SVM classifier is then used to evaluate DR severity based on the area and intensity of strong exudates and hemorrhages present in the images. The results proved that this work outperformed existing works in accuracy, with an average of 98% accuracy for both optic disks segmentation accuracy and blood vessels segmentation. The categorization of normal and pathological retinal images using Neural Networks achieved a 90% accuracy.

Different from the previous works, some more recent works sought to derive a quantitative measurement of severity in diseases of the retina (such as Prematurity Retinopathy and Diabetic Retinopathy) that can be used to keep track of disease progression during and subsequent treatment. Gupta et al., [20] showed that a measurable Prematurity Retinopathy severity score generated by the system of deep learning may be used to follow illness progression in a group of newborns with disorders that required treatment. The study suggested that a measurable score of ROP vascular severity produced automatically using deep learning may be applied to track disease retrogression in subsequent treatments.

The main findings of this work were: the progression of ROP's measurable score of vascular severity over time was linked to the evolution of clinically observable disease (prior to treatment) and decline (upon treatment), Newborns treated with Bevacizumab showed a higher severity score at the start of the treatment and experienced faster disease decline compared to those who are laser photocoagulation-treated, patients who required re-treatment after laser treatment had a higher score of ROP severity at the time of starting therapy and a faster rate of disease than those who were not. The technology proposed has the potential to track atypical and conventional disease progression patterns, detect the recurrence of illness, and provide an objective parameter to guide future therapy and post-treatment outcome. These findings might be applied to other medical and ophthalmologic conditions where treatment is presently based on subjective clinical criteria. To address the issue of subjectivity involved in clinicians performing the routine task of evaluating disease severity,

Li et al., [4] created a Siamese neural network used to assess a single instance of disease severity and measure transformation between repeated patient visits to the clinician on a regular basis. The study demonstrated this approach in two domains; retinal images of retinopathy of prematurity (ROP) and knee radiographs of osteoarthritis. The images used were evaluated and annotated with severity classes by expert clinicians. The researchers discovered that when compared to a group of unaffected (no presence of disease) reference images, the output of the Siamese neural network for an image correlates with the disease severity score (0.87 for ROP and 0.89 for osteoarthritis), both within and between clinical grading categories. In addition, using the Siamese neural network, paired images from the same patient at two different times were juxtaposed, producing an additional continuous way of assessing change. Importantly, the method used in the study does not necessitate manually pinpointing the pathology and merely requires two alternative labels for training (same against different). The study however, in training the Siamese network, failed to explore other loss functions like the margin ranked loss and triplet loss functions which could potentially incorporate more information in the training labels, thereby improving the performance of the model.

3 Methodology

In this section, the study describes the approach taken to assess the severity of eye diseases on a continuous scale for fundus retinal images captured both at single instances of time and for the same patient longitudinally.

3.1 Data Collection

Two datasets were used in the course of the study:

- Indian Diabetic Retinopathy Image Dataset (IDRiD)
- Digital Retinal Images for Vessel Extraction (DRIVE)

The strategy employed for data collection was sourcing images from the Indian Diabetic Retinopathy Image Dataset (IDRiD), the first dataset of this kind that is typical of the Indian populace. Moreover, it is the only dataset featuring pixel-level observations of typical diabetic retinopathy lesions and normal retinal frameworks. A retinal doctor at an Eye Center in Nanded, Maharashtra, India took the fundus photos. The images were taken with a 50-degree field of view Kowa VX-10 alpha digital fundus camera (FOV) and were verified by experts to be of adequate quality, clinical relevance, and containing a decent combination of disease grades indicative of diabetic retinopathy (DR). Of the thousands of examinations available, 516 images with a resolution of 4288x2848 pixels were chosen and saved in jpg format with an average size of roughly 800 KB. Expert annotations of typical diabetic retinopathy impairment and normal retinal components are included in the dataset. It also provides information regarding the severity of diabetic retinopathy and diabetic macular edema for each picture in the database, based on global clinical relevance criteria. The image data was downloaded as a compressed zip folder from the source database to a computer system. The 516 samples were divided into 413 and 103 training and testing samples respectively. Each image in the dataset is annotated with a label in the range of 0 to 5, which represents the diabetic retinopathy severity class corresponding to the image.

The Digital Retinal Images for Vessel Extraction (DRIVE) dataset is a dataset for retinal vessel segmentation. This consists of a total of 40 color fundus images; including 7 showing mild proportions of diabetic retinopathy and the other 33 images showing no presence of the disease. The images are divided into 20 training and 20 test images with corresponding image masks. The images were obtained from a diabetic retinopathy screening exercise in the Netherlands. A Canon CR5 non-mydratic 3CCD camera with a 45-degree field of view was used to capture the photographs (FOV). At 768 by 584 pixels, each image was taken utilizing an 8-bit-per-color plane. Each image has a circular field of view (FOV) with 540 pixels diameter. The cropped images near the FOV were used for this database. A delimited FOV image mask is provided for each image. The DRIVE database was created to allow for comparative studies on blood vessel segmentation in retinal images and has proven useful in several studies in this regard.

3.2 Data Preprocessing

The proposed method outlines the steps involved in transforming the data before feeding it into the image segmentation and Siamese neural networks. Image Augmentation: Achieving high performance on Convolutional neural networks often requires having sufficient data samples for training. Hence, several image augmentation techniques were applied to the DRIVE dataset which contains only 20 training samples to increase the training sample size. The following augmentation techniques in the Argumentation library in Python were employed for the dataset thereby increasing the sample size to 120.

- Horizontal Flip
- Vertical Flip
- Elastic Transform
- Grid Distortion

- Optical Distortion

During the augmentation, extra attention is paid to make sure that the masks are getting the same changes that are being applied to the images.

Histogram Equalization: According to Reza [21], Histogram equalization is a relatively basic approach for image improvement as captured real-time series of images, in their true form may lack excellent viewing quality owing to the absence of adequate lighting or intrinsic noise. A notable method of interest that is widely used for the enhancement of still images is Contrast Limited Adaptive Histogram Equalization (CLAHE). CLAHE is an enhanced version of Adaptive Histogram Equalization (AHE). Both surmount the limitations of standard histogram equalization. The algorithm is used to partition the images to be applied to the Siamese network into contextual sections and histogram equalization is applied to each one. This levels off the distribution of used grey pixel values and thus makes concealed features of the image more glaring. The complete grey spectrum is used to show the image.

Rotation/Resizing/Mirroring/Brightness (for too-dark images): In this pre-processing stage, Random rotation (range of 10°), horizontal flipping (50% probability), random pixel cuts (to 512×512 pixels), and random brightness and contrast variation (range of 3%) were applied to the photos. For validation and testing purposes, images were center cropped (to 512×512 pixels). Every color image was converted to a grey scale.

3.3 Image Segmentation

Image segmentation is a technique in image processing that partitions an image into regions based on particular features and properties. It is aimed at making the image easier for analysis and achieving better results in computer vision tasks.

To achieve the required granularity and specificity required for determining severity using the Triplet network, a convolutional neural network built with the U-Net architecture was trained to perform Retinal Vessel Segmentation.

Joshua et al., [22] pointed out that for precise and speedy detection of eye illnesses such as glaucoma, diabetic retinopathy, and macular degeneration, proper examination of retinal fundus images is critical. U-Net was chosen to perform image segmentation for the following reasons:

- It has been proven to perform well when applied to image segmentation tasks since it performs image localization by predicting the retinal images pixel by pixel.
- U-Net does not require a large amount of training data, making it perfect for biomedical image analysis because of the relative paucity of images in the field of biomedicine.

The basic U-Net architecture employed is shown in Figure 3. After image augmentation was used to significantly increase the data set, the images were sorted and passed to the U-Net model for training. The network was trained on a batch size of 2 and for 50 epochs with 55 steps per epoch. Afterward, the model is evaluated for important performance metrics.

3.4 Siamese Network Architecture

Three identical subnetworks with shared weights were used to build a Siamese neural network using a triplet loss function. Each subnetwork is a pre-trained ResNet-101 convolutional neural network trained on the ImageNet dataset and all networks are joined at the output. Figure 4 shows the Siamese Network.

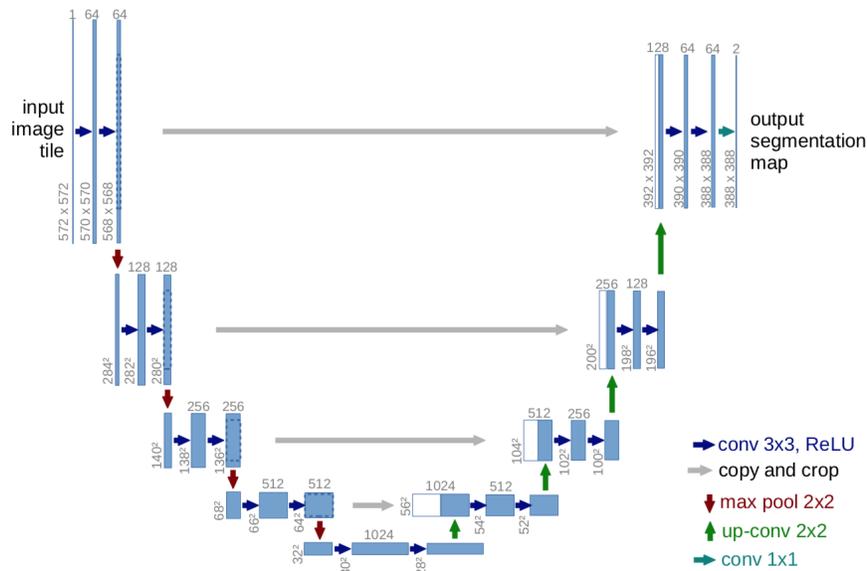


Figure 3: U-net architecture

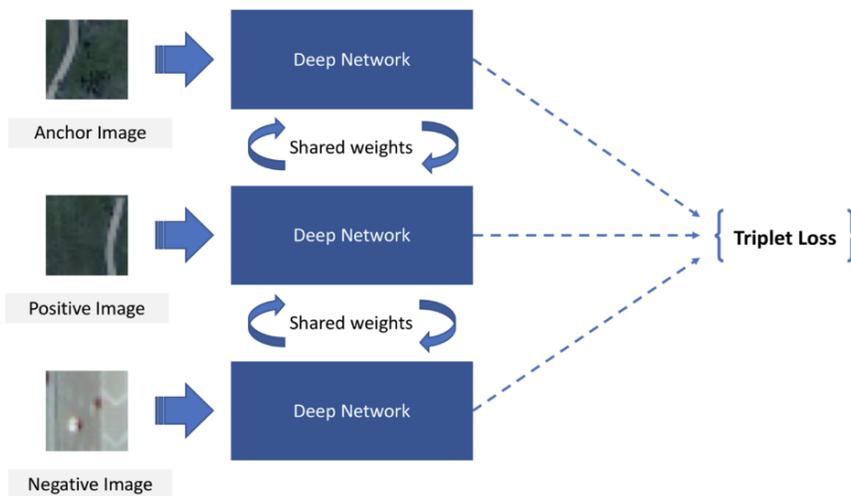


Figure 4: Siamese Network

3.5 Triplet Mining Algorithm

This algorithm was used to determine the selection of triplet samples based on the severity class of the annotated anchor image before being fed to the Siamese network. Given the triplet architecture of the Siamese network, the data is divided into triplet image pairs (anchor, positive and negative images) based on the proposed algorithm.

3.6 Severity Evaluation Method

The input to the Siamese network architecture is a pair of three images, an anchor image, a positive image (similar to the anchor), and a negative image (dissimilar from the anchor). Each image pair is passed through the pre-trained network which seeks to learn the distributed embedding of the images by the notion of similarity and dissimilarity to the anchor image.

The anchor (a), positive (p), and negative (n) images are passed through their respective convolutional neural networks, and during backpropagation weight vectors are updated using the shared architecture. The Euclidean distance between the images is then calculated from the final connected layer. This distance is representative of the difference between the images and is abstracted to represent the severity score for the image.

3.7 Model Deployment

The goal of building any system is to make it available to the end user, Deep learning models are no exception. Model deployment is one of the final and most challenging steps in any Machine Learning workflow. Models are usually deployed via API, web, or app interfaces which users can interact with for their intended use cases. The study employs Flask which is a web application framework to deploy the U-Net and Siamese networks for use in production. Flask's variety of tools and libraries make it easy to integrate Machine Learning models into its API.

3.8 Performance metrics

These metrics are used to evaluate the performance of trained models. In this study, we evaluate the performance of two models; the U-Net model used for Retinal vessel segmentation and the Triplet Siamese model. The U-Net model was evaluated on the:

- Accuracy: Model accuracy is a metric used to determine which model performs best at recognizing correlations and patterns between variables in a dataset based on the input, or training, data (see Equation 1, where tp is truly positive, fp is false negative, tn is true negative).

$$A = \frac{tp + tn}{tp + tn + fp + fn} \quad (1)$$

- Precision: Model precision measures what proportion of positive predictions was correct. It is defined as the number of true positives divided by the sum of the number of true positives and false positives (see Equation 2, where tp is truly positive, fp is false negative).

$$P = \frac{tp}{tp + fp} \quad (2)$$

- Recall/Sensitivity: The recall of a model is a measure of how well the model detects True Positives. It is the total number of cases correctly predicted as positive of the total positive cases. For this use case, recall measures the ability of our U-Net to correctly identify the location of white lines drawn for the mask generated from each eye fundus image (see Equation 3, where tp is truly positive, fn is false negative).

$$S_e = \frac{tp}{tp + fn} \quad (3)$$

Furthermore, the Siamese model used in this research is evaluated based on the correlation between its output and the disease severity class of the fundus images. All processes involved in the

research, including image preprocessing and model training, were carried out on a MacBook Pro (2015) with a 2.7 GHz Dual-Core Intel Core i5 processor having a Random-Access Memory (RAM) of 8 gigabytes. All development efforts were made in Python. Python is an interpreted multipurpose programming language. The substantial design philosophy used stresses code readability through the use of indentation, and its language components and object-oriented paradigms are aimed at assisting programmers in writing clear and logical code for both small and big projects. Python is the preferred programming language for Data Science tasks because it is simple to use, has good library availability, active community engagement, and a dynamic selection of libraries and resources. Inactive communities are less likely to maintain or develop their platforms, which is not the case with Python. The principal Python libraries used in the course of the implementation were Pandas, PyTorch, TensorFlow/Keras, Scikit-Learn, and Flask.

3.9 Why the Siamese Neural Network

The Siamese Neural Network was chosen for this investigation because of its capacity to predict better with fewer input data. Siamese networks have grown in popularity in recent years due to their capacity to learn from relatively little input. With the help of one-shot learning, a few images per class are enough for Siamese networks to recognize images. Siamese learning requires a small number of parameters (weights), hence training the network will likely converge faster and the network will be less volatile to noise. Using Siamese to build the model avoids the model computing the output for the input just once, and then using the cached results for the existing images in the database to quickly compute the similarity measures.

SNN focuses on learning embeddings (in the deeper layer) that group comparable classes/concepts together and hence decide on semantic similarity. Also, Siamese networks learn from semantic similarity rather than training mistakes or mispredictions. This helps the model to acquire even better embeddings that represent images from the support set and bring related ideas closer together in the feature space. Instead of only extracting static features using convolutions, the model learns ideas and seeks to comprehend why certain images are more similar than others by learning such a feature space, similar to how textual models learn word embeddings. The capacity to provide benchmark performance on extraordinarily little data is the most immediately relevant advantage. The problem of class imbalance disappears when the data need is minimized.

4 Results Analysis

The experimental results are presented for the model employed for image segmentation as well as the Siamese network outputs.

4.1 Results of Training and Testing U-Net model

The results of the prediction of a U-Net model which was trained on 120 retinal fundus images using dice loss: accuracy 95%, precision 38%, recall 95%, F1 score 55%. The high accuracy (95%) of the model shows its overall ability to correctly segment the retinal vessels from well-captured retinal fundus photographs. However, the low precision (38%) value results from the fact that Image segmentation tasks are not as straightforward as linear prediction tasks largely because image segmentation tasks require that the test image is properly captured, including the presence of proper lighting conditions and powerful lens cameras. Therefore, the precision value here would be a result of not being able to capture the image structure for certain poorly taken images. The high recall value (95%) implies that of the properly taken retinal images, the model could accurately produce a close-to-identical image mask for the larger number. This indicates that retinal fundus images taken with high-resolution cameras which produce fundus images with visibly seen vessels can perform well

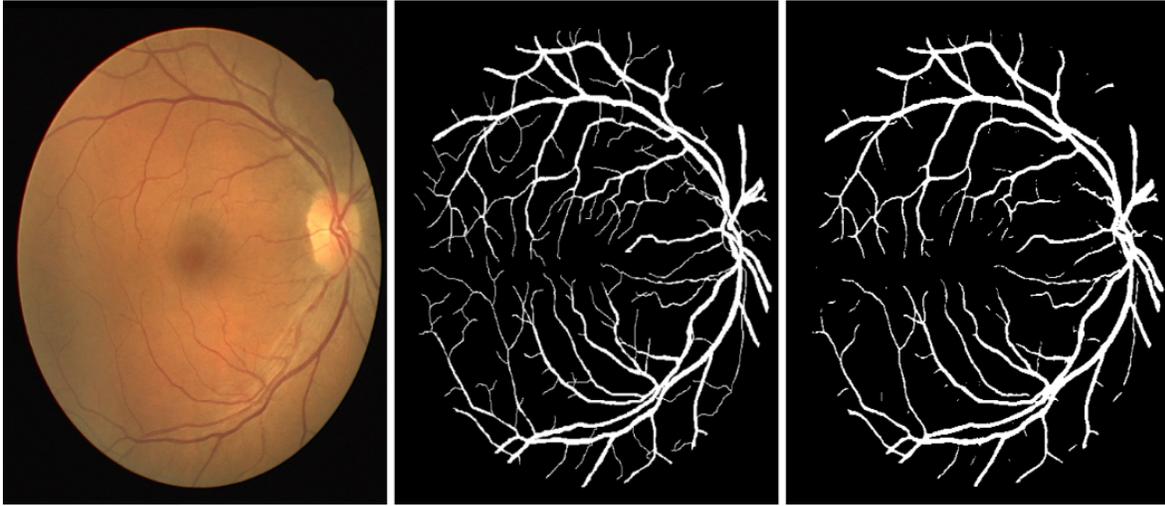


Figure 5: U-net fundus image and mask comparison

on this model. Figure 5 shows the comparison of the image masks generated by the trained U-Net models and the original annotated masks from the test set.

4.2 Severity Evaluation

This section of the work presents key findings about eye disease severity determination using trained Siamese neural networks on single images and images taken for the same patient at different time points.

4.2.1 Evaluation of disease severity for single images

Disease severity was evaluated for images taken at single time points by comparing the single image (anchor) to a pool (5) of randomly selected dissimilar (negative) images following the Triplet mining algorithm described earlier.

First, the single image is passed through the U-Net model described above then both anchor and negative images are passed through a Siamese network comprised of ResNet-101 models which generate an image embedding for each image.

The Siamese neural network is used to compute the Euclidean distance between the image pairs. The median of these computed Euclidean distances is interpreted as the measure of disease severity. These Euclidean distances are plotted against their numerical severity grade in the figure above. The findings in Figure 6 shows a correlation between the Euclidean distance of the image embeddings and the grade of disease severity assigned for the image. This implies the following:

- That the severity score generated is a valid representation of eye disease severity.
- Image segmentation enables the Siamese network to achieve better specificity in severity prediction.

4.2.2 Evaluation of disease severity change longitudinally

Evaluating a pair containing two images taken from a particular patient at different instances of time to determine the severity of disease and the change on a continuous scale occurs in two steps:

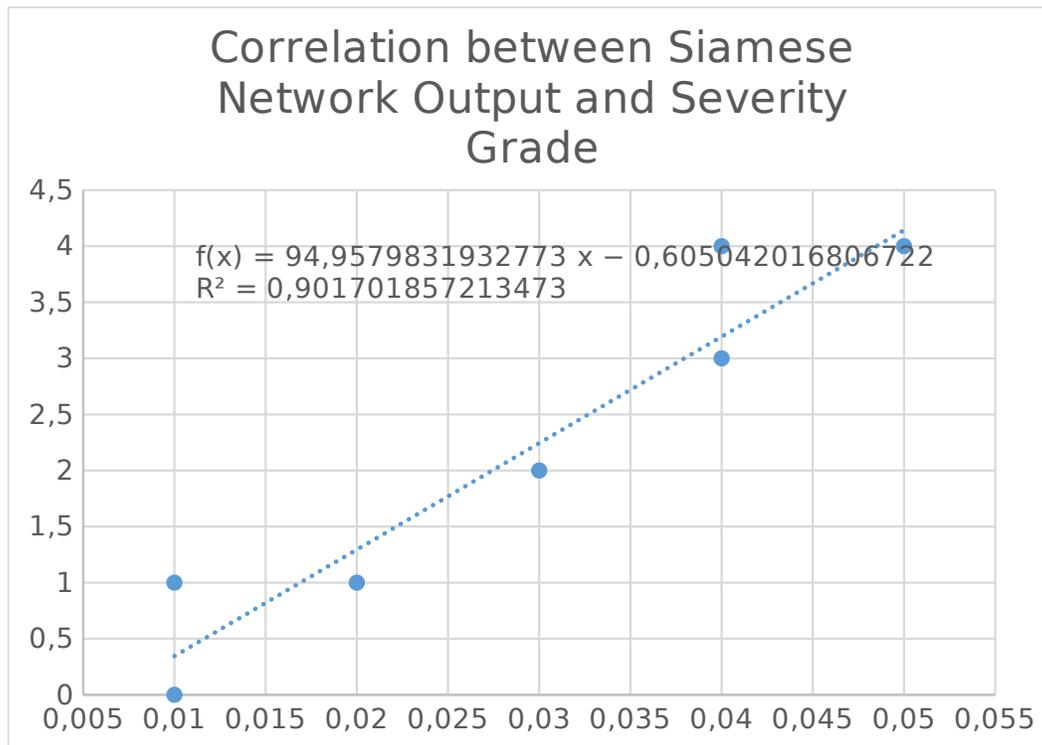


Figure 6: Scatter plot between the Siamese network output and severity

- Two images in the pair are inputted directly to the network which computes their Euclidean distance representing disease severity.
- The Euclidean distance for every single image in the pair is calculated using the earlier proposed algorithm. The computed difference between these Euclidean distances is then used to quantify disease severity and change.

From the scatter plot in Figure 6,

- It is observed moderate correlation between the longitudinal disease severity change between two images taken at different time points and the difference between the severity scores generated for both images using the earlier proposed method.
- This demonstrates that disease severity change in the same patient can reasonably be measured over time by finding the distance between any two image samples computed by the Siamese network.

5 Model Deployment

The proposed system is deployed via a Flask web application providing a comfortable interface for medical practitioners to work with. The web application is deployed to a production environment via Heroku for ease of access. The flask web application provides the following features:

- An interface where users can navigate between single disease severity evaluation and, multi-point disease severity evaluation.

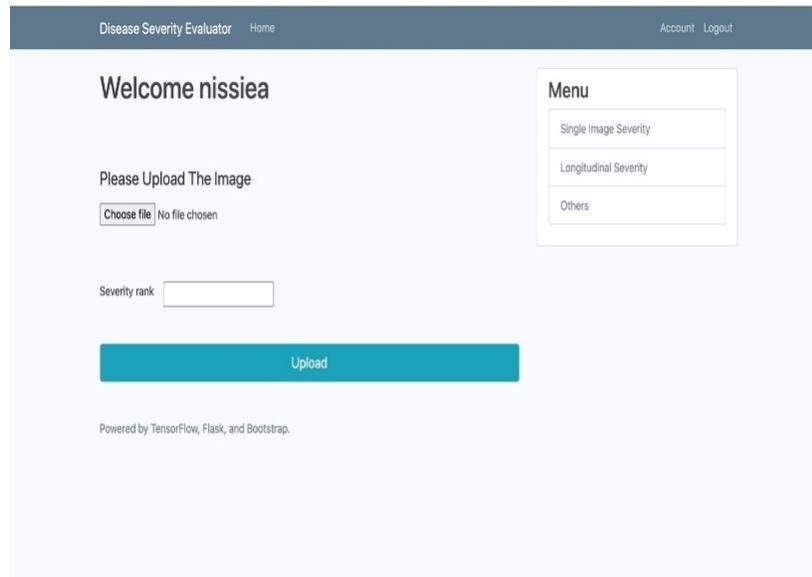


Figure 7: Web evaluator home page

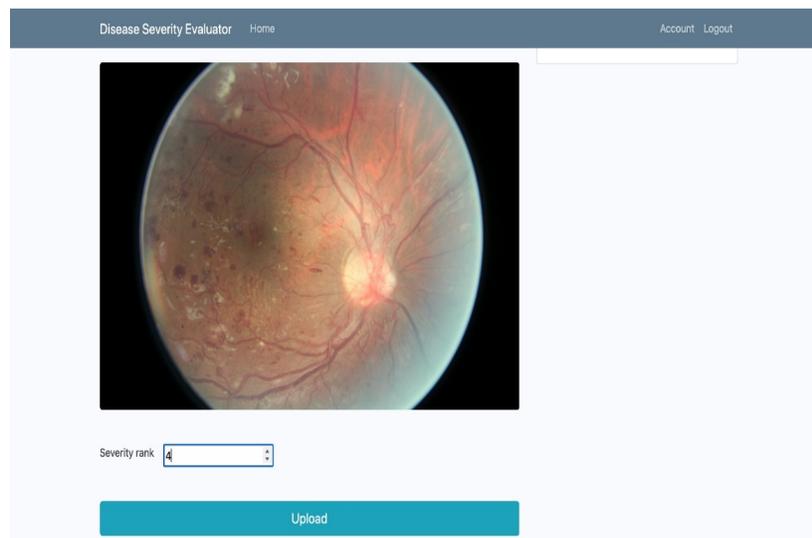


Figure 8: Web app fundus image upload

- Functionality that enables users to upload retinal fundus images to get a severity score back as a response.

Figures 7 and 8 show key features of the application.

6 Future Work

We anticipate a future in which continuous disease severity ratings may be utilized to more consistently quantify disease severity and follow progress over time. This might be useful in ordinary health

treatment as well as clinical research. Prior work on deep learning classification has investigated the benefit of adopting a continuous measure for performance evaluation to increase model accuracy, however, that classification approach ultimately predicts ordinal illness grades. Using measurements obtained from Euclidean distance output, our Siamese neural network technique presents a potentially generalizable solution for producing continuous disease severity grading scales for medical pictures.

This work has a number of limitations. Initially, the training labels were constructed from the ordinal illness severity class labels. Instead of labels assigned by expert comparisons of pictures, future research may integrate labels assigned by expert comparisons of images for training. In addition to potentially enhancing training performance, the comparison labels might be used with ordinal illness severity class labels for training, thus reducing the number of pictures that need to be annotated.

Second, we established the applicability of our method to two-dimensional imaging issues, despite the fact that typical medical imaging requires three-dimensional data. Further research will need to look at the applicability of this method to volumes. Finally, the models in this work were trained using a contrastive loss function on the Euclidean distance computed from the last completely connected layers. Other techniques, such as evaluating the loss function on earlier network layers or using totally different loss functions, such as the triplet loss function, margin rank loss, and other changes to the training methodology, may increase model performance. Alternative loss functions may include more information in the labels, such as magnitude and directionality, which might possibly boost learning even more.

7 Conclusions

The study demonstrated the use of Siamese neural networks to evaluate eye disease severity on a continuous scale. The wide range of literature reviewed in the course of this research revealed the majority of efforts geared towards evaluating disease severity evaluation have been done by employing machine learning models generating discrete severity categories, it also identified the failings of these systems. The study demonstrated the utility of employing image segmentation before disease severity evaluation and leveraging the ability of pre-trained models. The research met its key objectives by demonstrating in two scenarios that evaluating eye disease severity quantitatively better represents its true underlying continuous nature. Subsequent research can train models with the use of data with expertly assigned severity grades, thus limiting the subjective input of medical practitioners. Severity scores generated can be used automatically to reconstruct images that would predict the progression of the disease given the severity change across time. Investigating the applicability of this approach to 3-dimensional images which are increasingly prevalent in medical imaging and radiography.

Authors' Information

- **Muyideen Abdulraheem** is a Lecturer I in the Department of Computer Science, University of Ilorin, Ilorin, Nigeria. He obtained a Bachelor of Technology Computer Science at the Prestigious Abubakar Tafawa Balewa University Bauchi, in 1997, a Master of Science in Mathematics (Computer Option), and a Ph.D. Computer Science at the University of Ilorin, Nigeria. His research area is computer security, grid computing security, and data and information security.
- **Idowu D. Oladipo** received his B.Sc. (Edu.) degree in Computer Science from Ekiti State University, Ado-Ekiti, Ekiti State, Nigeria in 2005. He earned his M.Sc. and Ph.D. degrees in Computer Science from the University of Ilorin, Nigeria, in 2010 and 2018 respectively. Since 2019, he has been a Lecturer II with the Department of Computer Science, University of Ilorin, Ilorin, Nigeria. His research interests include Software Engineering, Bioinformatics, Information Security,

Artificial Intelligence, Cyber Security, and Computer Education. He is a member of the International Computer Professional Registration Council of Nigeria (MCPN) and Nigeria Computer Society (MNCS).

- **Sunday Adeola Ajagbe** is a faculty member at the Computer and Industrial Production Engineering Department, First Technical University, Ibadan, and a Ph.D. candidate at the Computer Engineering Department, Ladoke Akintola University of Technology (LAUTECH), both in Nigeria. He obtained a Master of Philosophy in the Computer Engineering Department at LAUTECH, MSc and BSc in Information Technology and Communication Technology respectively at the National Open University of Nigeria (NOUN), and his Postgraduate Diploma in Electronics and Electrical Engineering at LAUTECH. He specializes in Artificial Intelligence (AI), Natural language processing (NLP), Information Security, Data Science, the Internet of Things (IoT), Sensors, Communications, and Signal processing. He is also licensed by The Council Regulating Engineering in Nigeria (COREN) as a professional Electrical Engineer, and a student member of the Institute of Electrical and Electronics Engineers (IEEE), and the International Association of Engineers (IAENG).
- **Ghaniyyat B. Balogun** received her B.Tech. degree in Mathematics/Computer Science from the Federal University of Technology, Minna, Nigeria in 1998. She obtained a Master's degree in Business Administration (MBA), together with an MSc. in Mathematics/Computer Science and Ph.D. in Computer Science degrees from the University of Ilorin, Nigeria in 2003, 2008, and 2015 respectively. She is presently a lecturer in the Computer Science Department of the University of Ilorin, Nigeria. Her research interests fall in the areas of Data Structures, Big Data, Data Mining, Data Analytics, and Database Management Systems. She is a member of the International Computer Professionals Registration Council of Nigeria (MCPN), Nigeria Computer Society (NCS), the International Association of Engineers (IEANG), the Institute of Strategic Management of Nigeria (ISMN), Nigerian Women in Information Technology (NIWIT) among others.
- **Mulikat Bola Akanbi** is a lecturer at the Department of Computer Science, Institute of Information and Communication Technology, Kwara State Polytechnic, Ilorin, Nigeria. She obtained her B.Sc and M.Sc in Computer Science from the University of Ilorin in 2003 and 2014 respectively. She is currently a Ph.D. student in the same University. Her research interests include Information Security, Computer Networks, Data mining, Biometric authentication, and Template protection. She is a professional member of the Nigeria Computer Society (NCS), Academia in Information Technology (AITP), Nigeria Women in Information Technology (NIWIIT), and Women in Technical Education and Development (WITED).
- **Nissi O. Emma-Adamah** is a graduate of Computer Science from the University of Ilorin, Nigeria. He graduated in the year 2022.

Authors' Contributions

- **Muyideen Abdulraheem** participated in conceptualization, review, Validation, and Supervision.
- **Idowu D. Oladipo** participated in the writing of the original draft, methodology, and coding.
- **Sunday Adeola Ajagbe** participated in the methodology, resource management, writing of the original draft, editing, and supervision
- **Ghaniyyat B. Balogun** participated in resource management, project administration, and visualization.

- **Mulikat Bola Akanbi** participated in resource management and editing.
- **Nissi O. Emma-Adamah** participated in editing, review and supervision.

Competing Interests

The authors declare that they have no competing interests.

Funding

No funding was received for this project.

References

- [1] C. P. Wilkinson, F. L. Ferris III, R. E. Klein, P. P. Lee, C. D. Agardh, M. Davis, D. Dills, A. Kampik, R. Pararajasegaram, J. T. Verdager, *et al.*, "Proposed international clinical diabetic retinopathy and diabetic macular edema disease severity scales," *Ophthalmology*, vol. 110, no. 9, pp. 1677–1682, 2003.
- [2] A. Colomer, J. Igual, and V. Naranjo, "Detection of early signs of diabetic retinopathy based on textural and morphological information in fundus images," *Sensors*, vol. 20, no. 4, p. 1005, 2020.
- [3] S. A. Ajagbe, K. A. Amuda, M. A. Oladipupo, F. A. Oluwaseyi, and K. I. Okesola, "Multi-classification of alzheimer disease on magnetic resonance images (mri) using deep convolutional neural network (dcnn) approaches," *International Journal of Advanced Computer Research*, vol. 11, no. 53, p. 51, 2021.
- [4] M. D. Li, N. T. Arun, M. Gidwani, K. Chang, F. Deng, B. P. Little, D. P. Mendoza, M. Lang, S. I. Lee, A. O'Shea, *et al.*, "Automated assessment and tracking of covid-19 pulmonary disease severity on chest radiographs using convolutional siamese neural networks," *Radiology: Artificial Intelligence*, vol. 2, no. 4, p. e200079, 2020.
- [5] A. B. Rosenkrantz, R. Duszak Jr, J. S. Babb, M. Glover, and S. K. Kang, "Discrepancy rates and clinical impact of imaging secondary interpretations: a systematic review and meta-analysis," *Journal of the American College of Radiology*, vol. 15, no. 9, pp. 1222–1231, 2018.
- [6] J. P. Campbell, J. Kalpathy-Cramer, D. Erdogmus, P. Tian, D. Kedarisetti, C. Moleta, J. D. Reynolds, K. Hutcheson, M. J. Shapiro, M. X. Repka, *et al.*, "Plus disease in retinopathy of prematurity: a continuous spectrum of vascular abnormality as a basis of diagnostic variability," *Ophthalmology*, vol. 123, no. 11, pp. 2338–2344, 2016.
- [7] M. Kim, J. Yun, Y. Cho, K. Shin, R. Jang, H.-j. Bae, and N. Kim, "Deep learning in medical imaging," *Neurospine*, vol. 16, no. 4, p. 657, 2019.
- [8] S. A. Ajagbe, O. A. Oki, M. A. Oladipupo, and A. Nwanakwaugwum, "Investigating the efficiency of deep learning models in bioinspired object detection," in *2022 International Conference on Electrical, Computer and Energy Technologies (ICECET)*, pp. 1–6, IEEE, 2022.
- [9] J. Bromley, I. Guyon, Y. LeCun, E. Säcker, and R. Shah, "Signature verification using a "siamese" time delay neural network," *Advances in neural information processing systems*, vol. 6, 1993.
- [10] D. Chicco, "Siamese neural networks: An overview," *Artificial neural networks*, pp. 73–94, 2021.

-
- [11] D. S. W. Ting, L. R. Pasquale, L. Peng, J. P. Campbell, A. Y. Lee, R. Raman, G. S. W. Tan, L. Schmetterer, P. A. Keane, and T. Y. Wong, "Artificial intelligence and deep learning in ophthalmology," *British Journal of Ophthalmology*, vol. 103, no. 2, pp. 167–175, 2019.
 - [12] M. R. K. Mookiah, U. R. Acharya, C. K. Chua, C. M. Lim, E. Ng, and A. Laude, "Computer-aided diagnosis of diabetic retinopathy: A review," *Computers in biology and medicine*, vol. 43, no. 12, pp. 2136–2155, 2013.
 - [13] U. R. Acharya, C. M. Lim, E. Y. K. Ng, C. Chee, and T. Tamura, "Computer-based detection of diabetes retinopathy stages using digital fundus images," *Proceedings of the institution of mechanical engineers, part H: journal of engineering in medicine*, vol. 223, no. 5, pp. 545–553, 2009.
 - [14] K. A. Anant, T. Ghorpade, and V. Jethani, "Diabetic retinopathy detection through image mining for type 2 diabetes," in *2017 International Conference on Computer Communication and Informatics (ICCCI)*, pp. 1–5, IEEE, 2017.
 - [15] G. Mushtaq and F. Siddiqui, "Detection of diabetic retinopathy using deep learning methodology," in *IOP conference series: materials science and engineering*, vol. 1070, p. 012049, IOP Publishing, 2021.
 - [16] J. K. Adeniyi, A. E. Adeniyi, Y. J. Oguns, G. O. Egbedokun, K. D. Ajagbe, P. C. Obuzor, and S. A. Ajagbe, "Comparison of the performance of machine learning techniques in the prediction of employee," *ParadigmPlus*, vol. 3, no. 3, pp. 1–15, 2022.
 - [17] M. Leeza and H. Farooq, "Detection of severity level of diabetic retinopathy using bag of features model," *IET Computer Vision*, vol. 13, no. 5, pp. 523–530, 2019.
 - [18] M. U. Akram, S. Khalid, A. Tariq, S. A. Khan, and F. Azam, "Detection and classification of retinal lesions for grading of diabetic retinopathy," *Computers in biology and medicine*, vol. 45, pp. 161–171, 2014.
 - [19] N. B. Prakash, D. Selvathi, and G. R. Hemalakshmi, "Development of algorithm for dual stage classification to estimate severity level of diabetic retinopathy in retinal images using soft computing techniques.," *International Journal on Electrical Engineering & Informatics*, vol. 6, no. 4, 2014.
 - [20] S. Taylor, J. Brown, K. Gupta, J. Campbell, S. Ostmo, and R. Chan, "Imaging and informatics in retinopathy of prematurity consortium. monitoring disease progression with a quantitative severity scale for retinopathy of prematurity using deep learning," *JAMA Ophthalmol*, vol. 137, no. 9, pp. 1022–1028, 2019.
 - [21] A. M. Reza, "Realization of the contrast limited adaptive histogram equalization (clahe) for real-time image enhancement," *Journal of VLSI signal processing systems for signal, image and video technology*, vol. 38, pp. 35–44, 2004.
 - [22] A. O. Joshua, F. V. Nelwamondo, and G. Mabuza-Hocquet, "Blood vessel segmentation from fundus images using modified u-net convolutional neural network," *Journal of Image and Graphics*, vol. 8, no. 1, pp. 21–25, 2020.