جامعة بيروت العربية
BEIRUT ARAB UNIVERSITY

Diabetic Retinopathy Classification Using Advanced Algorithms: EyeCare Design

By

Abdulrahman Kadhim Kadhim, 202200059

Ibrahim Alkahlil, 202203177

Final Year Project II – BIME502

Submitted in Partial Fulfilment of the Requirements for the Degree of Bachelor in Biomedical Engineering

Department of Electrical and Computer Engineering

Biomedical Engineering

Supervised by

Dr. Amira Zaylaa

Fall 2025/2026

Acknowledgements

The authors would like to acknowledge Dr. Amira Zaylaa for her guidance and support throughout the length of this project. We also thank Dr. Mohamad Houri and Dr. Ahmad Araissi for providing practical insight into the field of ophthalmology, which enabled the design of a clinically significant system.

Table of Contents

Table of Figures

Table of Equations

List of Tables

Abstract

Diabetic retinopathy (DR) is a leading cause of vision impairment globally, especially among individuals with prolonged diabetes. Traditional diagnostic methods relying on manual retinal examination are time-consuming, subjective, and often inaccessible in remote regions. This project proposes a comprehensive, automated system for early DR detection and classification using advanced machine learning (ML) and deep learning (DL) algorithms. The performance of 4 state-of-the-art transformer-based architectures was thoroughly assessed in both DR screening and grading. ConvNeXt performed best in both settings scoring an accuracy of 99.18% and a quadratic weighted kappa (QWK) of 0.91 in binary and multiclass classification respectively.

To extend accessibility, a user-friendly Graphical User Interface (GUI) and a telemedicine module were developed via MATLAB, enabling real-time DR screening and grading. The GUI also displays the processing steps done by the algorithm, performs image segmentation, and provides the use with options for automated patient report generation. This integrated framework offers a scalable, accurate, and efficient solution to DR diagnosis, particularly benefiting underserved communities.

Chapter 1: Introduction

1.1. Project Background

Diabetic Retinopathy (DR), also known as diabetic eye disease, is a medical condition that arises as a complication of type 1 and 2 diabetes, causing damage to the retina [1]. Globally, approximately 22.27% of diabetes patients suffer from DR. After 30 years of living with diabetes, this figure increases to 100% and 63% in type 1 and type 2 patients respectively [2]. In the MENA region, DR prevalence is about 33.8% in diabetes patients [3]. On the national level, 16.96% of Lebanese population suffering from diabetes have the condition [4]. If left untreated, DR results in significant vision loss and even permanent blindness. With a steady increase in the rate of DR incidence and prevalence as shown in Fig. 1 [5], the early diagnosis of the issue has become a pressing concern in the world of healthcare.

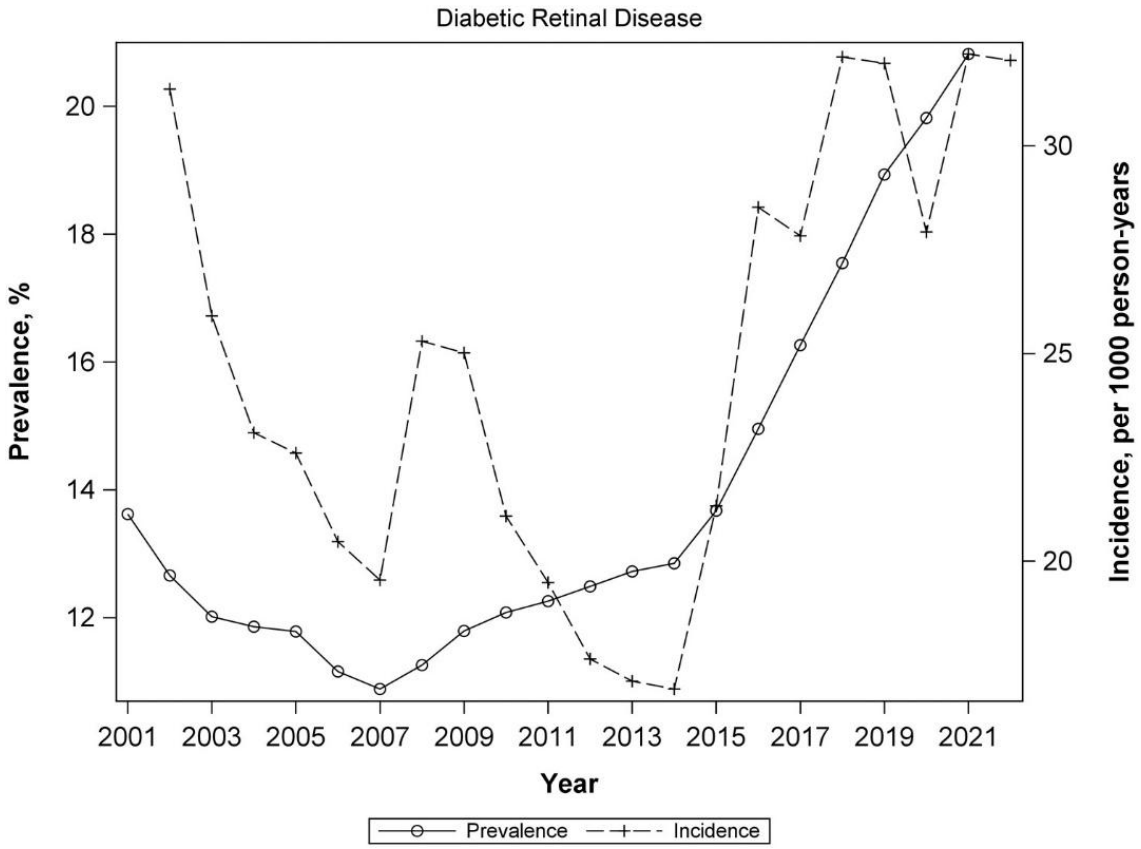

Figure 1-DR prevalence and incidence rate as a function of time in US DM patients [5]

The anatomy of the human eye could be split into 3 main parts shown in Fig. 2 [6]: an outer layer, middle layer and inner layer. The outer layer contains the cornea which is a transparent dome structure that focuses light into the eye, as well as the sclera which is the tough white layer that protects its sensitive structures. The middle layer notably houses the iris, which controls the amount of light entering the eye by dilating or contracting an opening known as the pupil. Finally, the inner layer is mainly comprised of the lens and retina. The lens is a flexible structure that changes shape to focus light into the retina for near and far vision. The retina is the light-sensitive part of the eye which contains photoreceptors capable of transforming light into an electrical signal, most notably in the macula. Moreover, photoreceptor cells are split into rods which work in low light and detect shades of black, white and grey, as well as cones which operate in bright light, and are responsible for the detection of color. Thus, in order to see, light enters through the cornea and pupil, is focused into the retina by the lens, and transformed into an electrical signal by the rods and cones of the retina which is then conducted to the visual cortex of the brain through the optic nerve [7]. DR disrupts this process by incurring damage to the retina which manifests in many forms.

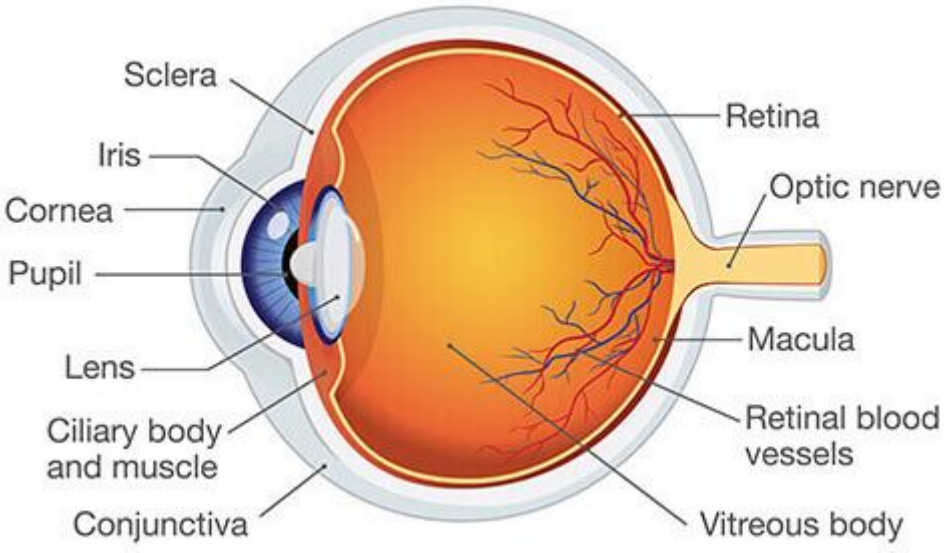

Figure 2-Anatomy of the eye [6]

The progression of DR could be classified into 2 broad stages, the first of which is non-proliferative diabetes-related retinopathy (NPDR). Damage to the retina first appears in the form of small bulges in the retinal blood vessels, known as microaneurysms. Over time, more abnormalities begin to appear in the eye such as retinal hemorrhages, hard exudates, cotton wool spots, and retinal edema. Eventually, the body grows new blood vessels to compensate for the poor blood flow to the retina, propelling the patient into the 2nd stage known as proliferative diabetes-related retinopathy (PDR). These new blood vessels often break and bleed leading to dark spots in the eye, effectively obscuring vision. These symptoms are illustrated in Fig. 3 [8].

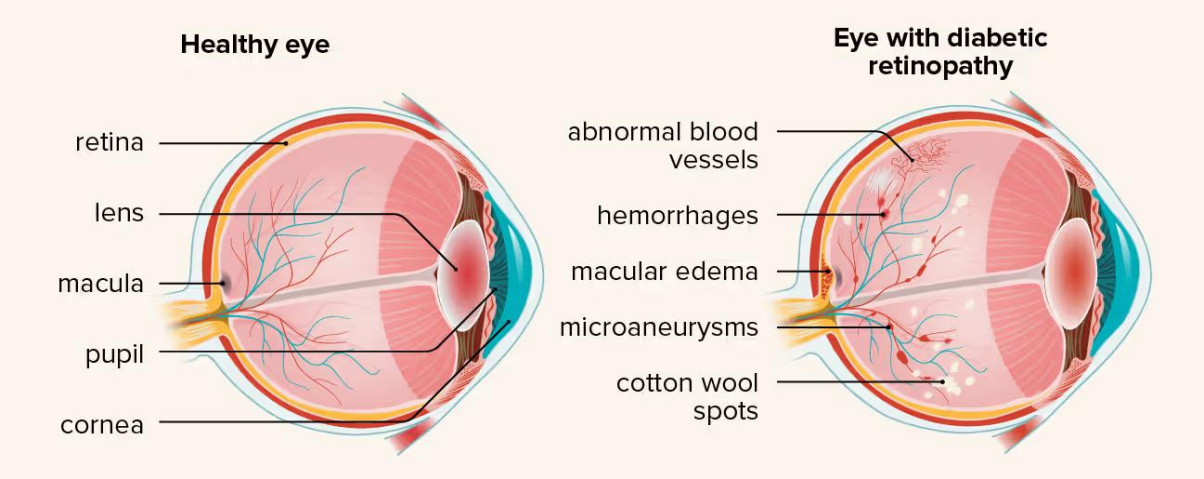

Figure 3-Healthy Eye vs Eye with DR [8]

Moreover, about half of those suffering from DR experience swelling of the macula, known as macular edema. In the case that the swelling occurs near the center of the macula, adverse vision disruptions occur ranging from blurred vision to the severe loss of the center of the visual field. Furthermore, the prolonged cycle of bleeding, scarring, and swelling may eventually result in retinal detachment. Traditional methods for DR diagnosis rely on the observations an optometrist makes, either during a clinical examination or on a retinal image. Clinical examination techniques most commonly include Direct Ophthalmoscopy in which a handheld ophthalmoscope is used to examine the retina through the pupil. As for retinal imaging techniques, fundus photography is a commonly used method which employs a digital retinal camera, to capture high resolution images of the retina. Moreover, Optical Coherence Tomography (OCT) is a high-resolution method which captures cross-sectional images of the retina, useful for the detection of retinal thickening, macular edema, and early DR changes. The technique could be paired with angiography (OCTA) to produce images that show capillary dropout, ischemia, and neovascularization more clearly [9].

Recent studies have shown the potential of employing machine learning algorithms, such as Deep Neural Networks (DNNs) and Convolutional Neural Networks (CNNs) in the process of analyzing retinal images and diagnosing DR. Although such methods have shown great promise, they remain limited by their computational intensity and use of non-exhaustive, unspecific datasets.

1.2. Problem Statement

Early detection and treatment are crucial in reducing the risk of vision loss. Using Clinical methods and conventional diagnostic methods such as manual grading of fundus images by eye experts is time-consuming, expensive, subjective and requires specialized training [10]. Moreover, subtle changes in retinal structure often go unnoticed by the ophthalmologist. This, paired with the unavailability of qualified medical professionals in rural areas constitutes a need for an automated diagnosis system that gives accurate, objective results based on a deep analysis of the retinal image.

1.3. Objective of the work

This project aims to develop a system for early detection and classification of DR using advanced algorithms. The system takes retinal images from various techniques to provide accurate and automated diagnosis, improving the early intervention and blindness prevention. Moreover, telemedicine will be used to facilitate remote screening, and DR detection will be facilitated more easily, especially in areas where people face difficulties diagnosing or having information about this disease. A mobile application developed to enable and monitor real time image analysis, diagnosis, and consultation between patients and healthcare professionals, and for an effective and smooth screening process.

1.3.1 Sub-Objectives

- Collecting and preprocessing a good dataset of fundus retinal images suitable for diabetic retinopathy classification and proper image enhancement, resizing, and normalization in favor of high-performing algorithm output.
- Implementing and evaluating various machine learning and deep learning algorithms, for automating diabetic retinopathy stages classification and calculating their accuracy, sensitivity, specificity, and processing power.
- Comparing the performance of the classification models developed using standard evaluation metrics.
- integrating telemedicine functionality to facilitate remote acquisition, transfer, and analysis of retinal fundus images for early diabetic retinopathy screening and diagnosis in remote and underserved communities.
- Developing a new, user-friendly Graphical User Interface (GUI) that allows health professionals to upload retinal images, receive automatic classification results, and see corresponding diagnostic information in real-time.

1.4. Project Organization

The project is divided into four main chapters: A general introduction, a literature review, materials and methods, and a conclusion. The first chapter (General Introduction) provides an overview of DR and the importance of its early detection and diagnosis. The second part expands on and discusses the most prevalent techniques used for DR diagnosis and segmentation. This includes imaging modalities, classifiers used, relevant features, and assessment techniques. The report contains a materials and methods chapter that highlights the features of the system being proposed. Finally, the topics are summarized with a conclusion. Citations, references, a list of figures and tables, along with the table of contents are included in the report to facilitate navigation and full understanding of the research.

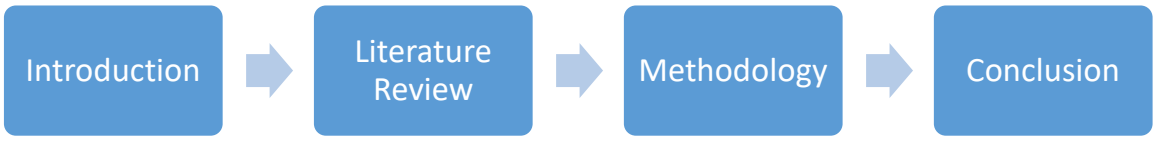

Figure 4-Report Organization

1.5. Timeline of Work Done

Table 1 Timeline of BIME-501 Spring 24/25

Ibrahim Alkhalil Abdulrahman Kadhim Kadhim		Spring Semester 2024/2025												
Project Task	Task Name	Week 4	Week 5	Week 6	Week 7	Week 8	Week 9	Week 10	Week 11	Week 12	Week 13	Week 14	Week 15	Week 16
T1	Define project scope and objectives													
T2	Collect relevant articles													
T3	Summarize article key points													
T4	Identify best practices and weak points													
T5	Determine design specifications													
T6	Consider design alternatives													
T7	Making prototype work plan													
T8	Write final report													
T9	Proofreading and revising													
T10	Prepare presentation slides													
T11	Present to the committee													

Chapter 2: Literature Review

2.1. Introduction

One of the major causes of retinopathy today is diabetes. Because diabetes has become a global problem, more and more people worldwide and in regional countries are at risk for visual problems and even blindness. Interestingly, damage to the retina often shows up years before diabetes is diagnosed. This can only be detected by a retinal screening, done at your diabetes clinic or by an ophthalmologist (eye specialist) or an optometrist who examines the retina.

2.2. Imaging Modalities Used for DR Diagnosis

Medical imaging is the technique of imaging the interior organs of the body, as well as the visual representation of the organs and tissues. Medical Imaging seeks to reveal the internal structures hidden by the skins and the bones. It refers to several different technologies to view the internal human body to diagnose, monitor and treat medical conditions.

While the primary method for evaluating DR requires ophthalmoscopy, there are various imaging modalities that are significant in the screening, evaluation and diagnosis. Many imaging techniques can be useful depending on the manifestation of DR [11]. Important imaging techniques to be familiar with including optical coherence tomography (OCT), fundus camera and digital slit lamp.

2.2.1. OCT and OCTA

Optical coherence tomography (OCT) and optical coherence tomography angiography (OCTA) are non-invasive imaging tests. They use light waves to take cross-section pictures of your retina. With OCT, your ophthalmologist can see each of the retina's distinctive layers and the optic nerve fiber layer. This allows your ophthalmologist to map and measure thickness and changes over time. These measurements help with diagnosis. They also guide treatment for glaucoma as well as retinal disease, like age-related macular degeneration and diabetic eye disease. Optical coherence tomography angiography (OCTA) takes pictures of the blood vessels in and under the retina. OCTA is like fluorescein angiography. But it is a much quicker test and does not use a dye [12].

Figure 5-OCT Machine

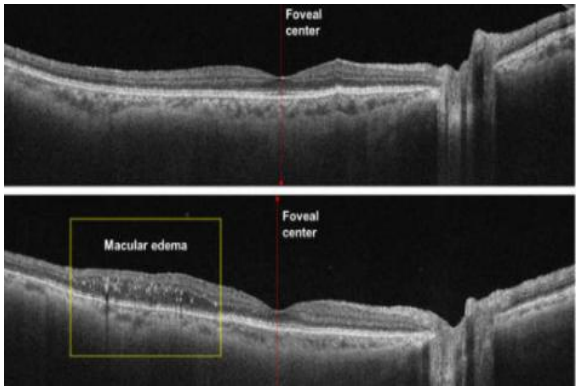

Figure 6-OCT images for normal and DR patients

The image above is an image of a 58-year-old woman without any chronic systemic diseases or eye diseases, wide field optical coherence tomography (OCT) of a horizontal scan through the center of the fovea (red line) revealing a normal retina. The image below Wide field OCT of a horizontal scan through the center of the fovea (red line) reveals marked thickening of the retina at the temporal quadrant of the retina (yellow box) [13].

2.2.2. Fundus Camera

Fundus photography is a simple, non-invasive medical test where an eye care specialist takes a picture of the fundus, the back inner wall of your eye. These photos let your eye specialist take a close look at several parts of your eye that are very important to your vision. Fundus photography can be part of a routine eye exam. Your eye care specialist can also use it for diagnostic eye exams when you have eye symptoms or vision changes. Diagnostic eye exams look for more specific signs of eye diseases or conditions [14].

Figure 7-Patient undergoing fundus photography [14]

Image A is color fundus image of the right eye for 58 years woman without any chronic systemic diseases or eye diseases. Image B is a color fundus image of the right eye of 45 years old man with signs of moderate non-proliferative DR. Features include microaneurysms (red arrows), hemorrhages (white arrows), hard exudates (HE, blue arrows) [13].

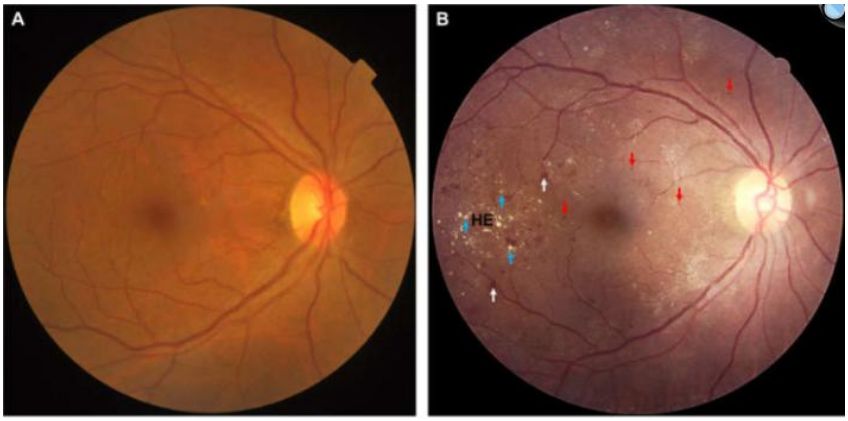

Figure 8-Fundus images for normal and abnormal patients [13]

2.2.3. Digital Slit lamp

A slit lamp exam is a test that lets an eye care specialist see every part of your eyes — including inside them. It's a common part of a routine eye exam. A slit lamp is a special microscope with a bright light attached to it that your eye care specialist will use to look at the different parts of your eyes. They'll adjust the light to see into and through the layers of your eyes. They'll check the overall health of your eyes and diagnose any issues or symptoms you're having. Your eye care specialist will probably dilate your pupils to perform a slit lamp exam. You can't drive with dilated pupils, so make sure to arrange transportation or have someone pick you up after your appointment.

Figure 9-Digital Slit Lamp in BME Lab at Beirut Arab University

2.3. Existing Pre-Processing Techniques

Before an image can be used for feature extraction, or as an input for a classifier, it must undergo the pre-processing phase, during which a variety of mathematical operations are applied to it. This is done to ensure that the image properties comply with the specifications of the classifier, and to optimize the performance of the algorithm. In medical imaging, specifically retinal imaging, pre-processing is crucial for reducing any variability in the training data introduced because of differences in acquisition conditions, illumination, noise and anatomical diversity.

2.3.1. Grayscale conversion

Retinal images are often captured in full color and processed in the RGB format, meaning each image is digitally described via 3 matrices, each representing the intensity of one color, blue, red and green. This introduces a significant amount of redundant data that is of no use in the classification process. Therefore, most algorithms start by converting the image into greyscale, as shown in Fig. 10 [15] .

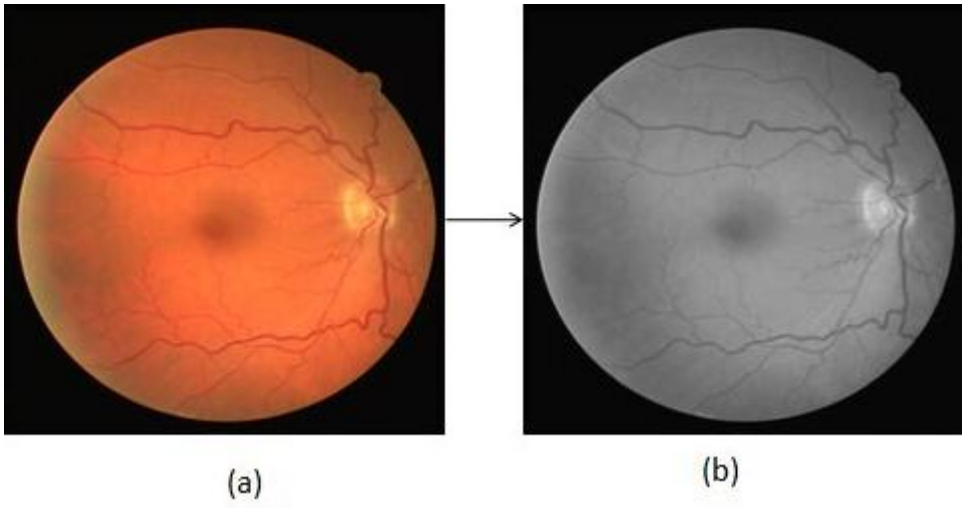

Figure 10-Greyscale conversion of RGB retinal image [15]

This means that the image is now described using a single matrix corresponding to color intensity, which is done using the following formula:

$$\boldsymbol{Grey = 0.299R + 0.587G + 0.114B} \qquad \textit{Equation 1}$$

Where:

- **Grey** is the greyscale pixel intensity value

- **R**, **G**, and **B** correspond to red, green and blue pixel intensity values respectively.

2.3.2. Resizing

DL models used for retinal image classification often have a specific image input size, meaning that all retinal images used in the dataset must comply with these dimensions before they can be used for training, testing or classification. This is done through the process of resizing which aims to change the dimensions of a retinal image while maintaining all data necessary for classification to work. There are many methods used to resize retinal images with varying results, such as:

A. Nearest Neighbor Interpolation

It is the simplest method for image resizing, works by assigning the intensity value of the nearest pixel in the original image to the pixel of the new image grid. Quick and reliable but can cause jagged or blocky results. Mainly used when speed is a critical part of the application.

B. Bilinear interpolation

$$\boldsymbol{I}(\boldsymbol{x}', \boldsymbol{y}') = \sum_{(i=0)}^{1} \sum_{(j=0)}^{1} \boldsymbol{w}_{ij} \cdot \boldsymbol{I}(\boldsymbol{x}_i, \boldsymbol{y}_j) \qquad \textit{Equation 2}$$

Where:

- (x′,y′) is the position in the resized image.
- (x,y) are the corresponding non-integer coordinates in the original image.
- Wij are weights based on the distances from the four surrounding pixels

This method performs linear interpolation along 2 axes, making for a smooth result with decent quality for most applications.

C. Bicubic interpolation

$$I'(x', y') = \sum_{i=-1}^{2} \sum_{j=-1}^{2} w_{ij}.I(x+i, y+i)$$ Equation *3*

Where:

- I(x,y): Interpolated pixel value
- I(x+i,y+j): Intensity of surrounding pixels
- wij: Weight computed using a bicubic kernel function, typically based on distance

This method estimates intensity value using the weighted average of 16 neighbouring pixels (4x4 grid), resulting in smoother gradients and higher quality image scaling.

2.3.3. Contrast Enhancement

Any of a variety of techniques meant enhance the difference between a region of interest and the background, making for a smoother feature extraction process. In retinal image processing, one of the most used contrast enhancements is CLAHE or Contrast Limited Adaptive Histogram Equalization as shown in Fig. 11 [16].

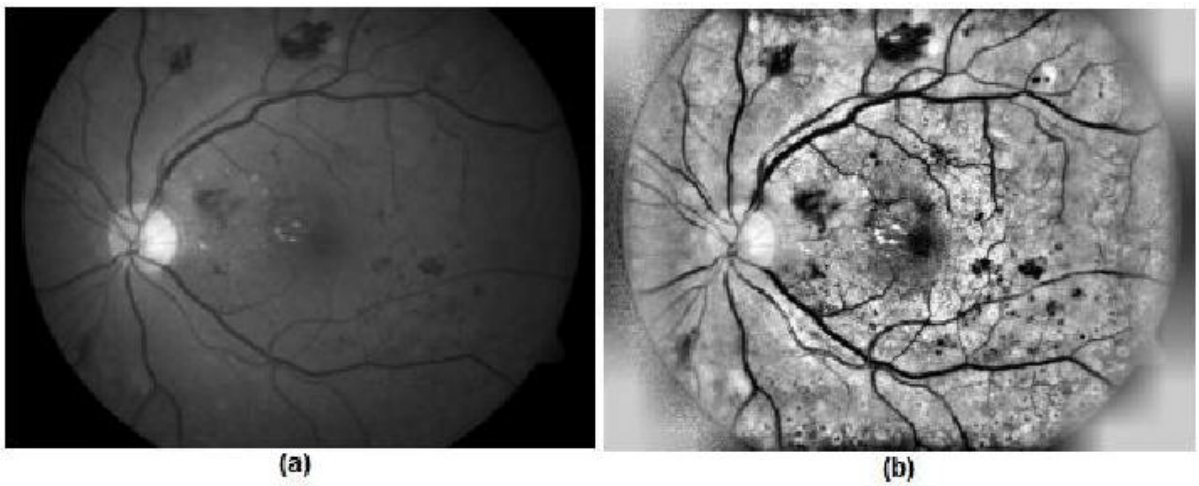

Figure 11-Retinal image before and after applying CLAHE [16]

It is a variant of histogram equalization that operates on small tiles of pixels rather than the whole image. It starts by dividing the image into tiles then computing the histogram for each tile. After that the histogram is clipped at a pre-defined limit which prevents over-amplification, and the excess pixels are redistributed uniformly across the histogram. Finally, each tile is equalized using the modified histogram, and neighboring ties are interpolated to avoid block artefacts from appearing. Its main formula is as follows:

$$I_{eq}(x,y) = \left(\frac{C(I(x,y)) - C_{min}}{M \times N - C_{min}}\right) \times (L-1) \qquad \textit{Equation 4}$$

Where:

I(x,y): Original pixel intensity at location (x,y) in the image tile.

I_eq(x,y): Enhanced (equalized) pixel intensity after CLAHE processing.

C(i): Cumulative Distribution Function (CDF) of the histogram up to intensity level i.

C(I(x,y)): CDF value evaluated at the original pixel intensity I(x,y).

C_min: Minimum non-zero value in the CDF, used for normalization.

M × N: Tile dimensions (number of pixels in the local processing region).

L: Number of possible intensity levels (e.g., 256 for 8-bit grayscale images).

L − 1: Maximum intensity value after equalization (e.g., 255 for 8-bit images).

2.3.4. Noise Reduction

In the context of medical imaging, noise is pixel variance that impedes on the description of the structure being imaged. Extracting features from noisy images leads to a great deal of classification disparity between them and non-noisy ones and does not represent the properties of the structure adequately. Therefore, noise reduction (Fig. 12 [17]) is a crucial step in any image classification workflow and can be done using a multitude of techniques.

Figure 12-Fringe noise reduction in retinal fundus image [17]

A. Gaussian Blur

Gaussian Blur is a type of smoothing filter used to remove high frequency components from images while maintaining soft edges. It works by convolving the given image with a Gaussian kernel, that gives greater weight to pixels near the center. The following formula describes weight distribution in a Gaussian kernel.

$$G(x,y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \qquad \textit{Equation 5}$$

Where:

- G(x,y) is the value of kernel pixel weight
- x, and y are the horizontal and vertical distances from kernel center respectively
- σ is the standard deviation of the Gaussian function, where a greater σ results in a more intense blur.

B. Median Filtering

It is a non-linear noise reduction technique which is particularly useful for impulse noise, which refers to isolated bright or dark pixels. By replacing each pixel with the median intensity of its neighborhood, the technique disposes of high frequency components while preserving sharp edges, making it useful in cases where minimal blur is tolerated.

C. Bilateral Filtering

In bilateral filtering, the spatial proximity and intensity similarity of pixels is used to calculate the filtered pixel value, which results in noise reduction while preserving the image edges. It can be described using the following formula:

$$I'(x,y) = \frac{1}{W_P} \sum_{I,J} I(i,j).f_r(|\,\mathrm{I(i,j)} - \mathrm{I(x,y)}\,|).f_s(\sqrt{(i-x)^2 + (j-y)^2}) \qquad \textit{Equation 6}$$

Where:

- I’(x,y) is the filtered pixel value at position x,y
- Wp is a normalization factor
- fr is the range kernel responsible for intensity similarity
- fs is the spatial kernel responsible to spatial proximity

2.3.5. Green Chanel Extraction

As previously mentioned, digital retinal images contain 3 channels or matrices: red, green and blue. Since green light penetrates only moderately into the retina, it illuminates surface and near-surface features most effectively, providing contrast between blood vessels and surrounding tissue, and highlighting retinal abnormalities. Thus, many retinal image classification and segmentation algorithm extract the green channel from the image (Fig. 13 [18]) and then use that as the basis for grayscale conversion.

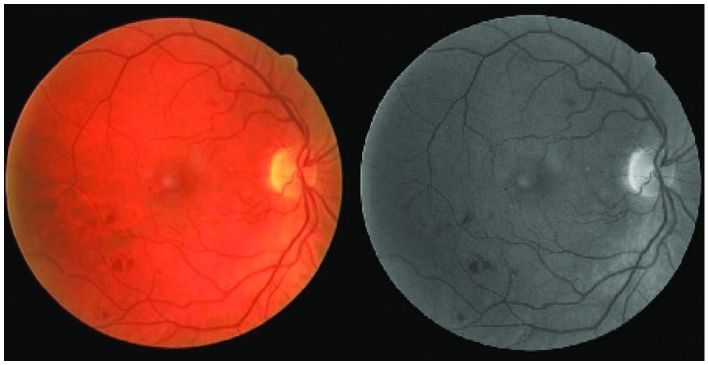

Figure 13-Retinal image and its green channel [18]

2.4. Existing Features Extraction Techniques

Feature extraction is one of the important tasks that increases the efficiency of the whole system. The feature explains some computable property of the given input image. In image analysis, feature extraction helps reduce the dimensionality of image data while preserving significant structural, textural, or statistical properties. It can be classified into many different types which are: color, shape, texture and frequency-based features. It is an important tool before classifying DR because it helps in taking many details from different types of retina images.

2.4.1. Fourier Transform

Fourier transform is mathematical technique that converts a signal from the spatial domain to frequency domain. In Images, it allows analysis of the images based on the frequencies such as (edges, repetitive structures). The Fourier transform is a representation of an image as a sum of complex exponentials of varying magnitudes, frequencies, and phases. It plays a critical role in a broad range of image processing applications, including enhancement, analysis, restoration, and compression. For a 2D grayscale image the Fourier transform is given by:

$$F(u,v) = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x,y) e^{-j2\pi(\frac{ux}{M}+\frac{vy}{N})} \qquad \textit{Equation 7}$$

Where:

- (x, y) is the intensity value at pixel position (x, y)
- $f(u, v)$ is the Fourier coefficient at frequency (u, v)
- M and N are the image dimensions
- j is the imaginary unit.

FT in DR classification

Low frequency components (center of spectrum) which represent smooth and slow areas in the image like the image background. And high frequency components (edges of the spectrum) represent sudden changes like edges, textures or fine details. So, in retinal images, the vessel structures and lesion textures introduce high frequency components, healthy background retina introduce low frequency components. So, by analyzing the frequency components of the retina image can help to take more details about blood vessels or lesions.

2.4.2. Principal Component analysis (PCA)

PCA is a dimensionality reduction and unsupervised machine learning method used to simplify large dataset into a smaller set while preserving as much information as possible. Reducing the number of variables of a data set naturally comes at the expense of accuracy, but the trick in dimensionality reduction is to trade a little accuracy for simplicity. Because smaller data sets are easier to explore and visualize and thus make analyzing data points much easier and faster for machine learning algorithms without extraneous variables to process. So, the steps of doing PCA are:

1. Data Standardization

Affected by feature scales, so mean centering and scaling are required

$$X = \frac{X - \mu}{\sigma} \qquad \textit{Equation 8}$$

Each feature is transformed to have zero mean value and unit variance.

2. Compute the covariance matrix to identify correlations

The covariance matrix captures how the features vary together.

$$C = \frac{1}{n-1} X^T X \qquad \textit{Equation 9}$$

Where:

- X is the standardized data matrix
- X^T is the transpose of matrix X
- C is the covariance matrix.

3. Compute Eigen Vectors and Values to identify principal components

Eigenvectors and eigenvalues are the linear algebra concepts that we need to compute from the covariance matrix in order to determine the *principal components* of the data.

Eigen values and eigen vectors come in pairs, every eigen vector has an eigen value, their numbers are equal to the data dimension.

$$\boldsymbol{C.v = L.v} \qquad \textit{Equation 10}$$

Where:

- v is the eigen vector
- L is the eigen value

4. Transforming the data

Project the original data into a new subspace

$$Z = X * W \qquad \textit{Equation 11}$$

Where:

- X is the original data matrix
- Z is the new transformed feature set with reduced dimensions

2.4.3. Haralick Features

Haralick features are second-order statistical texture features derived from the grey level co-occurrence matrix, a matrix that counts how often pairs of pixel values with a specific spatial relationship occur in an image. Haralick texture features are used to describe the "texture" of an image. If you are trying to quantify and represent the feel, appearance, or consistency of a surface, then Haralick texture features are a good starting point. Haralick texture features are computed using the Gray-Level Co-occurrence Matrix (GLCM). This matrix characterizes texture by recording how often pair of adjacent pixels with specific values occur in the image.

This is a grayscale image matrix to understand the way GLCM works.

Figure 14-Grey level matrix [19]

Then to construct the GLCM, look at pairs of adjacent pixels and count the number of times these two values appear next to each other.

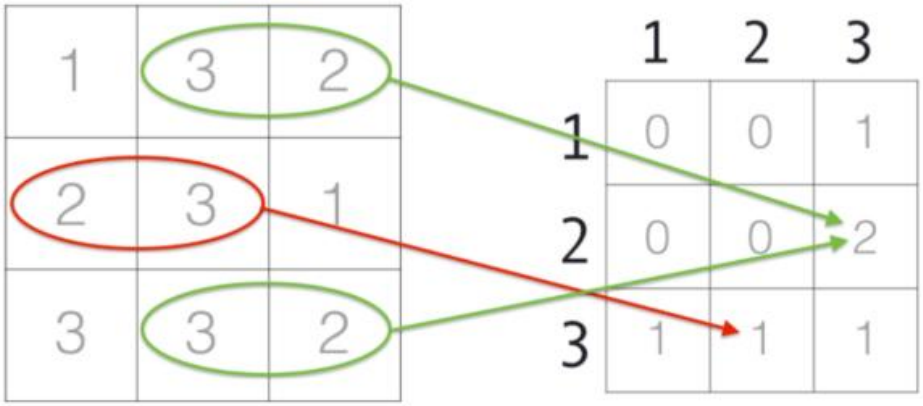

Figure 15-GLCM [19]

The pixel intensities 2 and 3 appear next to each other once, so they have an entry of one in GLCM. The pixel intensities 3 and 2 appear next to each other twice, so they have an entry of two in GLCM. It is not just limited to for recording the number of times the pixel intensities appear next to each other, also specifying four different directions of adjacency: left to right, top to bottom, top left to bottom right, top right to bottom left. After having four GLCMs Haralick features can be computed for each of the GLCMs. These values are simply statistics computed from the GLCM used to characterize and represent contrast, correlation, dissimilarity, entropy, homogeneity, and other desirable statistical properties [19].

2.5. Artificial Intelligence in Diabetic Retinopathy Detection

AI could be defined as an extremely broad category of computer science and engineering which aims to create intelligent software for the purpose of optimizing various processes. This is done through the design of systems and programs that simulate the process of human intelligence, effectively allowing them to think, learn and adapt autonomously with little to no human intervention [20]. This in turn makes it a more efficient and reliable option than human beings in a variety of tasks ranging from providing virtual assistance on a webpage to controlling the pumps and valves of the most intricate manufacturing facility.

The phrase "Artificial Intelligence" was coined in 1956 at the Dartmouth Conference, which marked the origin of AI as a separate field. Symbolic problem-solving and reasoning were the focus of early AI research. The pace of advance was derailed through so-called "AI winters" due to limiting constraints on computer resources and exaggerated hopes. But since the 21st century, it has sustained increasing popularity, simply because of leaps in computational strength, availability of extensive datasets, and improvements in algorithm and network architecture design. AI is now at the forefront of innovation, leading change across fields as varied as telecommunications, entertainment and finance. Its explosive growth produces not only technical challenges but serious ethical and social ones too [21].

The use of AI has become a growing practice in the ever-evolving medical field. As medical professionals begin to lose doubt in the reliability of AI, several companies have begun its integration into their software and machinery, and the bulk of research in medical technology seems to demonstrate its potential time and time again. For most parts, AI has 2 uses in the medical setting: prognosis and diagnosis. The former refers to the employment of an AI model for the prediction (regression) of the behavior of certain quantities such as predicting the hours

left before an insulin shot is needed or the days left for a terminally ill patient. Diagnosis is the ability of AI to deduce the type of condition a patient suffers from through analyzing various types of medical data such as symptoms, history, medical images, and using that to place the patient into the category they most accurately fit in (classification). This application has been of great interest to medical technology researchers in recent years, mainly due to its potential for providing more accurate diagnosis than the medical doctor for a fraction of the cost [22].

Therefore, the use of AI in DR detection should come as no surprise. Studies in recent years have employed various algorithms on images acquired through the previously discussed eye imaging modalities for the diagnosis and classification of the disease. These algorithms can be broadly split into two groups: Machine Learning (ML) and Deep Learning (DL). Notably, DL algorithms are a subset of ML algorithms used for more complex tasks. Thus, by labelling an algorithm as ML, one is referring to the set of ML algorithms complementary to those used in DL.

2.5.1. Existing Machine Learning Algorithms

ML is a broad branch of AI which allows systems to recognize patterns in data and thus make decisions without the use of any direct programming. This data-driven approach has proved to be very useful in applications where rule-based systems and mathematical models would be inefficient or impossible to create, such as image identification, speech processing, and anomaly detection.

Machine learning techniques are typically classified into three broad categories: supervised learning, unsupervised learning, and reinforcement learning. Supervised learning algorithms are trained using labelled data and learn to map inputs to recognized outputs. Unsupervised learning is the discovery of hidden patterns or structures in unlabeled data, and reinforcement learning allows the algorithm to learn and adapt its actions by experimenting with the environment in trial-and-error mode and receiving feedback in the form of reward or penalty. Algorithm selection and learning strategy depend on the problem type and available data.

2.5.1.1. Support Vector Machines

Support Vector Machines (SVMs) supervise learning algorithms and are commonly used for classification and regression. In their design, SVMs are committed to finding the optimal decision boundary—or hyperplane—about which data points of different classes are being separated with the largest conceivable margin. This makes them optimal for tasks that have to do with medical diagnosis. Data points lying closest to this boundary are called support vectors, and they play a crucial role in defining the position and orientation of the hyperplane. In linearly separable situations, the hyperplane in feature space optimally separates the classes while also being at a maximum distance from the nearest points of a class as seen in Fig. 16 [23]. When data is not linearly separable, SVMs apply a method called the kernel trick where data is mapped to higher space where a hyperplane which will be able to separate is calculated. Linear kernels, polynomial kernels, and radial basis function (RBF) kernels are the ones most used. Having such a capacity to manage linear as well as nonlinear relationships makes SVMs an efficient and all-around useful resource across a variety of fields such as image identification, bioinformatics, and text classification.

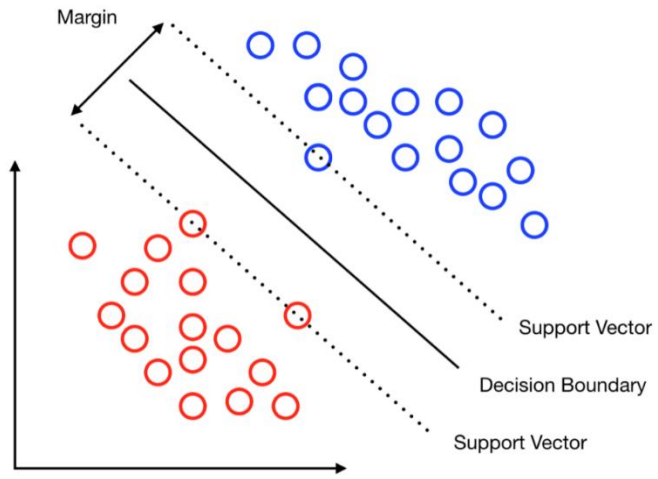

Figure 16-Example of hyperplane separating between 2 classes [23]

2.5.1.2. Random Forest

A random tree is a type of decision tree-based algorithm that aims to reduce overfitting by incorporating randomness into the tree construction process. Traditional decision tree algorithms create trees by selecting the best split at each node based on the entire dataset. Random trees use a random subset of features instead which helps overcome overfitting in highly specific datasets. The randomness of the tree construction process allows for the creation of several algorithms which can be used in an ensemble in classifiers such as random forest. Random Forest is a popular supervised machine learning algorithm that operates by constructing an ensemble of random decision trees (Fig. 17 [24]) during the training stage and then outputting the mode of the classes in classification or the mean prediction in regression of the individual trees. It belongs to a class of algorithms known as ensemble methods, which attempt to make better predictions by combining the output of several learners.

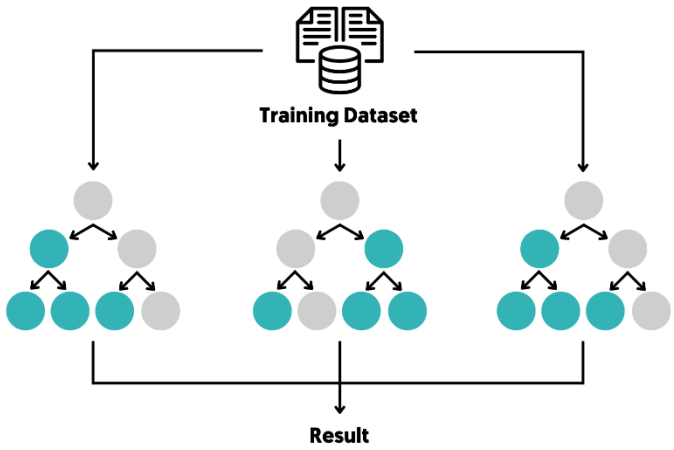

Figure 17-Random Forest algorithm diagram [24]

The idea of RF is to reduce the overfitting and large variance usually experienced by regular decision trees. When training, for every tree a random subset of the training set is considered (with bootstrapping) along with a random subset of features while splitting each node. Randomization introduces heterogeneity among the trees and provides a stronger model that is more immune to noise. Due to its precision, interpretability, and ability to handle high-dimensional data with minimal pre-processing, RF finds a wide range of applications.

2.5.1.3. K-Nearest-Neighbor

K-Nearest-Neighbor (KNN) algorithms are a subset of simple unsupervised ML algorithms mainly used for the classification of data into one of several defined classes. This is done through the calculation of the distance between a data point and previous training points then assigning the class label of its nearest feature space neighbours to it. The value of K (the number of neighbours considered) is an essential hyper-parameter that heavily influences the model's performance. Fig. 18 [25] showcases an example of a sample input being classified into 1 of 2 groups using KNN. Although KNNs are great for small datasets with limited dimensions, they can become computationally expensive for larger datasets, as calculation has to be done to classify each point, and the results of the classification process have to be stored. Moreover, the performance of the model can degrade in high-dimensional spaces in a phenomenon known as "the curse of dimensionality".

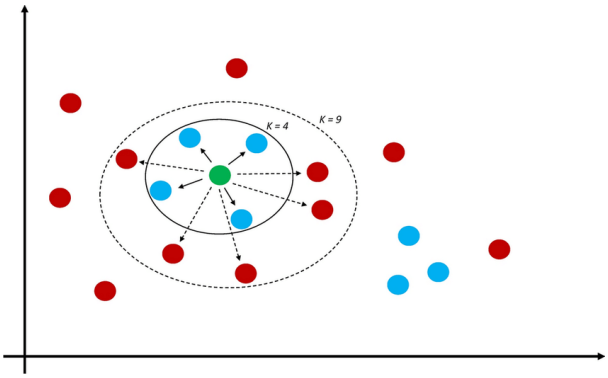

Figure 18-KNN algorithm diagram [25]

2.5.1.4. Probabilistic Neural Networks

A probabilistic neural network (PNN) is a type of feed forward neural network, that's mainly used for classification tasks. They are largely based on Bayesian decision theory and use kernel-based estimation to calculate the probability distribution function (PDF) of each class. After that, the PNN calculates the probability that the input belongs to each class and assigns it to the one with the highest probability. This process allows for the classification to be accurate, especially when the data used to train the model is well distributed. PNNs usually consist of four layers: input, pattern, summation and output (Fig. 19 [26]. The pattern layer contains one neuron for each training sample and calculated a similarity measure between the input and each sample. The summation layer then aggregates results by class and the output layer chooses the class with the highest likelihood. PNNs have the advantage of a very swift training process since they don't require any iterative weight updates. However, they can be

computationally intensive with large datasets, since every training sample essentially becomes a node in the network.

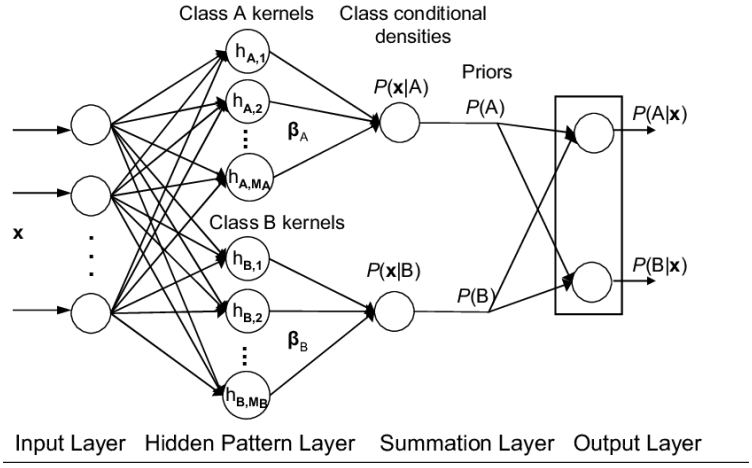

Figure 19-Probabilistic Neural Network Architecture [26]

2.5.2. Existing Deep Learning Algorithms

DL is one of the subsets of ML, which mainly deals with algorithms imitating the process of the human brain and nervous system called neural networks. Deep learning models consist of many interconnected levels with each having nodes known as neurons where the data is processed. Each level translates the data into a more abstract form so that the system can comprehend and adapt based on complex forms and features. DL has enabled us to make breakthroughs in those areas where mathematical modelling and traditional ML algorithms fail such as natural language understanding, computer vision and generative work. This is because DL can automatically extract relevant features from a sample without the need for mathematically complex and highly specialized features. That is at a price, however, since DL algorithms require high computational expense and large datasets to operate.

2.5.2.1. Convolutional Neural Network

Convolutional Neural Networks or CNNs are a class of deep networks that are specially designed to process data with grid-like topology, such as images. In contrast to the regular fully connected neural networks, CNNs use a mathematical technique called convolution in order to automatically extract spatial features from data inputs. This operation is done by applying filters or kernels that go over the pixels of the input image and detect local patterns such as edges, textures, and shapes. The result is a feature map that is an accurate representation of the spatial relations in an image.

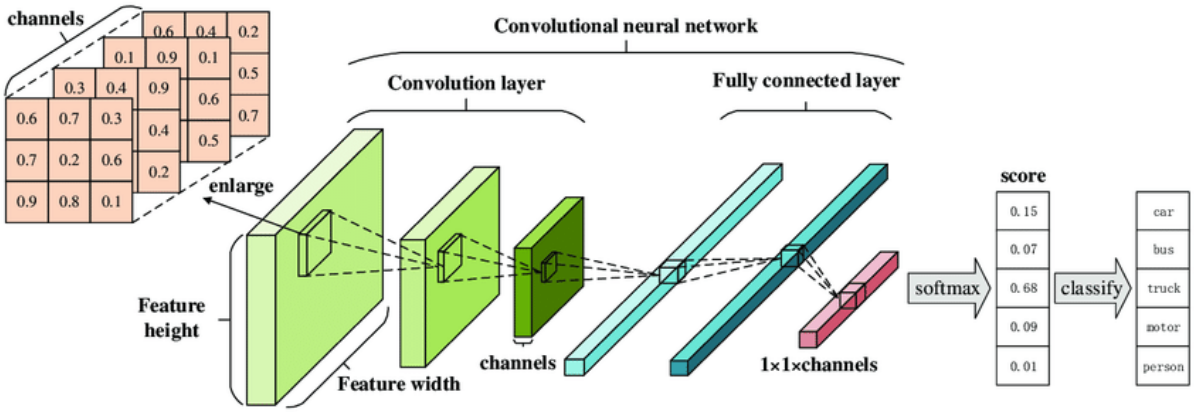

Figure 20-Typical CNN Architecture [27]

The basic CNN structure is composed of 3 layers of different kinds: convolutional layers for feature extraction, pooling layers to reduce the dimensions of the samples, and fully connected layers for carrying out the end regression or classification of the input image (Fig. 20 [27]). The ability of a CNN to learn a hierarchical pattern starting from low-level edges to high-level complex shapes renders them highly beneficial for use in computer vision applications.

2.5.2.2. Recurrent Neural Networks

Recurrent Neural Networks (RNNs) are a specific neural network which is utilized in processing sequential information, such as time series data, text, or speech. Unlike feedforward networks, RNNs do have a loop of feedback to allow information to persist from time step to time step as shown in Fig. 21 [28], so they are most suited for instances where context and order are of importance. With every step, an RNN processes an input and passes it with its previous hidden state to produce a new hidden state, therefore effectively possessing some kind of memory.

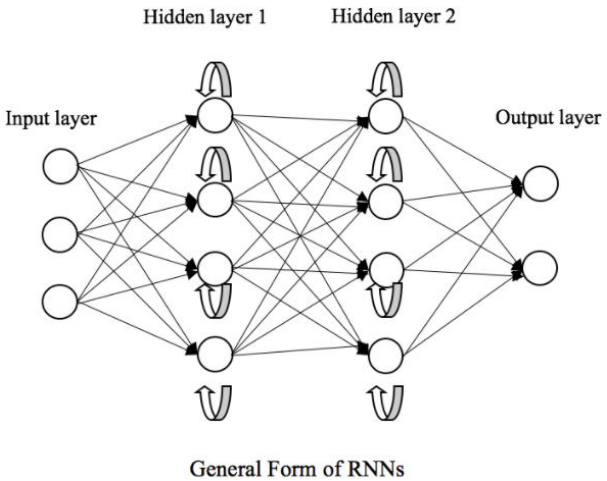

Figure 21-RNN architecture [28]

Although they are theoretically straightforward, typical RNNs are plagued by the inability to learn long-term dependencies due to vanishing and exploding gradients during training. In a bid to solve this issue, even more sophisticated architectures such as Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) networks were developed. The networks have structures that allow them to selectively forget or remember with time, which help overcome the constraints of previous models. As such they perform better in tasks such as language modelling, machine translation, and sequential signal classification.

2.5.2.3. Deep Belief Networks

A Deep Belief Network is a generative graphical model containing several layers of stochastic, latent variables. Networks are constructed using the technique of stacking several layers of Restricted Boltzmann Machines (RBMs) on one another as illustrated in Fig. 22 [29]. The RBM trains to capture the features of a layer below it and sends the output to serve as the next layer's input. The primary goal of DBN is to learn hierarchical representations of the data through unsupervised pretraining to initialize the network correctly before fine-tuning for supervised learning.

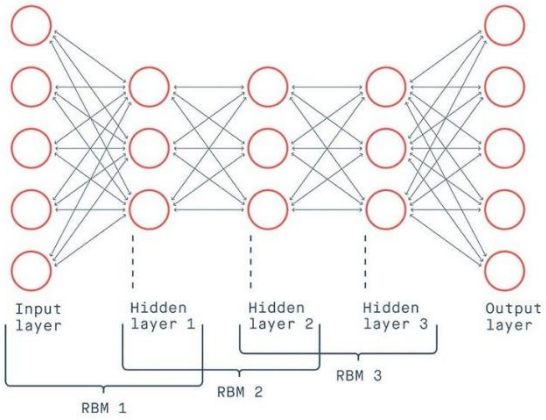

Figure 22-DBM architecture [29]

There are typically two phases involved in training DBN. The first is referred to as greedy layer-wise pretraining and it is where unsupervised learning is used to train individual RBMs so they can accurately model the distribution of their input. This trains DBN to progressively detect more abstract features at deeper layers. In the second phase, the entire network is refined using backpropagation and labelled data such that it can perform classification tasks. DBNs were among the initial deep learning models and helped to prove the power of deep architecture, especially in uses such as image recognition and speech processing.

2.6. State of the Art Detection Algorithms

In 2025, Zaylaa et al. [30] used the MATLAB software to evaluate the performance of 8 Deep Neural Networks (DNNs) of various depths and sizes in the classification and diagnosis of DR. The study also provided a framework for the integration of AI models into the software of the digital slit lamp shown in Fig. 1, paving way for the use of autonomous decision making in optometry. The paper used a publicly accessible dataset comprised of retinal images captured using fundus photography at the Aravind Eye Hospital in India. After the pre-processing of data which included image resizing and equalization, a conventional 80/20 split was used to train and test each model. The networks were then evaluated based on accuracy, sensitivity, specificity, precision, and Receiver Operator Characteristic (ROC) curve. The results demonstrated that GoogleNet exhibited the best sensitivity (99.44%), but ResNet-50 was the best in overall DR detection due to scoring 1.38% and 1.74% higher than GoogleNet in specificity and precision respectively.

In 2020, Gayarthi et al. [31] presented an automatic classification system for DR using a combination of Haralick and Anisotropic Dual-Tree Complex Wavelet Transform (ADTCWT) features. The study compared the performance of several classifiers such as Support Vector Machine (SVM), Random Forest, Random Tree, and J48 in binary and multiclasss classification using the aforementioned features. Image datasets were obtained from MESSIDOR, KAGGLE, and DIARETDB80. Haralick features were used to infer information about image texture, whereas ADTCWT reliably extracted directional information from the fundus images used, making for a novel hybrid feature extraction approach in DR classification. The results demonstrate that the Random Forest algorithm has the best accuracy among the tested classifiers scoring 99.7% for binary and 99.82% for multiclass classification. Despite the promising results, the algorithms used may be computationally intensive, and thus the possibility of clinical implementation remains undiscovered.

Ratanapakorn et al. [32] worked on developing a MATLAB Graphical User Interface (GUI) that classified fundus images into normal, NPDR and PDR. The software used image processing techniques such as image segmentation and thresholding to extract abnormalities in retinal images ranging from microaneurysms (MAs) to Neurovascularization (NV), and then classify the images based on the severity of the DR pathologies extracted. The GUI, shown in Fig. 23, provided a clear and concise method of viewing DR diagnosis and highlighted regions of interest in the fundus images. Moreover, the software was highly accurate in DR detection (96.25%) and exhibited high sensitivity (98%). Despite such promising results, the study was limited in that DR patients may have lesions like DR pathology which contribute to classification error. Furthermore, the study used a dataset comprised of 95% of DR images which may have led to inflated accuracy results, and the software had low accuracy in distinguishing NDPR from PDR (66.58%).

Figure 23-GUI developed in the study [32]

Suedumrong et al.'s (2024) paper [33] explains various deep learning approaches to detecting diabetic retinopathy (DR) with a focus on the performance of convolutional neural networks (CNNs). CNN architectures such as Inception V3, VGGNet, and ResNet have been widely employed in the literature to classify retinal images, primarily from EyePACS and Messidor datasets, with good classification accuracy but limiting DR severity largely to binary levels (No DR vs. DR) rather than the five-class classification needed for early detection. This Research highlights that DR detection at an early stage remains difficult since microaneurysms and tiny hemorrhages are inconspicuous and are disregarded by models trained on RGB images

alone. To enhance diagnostic accuracy, preprocessing techniques such as contrast enhancement, data augmentation, and background removal have been proposed, with background removal being highly effective in highlighting key retinal features. The paper further contrasts a number of CNN-based approaches, with the note that while state-of-the-art studies report competitive accuracy, the compromise between model performance and computational efficiency remains a core concern. The study contributes to the literature by demonstrating that the integration of advanced preprocessing techniques significantly improves the performance of CNNs in DR detection with validation accuracy of 90.60%.

Senapati, A., Tripathy, H. K., Sharma, V., & Gandomi, A. H. (2024) [34] provides a critical analysis of the development of AI-based detection of diabetic retinopathy (DR), more specifically ML and DL algorithms. It provides the necessity of DR screening automatization due to the increasing population of diabetes patients and the limitation of manual assessment by ophthalmologists. Different Machine learning methods such as SVM, KNN, and decision trees have been used, and Deep learning methods like CNNs, ResNet, VGG-16, and transfer learning methodology have proven promising. Recent literature, issues such as class imbalance, computational efficiency, and availability of the datasets also have been founded, alongside the need for more scalable, efficient, and cost-effective AI-driven solutions. It ensures improving feature extraction, classification, and real time Diabetic retinopathy detection for early detection and diagnosis and better patient outcomes.

The Kim, Mishra, and Sen (2022) [35] articulates teleophthalmology's role in enhancing eye care service delivery, particularly to distant and underprivileged populations. Teleophthalmology utilizes information and communication technology (ICT) to reduce the gap between patients and healthcare providers, reducing the need for transport and increasing access. AECS has successfully established teleophthalmology at the primary and secondary levels of healthcare, sharing with vision centers doing over 2800 teleconsultations each day. According to research studies around 15–17% of the patients in AECS vision centers require referral to tertiary centers. New technologies, including AI based softwares have increased the screening process of DR, have offered early diagnosis and treatment for patients. However, the challenges such as image quality issues, high initial investments and the security of the data concerns remain significant challenges to the widespread acceptance.

Pal et al. [36] presented a novel retinal vessel segmentation technique that aimed to enhance segmentation accuracy using morphological operations combined with the iterative rotation of structuring elements. The method was meant to construct clear outlines of retinal vessels from fundus images to aid optometrists in the diagnosis process. A 2d wavelet transform paired with Contrast Limited Adaptive Histogram Equalization (CLAHE) was used for the pre-processing of images to enhance edge contrast before segmentation. After that, the authors used a gray-level hit-or-miss transform with multi-structuring element to detect vessels of varrying widths across 12 different angles (0 to 165 in 15 steps). Finally, hysteresis thresholding and morphological operations were used to eliminate noise and remove unwanted areas from the image. The methodology was validated using the DRIVE dataset which included retinal images paired with their ground truth vessel map, which highlighted all the pixels corresponding to blood vessels. The algorithm achieved a maximum accuracy of 95.65%, and an average accuracy of 94.31%, making it a reliable choice for vessel segmentation. However, the work is limited using a predefined structuring element, which may lead to inaccurate results if used with fundus images taken under diverse conditions. Although the method was meant to aid medical specialists in making observations for the diagnosis of DR and other retinal conditions, it could also be used for the pre-processing of images used in automatic classification techniques.

The paper by Li et. al [37] describes a reinforcement local description-based approach for retinal blood vessel segmentation. The novel algorithm addresses problems such as intensity variations and vessel discontinuities. A line set-based feature extraction method was used which included factors such as local shape, intensity and morphological gradient features. A svm classifier was trained on the aforementioned features to classify image pixels into blood vessels and background, and an additional morphological post-processing step was used to reconnect blood vessels. The algorithm was tested on both the DRIVE and STARE datasets and achieved a remarkable accuracy of 96.3% and 97.1% respectively. However, like the previous paper, the method relies on highly specific features which may not generalize well with different images. Moreover, although post-processing improves vessel connectivity, it may produce artefacts in noisy images.

The review of Akhtar et al. [38] (2025) presents some deep learning methods for detecting diabetic retinopathy (DR), emphasizing their classification accuracy and limitations. Previous work has explored CNN-based models, transfer learning, and hybrid models, with test accuracy ranging from 86.08% to 98.48%, depending on the complexity of the dataset and preprocessing techniques. Methods utilizing ResNet, DenseNet, and ensemble learning techniques have improved DR classification but are often confronted with class imbalance, overfitting, and lower specificity. In the relief of these challenges, Akhtar et al. (2025) proposed the RSG-Net model containing convolutional layers, max pooling, batch normalization, and dropout for stability and accuracy. Preprocessing techniques such as Histogram Equalization and Gaussian Blur were applied to improve image quality, and data augmentation techniques corrected class imbalance. For the Messidor-1 dataset, RSG-Net reported 99.36% four-stage classification and 99.37% binary classification accuracy, outperforming other state-of-the-art models. These results make RSG-Net a highly precise and efficient computerized DR detection tool with the potential to be used in clinical settings.

Tabacaru et al. (2024) presents an effective machine learning model for DR classification, highlighting image preprocessing, feature extraction, and classification strategies. The work employs fundus eye images with preprocessing techniques enhancing contrast and extracting most important features before classification by machine learning techniques. The proposed model is 92.9% accurate, with an F1-score of 90.2% and an AUC of 94.1%, and this is substantiated using a bootstrap statistical approach. The novelty of the research lies in its systematic approach to feature manipulation and categorization, enabling the early detection of DR. The research contributes to the construction of AI-driven diagnostic tools for telemedicine and automatic ophthalmic screening, strengthening the role of machine learning in medical image analysis [39].

Qian et al. (2022) investigate the efficacy of an AI-powered computer-assisted DR grading and training program for assisting medical students in the learning of manual DR detection. The study employed a dataset of real-world Chinese diabetic fundus images and implemented deep learning classifiers. The AI system demonstrated excellent accuracy (96.5%), sensitivity (96.5%), and specificity (96.6%) in case detection of moderate or worse DR. When incorporated into medical education, the system boosted the diagnostic efficiency of junior residents and medical students, increasing their ability to identify DR from 91.5% to 95.4%. This study shows the potential for AI in enhancing early DR identification capacity for the non-

healthcare specialists and in ophthalmic education. This research aims to increase AI in the medical schools and enhance it in the decision making in the field of diabetic retinopathy (DR) classification and early detection [40].

In 2018, Gharaibeh et al. [41] proposed an image processing framework which can be used to enhance the detection efficiency of certain DR indicators such as microaneurysms, hemorrhages, and exudates. The pre-processing stage involved many operations such as HSI color space conversion, non-linear Wiener filtering to reduce noise, and CLAHE to enhace contrast. Blood vessel segmentation is then performed using a Constrained Possibilistic Fuzzy C-Means algorithm (SCPFCM), and feature extraction and selection were done via a DBN. Classification was then performed using a hybrid SVM-GA algorithm which combined SVM with genetic algorithm optimization. After using the DIARETDB1 retinal image dataset for evaluation, the results showed a higher sensitivity (99%), specificity (96%), and accuracy (98.4%) compared to traditional classifiers like PNN and standard SVM, which showcased the efficacy of the proposed algorithm in DR diagnosis.

Mutawa et al. [42] presented a novel DL algorithm for DR diagnosis, which combined CNNs with an RBF classifier. The algorithm utilized up-to-date pre-processing methods such as green-channel extraction, CLAHE, morphological operations, and Otsu's thresholding for accurate segmentation of vessels. MS-DRLBP was utilized for feature extraction in this study. After using publicly available data sets for testing (STARE, HRF, and FFA), the proposed CNN-RBF hybrid model achieved a top accuracy of 97.3% on STARE and outperformed traditional classifiers like Naïve Bayes, SVM, and ANFIS. The study demonstrated the efficiency of using randomization-inspired learning in medical image classification, and once again showcased the broad range of applications of hybrid models in medical diagnosis.

Table 2 Literature Review Summary Table

Year	Author	Methodology	Datasets	Classifier	Accuracy
2025	Zaylaa et al.	Evaluated the performance of 8 DNNs of various depths for DR detection. Using Pre-processing techniques like resizing and histogram equalization.	APTOS 2019	Google-Net, Res-Net-50, Other DNNs	Google-Net sensitivity 99.4%, Res-Net: higher specificity (+1.38%) and precision (+1.74%)
2025	Akhtar et al.	RSG-Net CNN with batch normalization, max pooling. Used histogram equalization, Gaussian Blur, and data augmentation.	Messidor-1	RSG-Net (CNN based)	99.36% for 4 class classification , 99.37% binary classification.
2025	Pal et al.	Developed morphological segmentation with iterative rotation of structuring elements. 2D wavelet transform and CLAHE.	DRIVE	Morphological operators, thresholding	94.31%
2024	Suedumrong et al.	Analyse CNN DR detection (Inception V3, VGGNet, ResNet), Preprocessing techniques involved using contrast enhancement, data augmentation, and background removal to highlight features.	DRIVE	InceptionV3, VGG Net, ResNet	90.60%
2024	Tabacaru et al.	Used ML pipeline, using contrast enhancement, feature extraction, and bootstrap statistical validation.	Fundus images	ML classifier	92.9% accuracy, F1:90.2%, AUC: 94.1%
2024	Li et al.	Introduced reinforcement local descriptor approach using line-set features (shape, intensity, morphological gradients). Used SVM for classification with morphological	DRIVE, STARE.	SVM and morphological operations.	96.3%(Drive),97.1%(STARE)

		post-processing to reconnect vessels.			
2022	Qian et al.	Developed an AI system to train medical students real world Chinese fundus images. Evaluating its ability in DR diagnosis and improving learning outcome.	Chinese Fundus Dataset	Deep learning-based classifier.	96.5%
2022	Kim, Mishra, Sen	Teleophthalmology system deployed in India, Combining ICT and AI tools to enable DR screening in rural areas.	-	-	-
2020	Gayarthi et al.	Hybrid feature extraction combining Haralick texture and Anisotropic Dual-Tree Complex wavelet Transform. Evaluation of multiple classifiers in binary and multi-class classification	MESSIDOR, DIARETDB80	Random Forest, Random tree, SVM, J48	Random Forest:99.7% Binary), 99.82% (multiclass).
2024	Senapati et al.	Surveyed ML and DL algorithms, and analysed issues in current systems	Various datasets	SVM, KNN, decision trees, KNNs	-
2025	Mutawa et al.	Hybrid CNN-RBF model with MS-DRLBP features and Otsu-based vessel segmentation, trained on multiple datasets	STARE, HRF, FFA	CNN-RBF, Naïve Bayes, RBF, ANFIS, NN, SVM	97.30% (CNN-RBF on STARE)
2019	Ratanapakorn et al.	Developed a MATLAB GUI to diagnose DR and highlight ROI in the retina	Private 400 image dataset	Unspecified	96.25%
2018	Gharaibeh et al.	Used a Hybrid ML algorithm with sophisticated pre-processing techniques	DIARETDB1	Hybrid SVM-GA algorithm	98.4%

Chapter 3: Methods and Materials

3.1. Introduction

This chapter details the design of the proposed system, expanding on the materials and methods used in its development. The system will be primarily composed of two parts: a digital slit lamp modified to capture retinal images, and a laptop or personal computer that performs all the processing and calculation steps required for the diagnosis and classification of DR. These steps include but are not limited to pre-processing, input classification, and image segmentation. In terms of classification, 4 state of the art DL algorithms are used in the development of the framework to provide the highest degree of accuracy and precision.

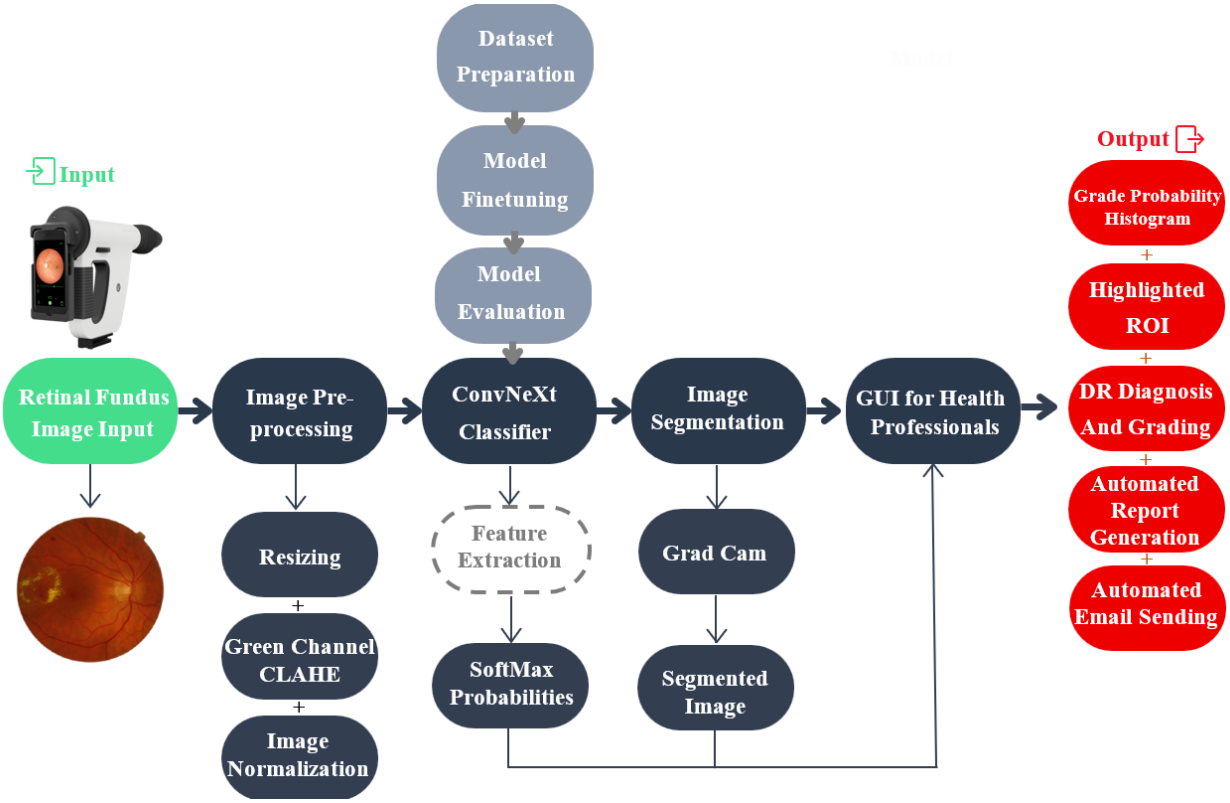

Figure 24-Block Diagram of the Proposed Framework

Additionally, a GUI that summarizes the findings of the automated diagnosis in a clear and concise manner is available for the medical staff, and the result may be relayed to patient and hospital staff via telemedicine.

3.2. Datasets Used

In order to finetune and test the classifiers proposed in this methodology, a sizeable number of retinal images is required along with their annotations to construct the ground truth. Datasets used in this project will include the publicly available APTOS 2019 dataset, as well as a private set of fundus images acquired locally.

3.2.1. APTOS 2019

Accessible via Kaggle, this dataset contains a total of 3662 retinal images mainly acquired at the Aravind Eye Hospital in India. Each of these images was expertly annotated by a group of medical professionals on a severity scale of 0 to 4 (0-Normal; 1-Mild; 2-Moderate; 3-Severe; and 4-Proliferative). Image distribution across the dataset is described in the table below:

Table 3 Image Distribution by Class in the APTOS 2019 Dataset

Class	Number of Images	% of Total
No DR	1805	49.29
Mild DR	370	10.1
Moderate DR	999	27.28
Severe DR	193	5.27
Proliferative DR	295	8.05

For multiclass classification or DR grading, each of the algorithms used in this study is trained and tested on all 5 classes as annotated in the dataset. However, in the case of binary classification, the last 4 classes are reduced into a single “DR” class which indicates the presence of the disease. As such, the distribution of the dataset becomes as listed in the following table:

Table 4 Image Distribution in the APTOS 2019 Dataset in Binary Classification

Class	Number of Images	% of Total
No DR	1805	49.29
DR	1857	50.71

Moreover, in both multiclass and binary classification, the dataset is split into 2 sets for training and testing containing 80% and 20% of the total images respectively.

3.3. Image Pre-Processing

The raw input image has to be pre-processed before being fed into the algorithm to enhance its quality and ensure it complies with the model's standards. This step can include techniques like:

1. **Resizing:** is a method done by adjusting the image size to fit the image model requirements, prevents model distortion and ensures consistency across the training samples. Images in our framework are resized to 224x224, which is the size expected by all of the 4 algorithms being tested.

2. **Normalization:** Is a method done by adjusting the image value pixels to a common scale. It aims to improve contrast by expanding a narrow range of input intensity values into a wide (stretched) range of output intensity values (usually the full range of gray values that can be displayed). Images in our system are normalized across all 3 channels in terms of mean and std. The values used for normalization include mean= [0.485, 0.456, 0.406], std=[0.229, 0.224, 0.225] for the red, green and blue channels respectively.

3. **Image enhancement:** Method used to improve contrast and highlight retinal features, in which Contrast limited adaptive histogram equalization (CLAHE) is used to enhance the visibility of vessels and lesions. Green channel enhancement can be used to isolate the vessels and hemorrhages and carry the most relevant anatomical information. Noise removal is also required for image enhancement in which it smooths artifacts and illumination irregularities. Edge sharpening and filtering, which is optional but better for pre-processing, can improve the lesion boundaries.

3.4. Model Development

This part elaborates on the 4 DL algorithms proposed for DR detection and grading, highlighting the strong points of each as well as their limitations.

3.4.1. ConvNeXt

As the field of image classification and segmentation continues to advance, traditional CNNs such as ResNet started to fall behind transformer-based vision models. To overcome this issue, researchers at Meta AI redesigned the architecture of ResNet by applying incremental changes so it can better match the performance of transformers while retaining the efficiency and simplicity of a typical CNN. The result was ConvNeXt which is a deep learning architecture inspired from transformer models [43].

ConvNeXt differs from the typical CNN in a multitude of ways. First, the network starts with 4x4 convolutions, like ViTs to reduce image resolution early, instead of the typical multiple 3x3 convolutions. Moreover, its inverted bottleneck layers, borrowed from MobileNet allows for efficient operations between channels, while normalization is performed on entire layers rather than batches to improve the training process in the case of small batch sizes. Finally, it applies 7x7 depth-wise convolutions instead of the usual 3x3, a change inspired by global attention in ViTs. These changes made to the architecture allowed to match or exceed the performance of transformers in benchmark tests. Most notably ConvNeXt achieved an accuracy of up to 87.8% on the ImageNet-1K benchmark test [43], surpassing many transformer models. Some of its pros and cons are listed below.

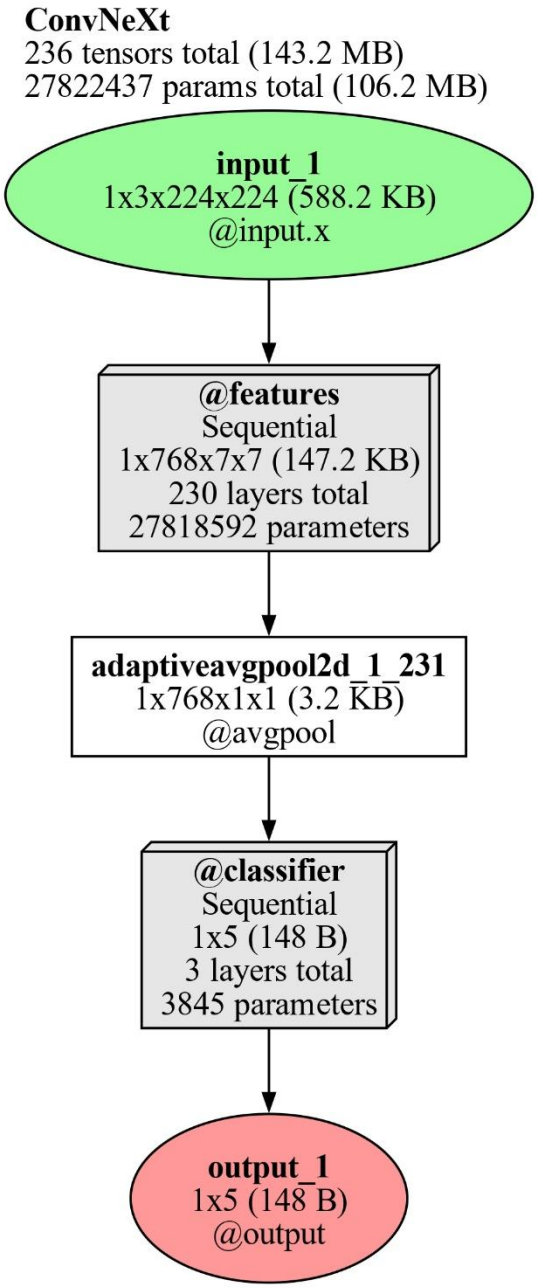

Figure 25 ConvNeXt Tiny Architecture with Custom Classification Head

3.4.2. Vision Mamba

Developed as an extension of the Mamba DL architecture, vision mamba adapts its principle of efficient sequence modelling to images, marking a significant advancement in the field of computer vision. It utilizes a state-space model (SSM) based architecture optimized for visual representation learning. The model is mainly used for object detection, as well as image classification and segmentation. Some of its key architectural features include bidirectional state space modelling, meaning that it processes visual data both forward and backwards, which allows it to efficiently capture context from images without the complexity of self-attention. Moreover, the model utilizes positional embedding to maintain spatial relations within images and has a hierarchical design which allows it to capture features at multiple levels of abstraction [44]. In terms of performance, the model proved to be a formidable algorithm, with ViM-Base version scoring a top 1 accuracy of 81.9% on ImageNet-1K [45], which showcased its efficacy in image classification tasks. Hybrid models based on the vision mamba architecture scored even higher such as NVIDIA's MambaVision-L3-512-21K which achieved a top 1 accuracy of 88.1% [46].

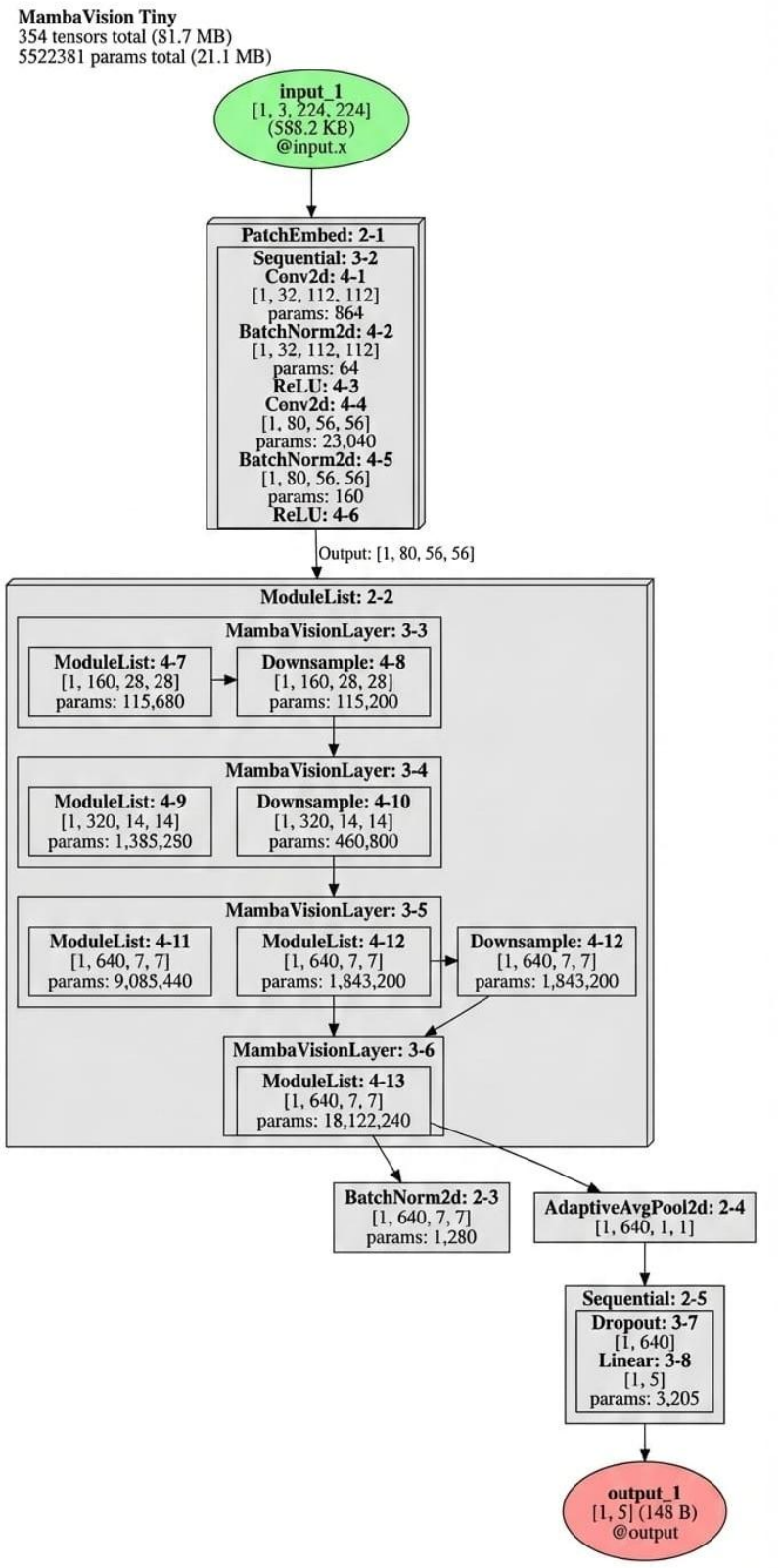

Figure 26 MambaVision Tiny Architecture with Custom Classification Head

3.4.3. Vision Transformers (Vit) Network

ViT has recently emerged as a competitive architecture to CNNs that are currently state of the art (SOTA) in different image recognition computer vision tasks. ViT models outperform the current SOTA CNNs by almost x4 in terms of computational efficiency and accuracy. ViT adapted the transformer for computer vision tasks like image classification, object detection and semantic image segmentation. Which has achieved a highly competitive market in computer vision.

In VITs images, images are represented as sequences, and they are class labels for the images predicted in which allow the models to learn the structure of the image independently [47]. The model achieved an 98% top 1 accuracy on the benchmark [48].

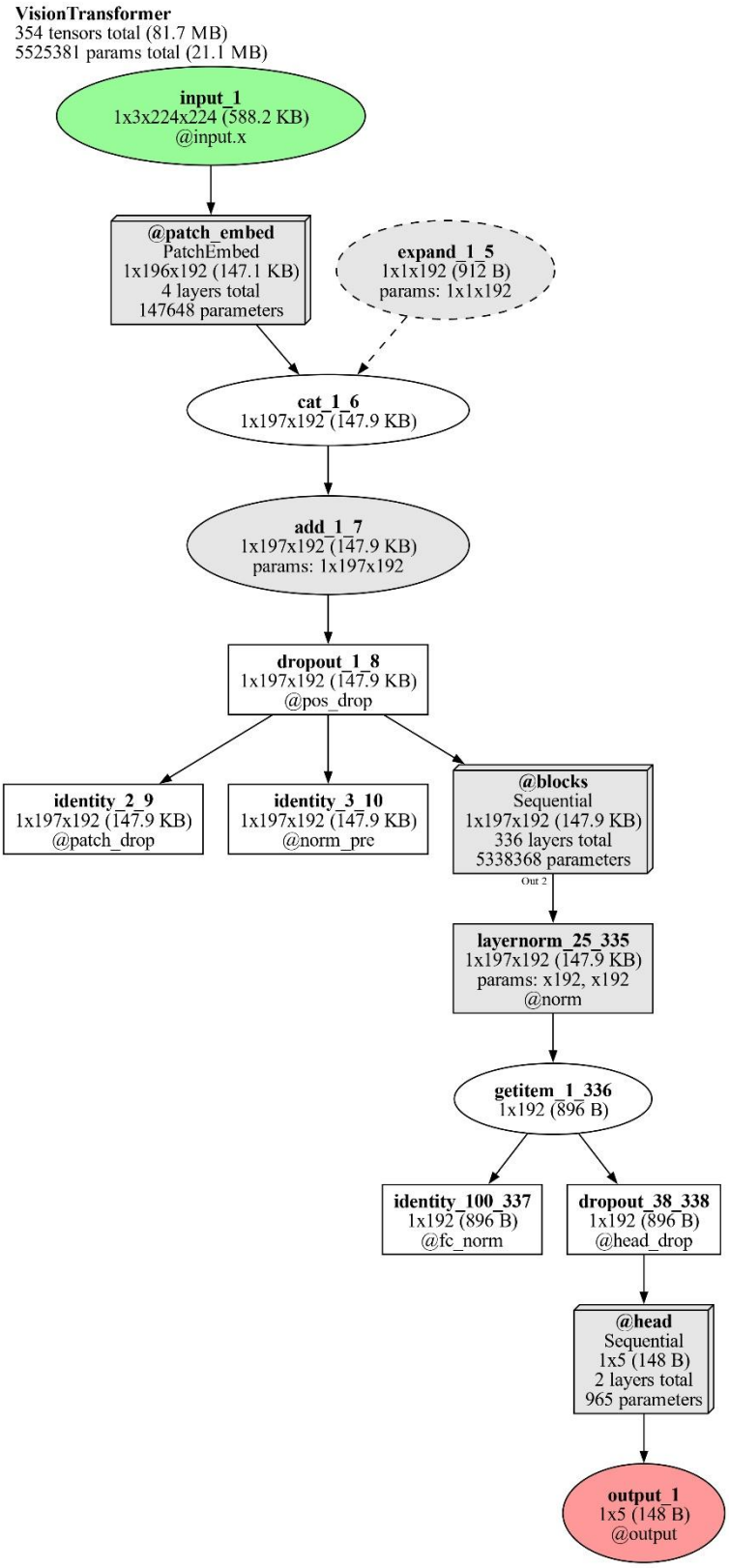

Figure 27 VIT Tiny Architecture with Custom Classification Head

3.4.4. Swin Transformer

Swin transformer which refers to shifted window transformer, it is hierarchical transformer that processes the image efficiently, it introduces mechanisms of window-based attention and shifted windows which improves the performance and scalability for high resolution images like HDR photos. Swin is a type of ViT architecture, it was designed to handle computer vision tasks like image classification, object detection [49].

The name “Swin” came from its window-based attention which shifts across the image to extract features. Swin Transformer designed to address challenges faced by other traditional transformers regarding inefficient working with high resolution images. So, it achieves a balance between high accuracy and computational efficiency. The swim transformer architecture is built on a combination between the hierarchical design and window-based self-attention for efficient working and feature extraction. The model achieved a top 1 accuracy of 87.3% on the ImageNet-1k benchmark [50].

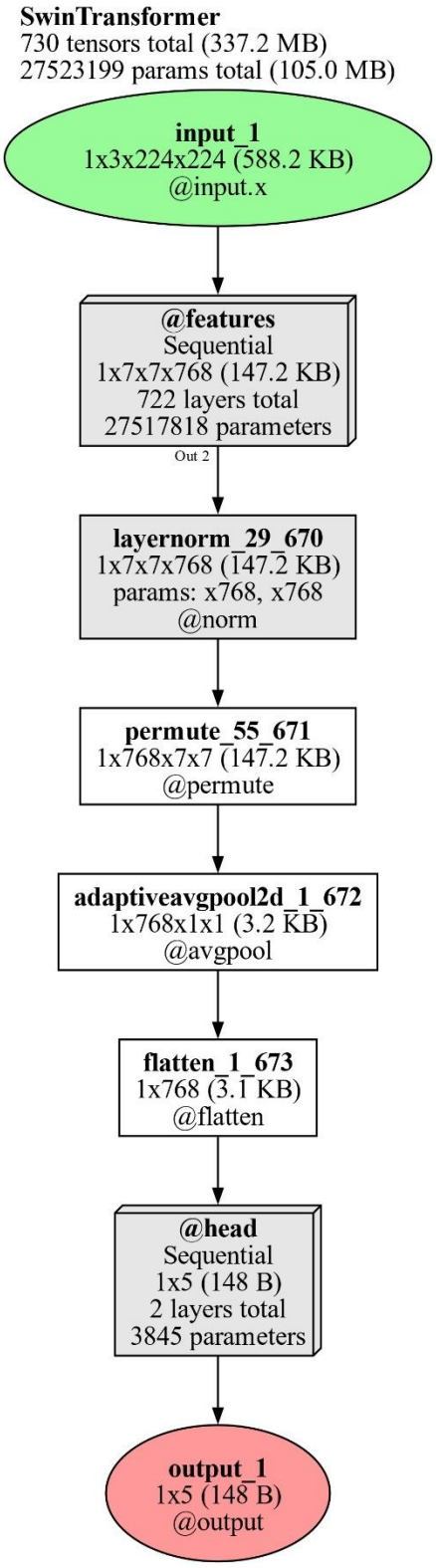

Figure 28 Swin Transformer Tiny Architecture with Custom Classification Head

Table 5 Summary of the 4 Algorithms

Model	Year	Description	Top-1 Accuracy on ImageNet-1K
Swin Transformer	2021	Hierarchical vision transformer with shifted windows for efficient and scalable learning.	87.3% (Swin-L, 384×384)
ConvNeXt	2022	A modernized ConvNet architecture that competes with transformers by integrating design elements from them.	87.8% (ConvNeXt-L, 384×384)
Vision Transformer	2020	First major model to apply pure transformer architecture to image classification.	89% (ViT-L, pretrained on CLIP + fine-tuned)
MambaVision	2024	A hybrid vision architecture combining Mamba (state-space models) with transformers for scalable vision tasks.	88.1% (MambaVision-L3-512-21K)

The table above provides a summary of the importance of each algorithm, its release date, as well as its classification performance. The table below summarizes the properties of each of these networks.

Table 6 Properties of the 4 Algorithms

Model	Size (MB)	Parameters (Millions)	Input Image Size
Swin Transformer	105	27.5	224 by 224
ConvNeXt	106.2	27.8	224 by 224
Vision Transformer	21.1	5.5	224 by 224
MambaVision	124.4	31.1	224 by 224

3.4.5. Logistic Regression

Logistic regression is a supervised machine learning algorithm used for classification tasks. It predicts the probability that an instance belongs to class or not. It is a statistical algorithm that analyzes the relationship between data factors. This algorithm uses binary classification where sigmoid function is used in which it takes input as an independent variable and produces a probability value between 0 and 1 [51]. Logistic Regression can be classified into three types:

- **Binomial**: Binomial logistic regression can have two types of dependent variables such as 0 and 1.
- **Multinomial**: In multinomial type, there can be three or more unordered types of dependent variable.
- **Ordina**l: In ordinal type, there can be three ordered types of ordered dependent variables.

The most important function that forms the core of logistic regression is sigmoid function. It is a function that is used to map the predicted values to probabilities. It maps any real value into a value between 0 and 1. The logistic Regression model transforms the linear regression system continuous value output into categorical output value output using the sigmoid function. So let the independent input features be a matrix X, and the dependent value be Y as a binary number 0 or 1 for class 1 and class 2 respectively. Then apply the multi-linear function to the input variable X. Which is:

$$z = (\sum_{i=1}^{n} wixi + b) \qquad \textit{Equation 12}$$

Where:

- wi= [w, w2, w3....] are the weights or the coefficients
- xi is the observation of x
- b is the bias term or intercept [51].

3.5. Training Procedure

Since all algorithms to be used in this system are transformer-based, there is evident overlap between the procedures and methods used to train each model. Regardless, the specifics of training for both binary and multiclass classification are illustrated using the tables below:

Table 7 Training Hyperparameters for Binary Classification

Hyperparameter	Head LR	Feature/Block LR	LR Scheduler	Loss Function	Optimizer	Early Stopping Patience (Epoch)	Batch Size
Value	3e-4	3e-5	Cosine Annealing	BCEWithLogitsLoss	AdamW	9	64

The simplicity of the classification task in the binary setting made it so that the training hyperparameters were almost identical across all 4 models. Some slight differences may remain in terms of pre-processing and weight decay.

Table 8 Training Parameters for Multiclass Classification

Network Model	Swin Transformer Tiny	ConvNeXt Tiny	VIT Tiny	MambaVision Tiny
Head LR	2e-4	3e-4	3e-4	3e-4
Feature/Block LR	5e-5	3e-4	3e-5	3e-5
LR Scheduler	ReduceLROnPlateau	Cosine Annealing	Cosine Annealing	ReduceLROnPlateau
Loss Function	Focal Loss	Focal Loss	Focal Loss	Focal Loss
Weight Decay	1e-4	0.02	0.05	0.02
Optimizer	AdamW	AdamW	AdamW	AdamW
Early Stopping Patience (Epoch)	13	12	12	13
Batch Size	64	64	64	64

As observed in table 8, due to the similarities in model architecture, hyperparameter values were also similar. This is apart from the LR scheduler which was more model dependent, and the early stopping patience that would consequently change with a change in scheduler choice. This is done since the ReduceLROnPlateau has an internal patience which denotes the number of epochs without improvement before LR is reduced. This patience was often set to 3 during training, consequently raising the Early Stopping Patience to 13, gave the algorithms one more "life" before early stopping is triggered. Note that Early Stopping Patience in all cases is higher than its binary counterpart as models take longer to converge during multiclass training. Additionally, a batch size of 64 was used for all models to ensure that each gradient update sufficiently aids in generalization.

Although standard cross entropy loss performs great in multiclass settings, the nature of the dataset, characterized by its small size and high imbalance, forced the use of focal loss to achieve better results. Focal Loss is a modified version of cross entropy loss that penalizes the model for misclassifications occurring in minority classes. Besides focal loss, loss trimming was also applied to limit the impact of potentially mis-annotated samples from largely influencing training loss. Different models benefited from different loss parameters as illustrated in table 9.

Table 9 Loss Parameters in Multiclass Classification

Loss Parameters	Swin Transformer Tiny	ConvNeXt Tiny	VIT Tiny	MambaVision Tiny
Gamma	2	1	2	2
Label Smoothing	0.05	0.1	0.05	0.05
Trim Ratio	0.05	0.05	0.05	0.05

3.6. Experimental Setup

All models, in both binary and multiclass setting, were finetuned and tested using the same hardware environment described below:

Table 10 Training Device Specifications

Property	RAM (GB)	CPU	GPU
Specification	16	Intel® Core™ i7-14650HX	NVIDIA GeForce RTX 4060

As for the software environment used, it was mainly the same except for the one used to train MambaVision which had components that were incompatible with windows.

Table 11 Training Software Specifications

Software Specification	Operating System	Containerized	Language	Framework
MambaVision	Linux (Ubuntu using WSL 2.6.1.0)	Yes	Python 3.11.11	Torch 2.6.0+cu126
Other Algorithms	Windows 11 (25H2)	No	Python 3.11.5	Torch 2.5.1+cu121

3.7. Model Evaluation

After the model is trained and fine-tuned, testing will be carried out on a separate validation set. This will be done to measure how accurately it classifies DR images and distinguishes between different grades of the disease. Evaluation will be mainly split into 2 types: quantitative and qualitative. Quantitative evaluation produces objective numerical figures that accurately reflect the model's performance, whereas qualitative assessment is concerned with the generation of graphs and charts that allow visual inspection and interpretation of results.

3.7.1. Quantitative Evaluation

1. **Accuracy:** Proportion of total correct predictions

$$\text{Accuracy}=\frac{\text{TP}+\text{TN}}{\text{TP}+\text{TN}+\text{FP}+\text{FN}} \qquad \textit{Equation 13}$$

2. **Sensitivity (Recall):** Ability to identify positive DR cases

$$\text{Sensitivity}=\frac{\text{TP}}{\text{TP}+\text{FN}} \qquad \textit{Equation 14}$$

3. **Specificity:** Ability to identify negative DR cases (Normal)

$$\text{Specificity}=\frac{\text{TN}}{\text{TN}+\text{TP}} \qquad \textit{Equation 15}$$

4. **Precision:** Proportion of correct positive predictions

$$\text{Precision}=\frac{\text{TP}}{\text{TP}+\text{FP}} \qquad \textit{Equation 16}$$

Where: *TP* is True positive, *TN* is true negative, *FP* is false positive, *FN* is false negative

5. **F1 Score:** Harmonic mean of precision and recall

$$\text{F1}=\frac{2(Precision \cdot Recall)}{\text{Precision}+\text{Recall}} \qquad \textit{Equation 17}$$

6. **Macro F1 Score:** The average F1 score of several classes

$$\text{Macro F1}=\frac{1}{K}\sum_{i=1}^{K} F1_i \qquad \textit{Equation 18}$$

Where: K is the total number of classes, and i is the class number

7. **Weighted F1 Score:** Weights each class by its supports

$$\text{Weighted F1}=\sum_{i=1}^{K}\left(\frac{n_i}{N}\cdot F1_i\right) \qquad \textit{Equation 19}$$

Where: n_i is the number of true samples in class i, N is the total number of samples

8. **Entropy:** Measures uncertainty in probability distribution of the model's output

$$H(p) = -\sum_{i=1}^{K} p_i \log(p_i) \qquad \textit{Equation 20}$$

Where: p_i is a component of probability vector p of K components

9. **Kurtosis:** Measures "Tailedness" of a distribution

$$Kurtosis = \frac{1}{N}\sum_{i=1}^{K}\left(\frac{p_i-\mu}{\sigma}\right)^4 \qquad \textit{Equation 21}$$

Where: μ is the mean of the probability distribution and σ is its standard deviation

10. Quadratic Weighted Kappa: Measures agreement between two raters

$$QWK = 1 - \frac{\sum_{i,j} w_{ij} O_{ij}}{\sum_{i,j} w_{ij} E_{ij}} \qquad \textit{Equation 22}$$

Where: w_{ij} is the weight penalizing distance between 2 classes *i* and *j*

O_{ij} is the proportion of samples where rater A gave class *i* and rater B gave class *j*

E_{ij} is the proportion expected by chance

3.7.2. Qualitative Evaluation

1. **ROC Curve:** A Receiver Operating Characteristic curve is a curve drawn to illustrate the performance of a binary model across all classification thresholds. The performance is reflected through the true positive rate on the y-axis and the false positive rate on the x-axis. A theoretically perfect performance would have a point (typically the top left of the graph) where the true positive rate is 1 and the false positive rate is 0. This would mean that the model managed to find a classification threshold that maximizes sensitivity while minimizing misclassification. In this case the AUC or the area-under-curve would be equal to 1. Although such a qualitative metric is primarily used for binary classification, it can also be used to evaluate multiclass performance. This is done by collapsing all classes besides the

class of interest into one class. In this case the ROC curve is referred to as a “One-vs-Rest” Curve.

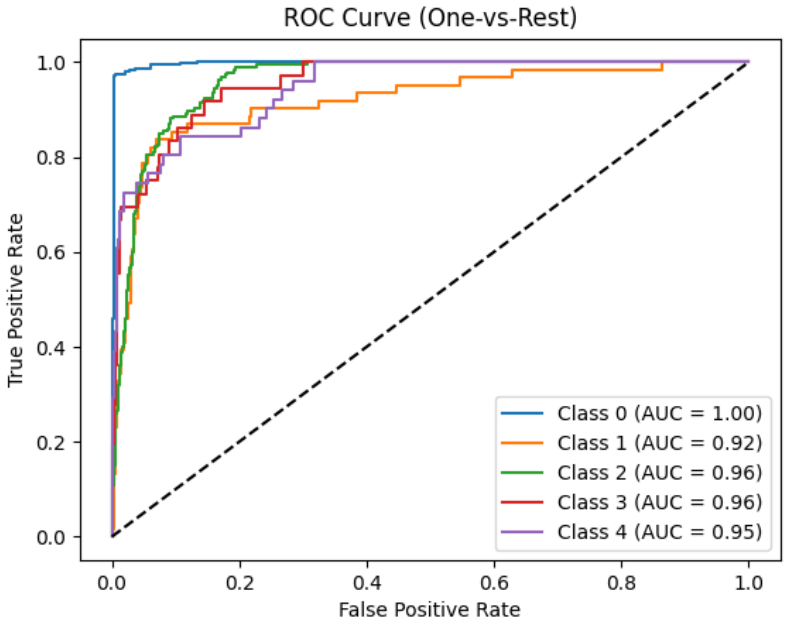

Figure 29 Example of One-vs-Rest ROC curve generated during training

2. **Confusion Matrix:** A confusion matrix is one of the most important metrics in assessing the accuracy of multiclass classification. As the name suggests, the matrix illustrates which classes the model ‘confuses’ with others by generating a heatmap with the true class of the sample on the y-axis and the predicted class on the x-axis. As such, a perfect confusion matrix would show a diagonal line extending from the top left corner to the bottom right. This plot helps understand model behavior and allows the researcher or engineer to adjust the training parameters as well as the pre-processing pipeline such that the model stops confusing similar classes.

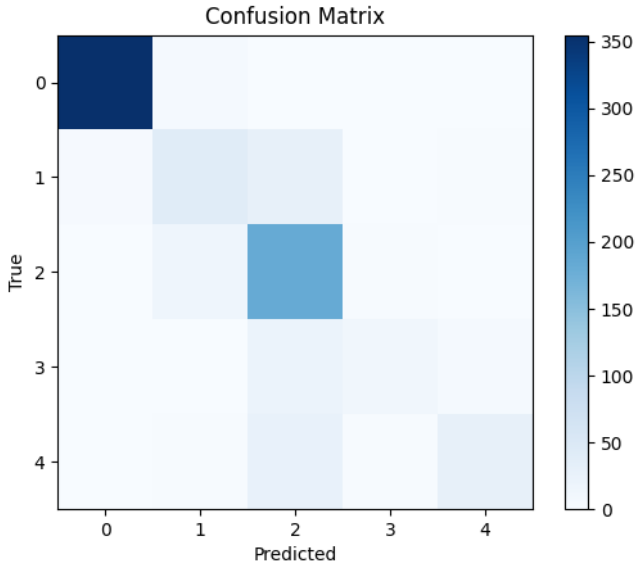

Figure 30 Example of Confusion Matrix Generated During Training

3. **Accuracy and Loss vs Epoch:** During training, especially at the beginning of the process, the loss and accuracy the model scores on the validation set can vary quite significantly. This is why constructing a plot that tracks how these 2 metrics trend as training epochs go on can help in understanding the training behavior of the model and provides evidence to go by in case overfitting is suspected.

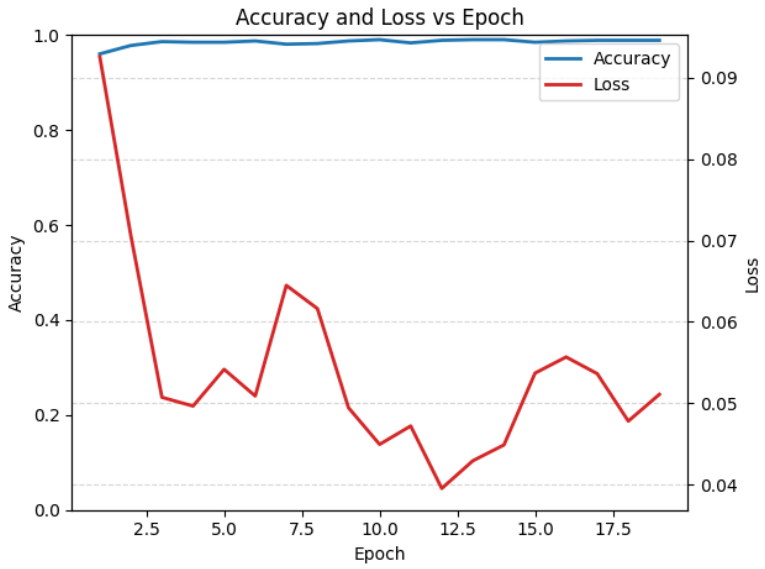

Figure 31 Example of Accuracy and Loss vs Epoch Graph Generated During Training

4. **Box Plot:** Since the final layer of each algorithm contains an activation function which generates probability scores (0-1 range), tools used to describe probability distribution can also be used here. Box plots are most commonly used in multiclass performance evaluation to reflect class distribution in the test set as well as the confidence exhibited by the model during classification. A lower probability score is indicative of lower confidence, thus a wide box plot for a particular class indicates the model lacks confidence when classifying its samples.

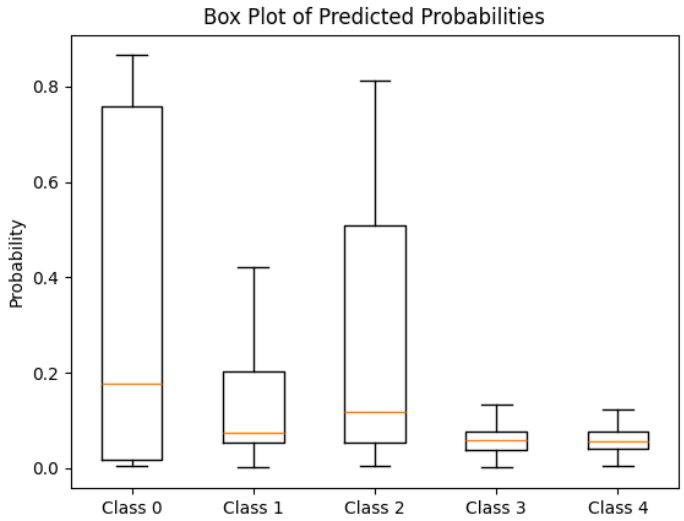

Figure 32 Example of Generated Box Plot During Training

5. **Probability Histogram:** An alternative way to describe a probability distribution. It is used in the case of binary classification with a roughly equal class sample distribution to assess the confidence of the model's predictions.

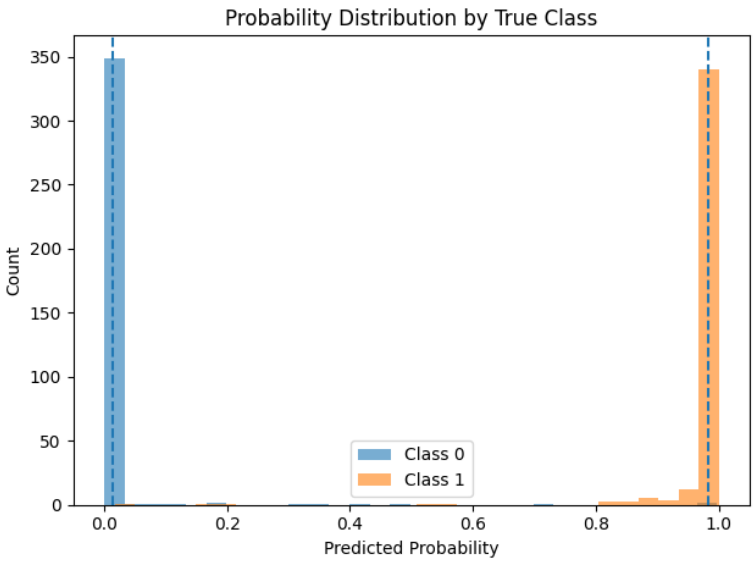

Figure 33 Example of Probability Histogram Generated During Training

3.8. Image segmentation

Image segmentation is used in the prototype to clearly display and highlight retinal structures that correspond to DR pathology. This directs the attention of the medical professional to the retinal features that may have influenced the decision of the classifier.

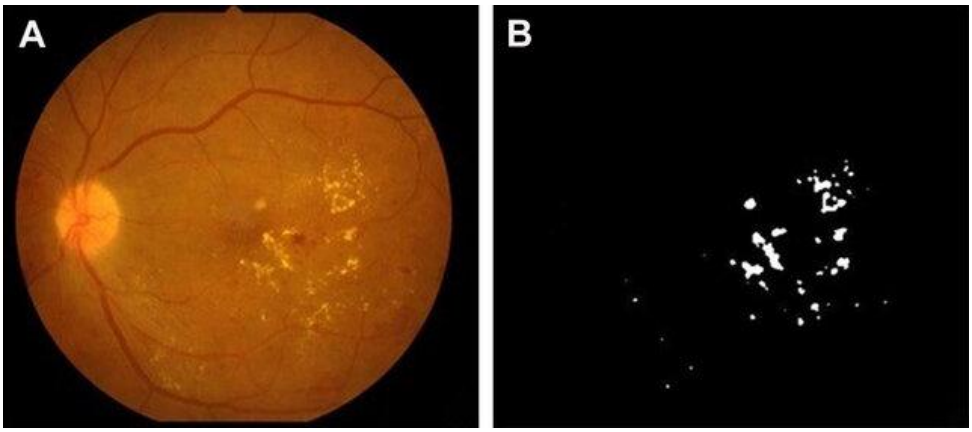

Figure 34 Fundus Image (A) and a Binary Mask Highlighting its Lesions (B)

Since the proposed framework makes use of advanced DL algorithms, it would only be appropriate to perform segmentation in a manner that holds the same standard. This is why the segmentation will be performed using Grad-Cam. Grad-Cam short for Gradient-Camera, is a visualization technique used to highlight the regions of the image that most influence the gradient of the model during classification, and thus its decision [52]. The tool generates a heatmap of gradient activation in the algorithm used which is then overlayed on the original image, providing the user with insight into the inner-workings of the classifier which increases trust. Typically, the final few layers of the classifier are included in gradient visualization, as these are the ones that most accurately correspond to DR pathology.

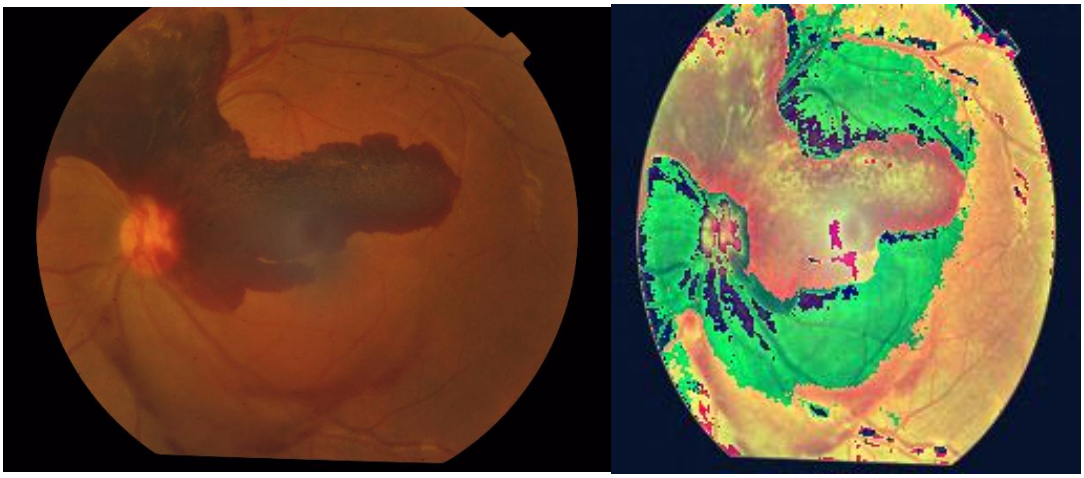

Figure 35 A CFP Along with its Grad-Cam Segmented Image

3.9. Graphical User Interface

A GUI is developed to allow clinicians, technicians or users to interact with the model with limited background knowledge; in other words, it is a program that enables a person to leverage these advanced DL models through simple buttons. To enable doctors or technicians to engage with the model without requiring technical expertise, an intuitive interface design has been planned. The GUI is to be designed in its entirety in MATLAB and then interfaced with python modules as needed.

Some of the functions the GUI intends to serve:

1. Allows the user to upload a desired retinal image for analysis
2. Allows the user to insert patient information
3. Automatically obtains classification results such as diagnosis and DR severity
4. Displays attention maps or Grad-Cam overlays (to explain model decisions)
5. Automatically generates a patient report to be saved locally
6. Automatically sends record to email address specified by the user

The last option is intended to serve as the telemedicine contribution of the project which will facilitate the process of health record digitalization as well as communication between different clinics and doctors. The telemedicine module can also potentially be used to send recommendations and guidelines to each patient automatically generated on a case-by-case basis.

3.10. Ethical Considerations:

This study was conducted using a publicly available retinal fundus image dataset that was collected and released in accordance with established ethical guidelines. The dataset consists of anonymized images, with all personally identifiable information removed prior to public release, thereby minimizing risks related to patient privacy and data confidentiality. Ethical approval and informed consent were obtained by the original data collectors, and no direct patient interaction was involved in this work.

The proposed system is intended strictly as a research prototype and is not designed to function as a standalone diagnostic tool. While the models demonstrate strong performance in both binary screening and multiclass grading tasks, automated predictions should not replace

clinical judgment. In real-world settings, such systems should be used only as decision-support tools under the supervision of qualified medical professionals.

Potential sources of bias must also be acknowledged. The dataset represents a specific population and imaging conditions, which may limit generalizability to other demographic groups or acquisition devices. Additionally, class imbalance may influence model behavior in multiclass classification scenarios. These factors highlight the importance of further validation on diverse, real-world datasets prior to any clinical deployment.

Finally, responsible deployment of medical AI systems requires transparency, interpretability, and continuous monitoring. Future work should consider cross-dataset validation, calibration analysis, and collaboration with clinical experts to ensure ethical and safe application of such technologies.

Chapter 4: Results

4.1. Quantitative Training Results

Table 12 Statistical Results for Binary DR Classification

Network Model	Swin Trans-former Tiny	ConvNeXt Tiny	VIT Tiny	MambaVision Tiny
Accuracy (%)	**99.18**	**99.18**	98.36	97.81
Sensitivity (%)	**100**	99.19	99.19	98.66
Specificity (%)	98.34	**99.17**	97.51	96.95
Precision (%)	98.41	**99.19**	97.62	97.09
F1 Score (%)	99.2	**99.19**	98.4	97.86

Table 13 Training Behavior Summary Table for Binary DR Classification

Network Model	Swin Trans-former Tiny	ConvNeXt Tiny	VIT Tiny	MambaVision Tiny
Total Training Time (m)	34.65	36.74	35.57	**25.99**
Average Time Per Epoch (s)	122.29	122.46	**85.37**	97.48
Best Epoch	8	9	16	**7**
Training Loss at Best Epoch	0.03	**0.02**	0.06	0.04
Validation Loss at Best Epoch	**0.03**	**0.03**	0.06	0.06

Although the Swin transformer demonstrated perfect sensitivity for DR detection (100%), ConvNeXt has it beat for the title of best DR detection algorithm. This becomes apparent when considering that ConvNeXt has the highest specificity, precision and F1-score among all 4 tested algorithms (99.17%, 99.19%, and 99.19%). Furthermore, ConvNeXt ties with the Swin transformer for best detection accuracy (99.18%). On the other hand, MambaVision exhibited the lowest scores in all evaluation metrics among the 4 algorithms.

In terms of training behavior, all models took about 35 minutes apart from MambaVision which required only 25.99 minutes to converge. However, the shortest average time per epoch belonged to VIT at 85.37 seconds with MambaVision closely trailing behind at 97.48 seconds while Swin and ConvNeXt each took about 122 seconds. Additionally, VIT only registered its best epoch at 16 while all 3 other models had their best epoch before epoch number 10. Finally, all 4 algorithms exhibited negligible training and validation loss at their best respective epochs, which was below 0.1 in all cases.

Table 14 Statistical Results for 5 Class DR Grading

Network Model	Swin Transformer Tiny	ConvNeXt Tiny	VIT Tiny	MambaVision Tiny
QWK (%)	90.65	**91.2**	90.13	90.24
Accuracy (%)	**84.85**	83.76	84.31	82.94
Macro F1 (%)	69.61	67.91	**70.04**	63.91

Table 15 Training Behavior Summary Table for 5 Class DR Grading

Network Model	Swin Trans-former Tiny	ConvNeXt Tiny	VIT Tiny	MambaVision Tiny
Total Training Time (m)	87.76	**40.78**	50.94	59.59
Average Time Per Epoch (s)	122.45	122.33	**84.9**	96.63
Best Epoch	30	**8**	24	24
Training Loss at Best Epoch	**0.17**	0.45	0.25	0.22
Validation Loss at Best Epoch	**0.07**	0.11	**0.07**	**0.07**

ConvNeXt extended its dominance to multiclass DR classification where despite scoring a modest accuracy of 83.76%, it exhibited the highest qwk score among the 4 algorithms (91.2%). The highest accuracy was achieved by the Swin transformer (84.85%), whereas the best macro F1-Score belonged to VIT (70.04%). Notably, MambaVision once again scored considerably lower than the other algorithms in accuracy as well as macro F1-Score (82.94% and 63.91% respectively).

Training data shows that ConvNeXt managed to achieve the highest qwk despite having the shortest total training time (40.78 minutes) and registering the 8th epoch as its best. Moreover, the algorithm had the highest training loss of 0.45 at its best epoch. On the other hand, the longest total training time was claimed by the Swin transformer (87.76 minutes) as well as the lowest amount of training loss at its best epoch (0.17). Additionally, VIT registered the shortest average time per epoch once again (84.9 seconds), and all algorithms exhibited negligible validation loss at their best epochs (<0.1), except for ConvNeXt (0.11).

4.2. Qualitative Training Results

4.2.1. Binary Classification

4.2.1.1. Swin Transformer Tiny

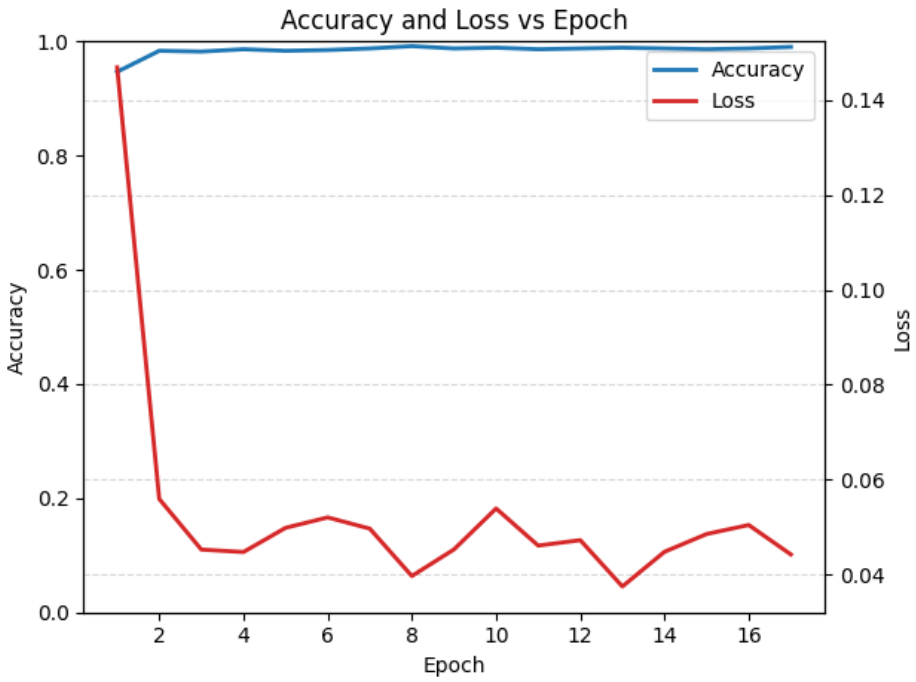

Figure 36 Swin Tiny Accuracy and Loss vs Epoch for Binary Classification

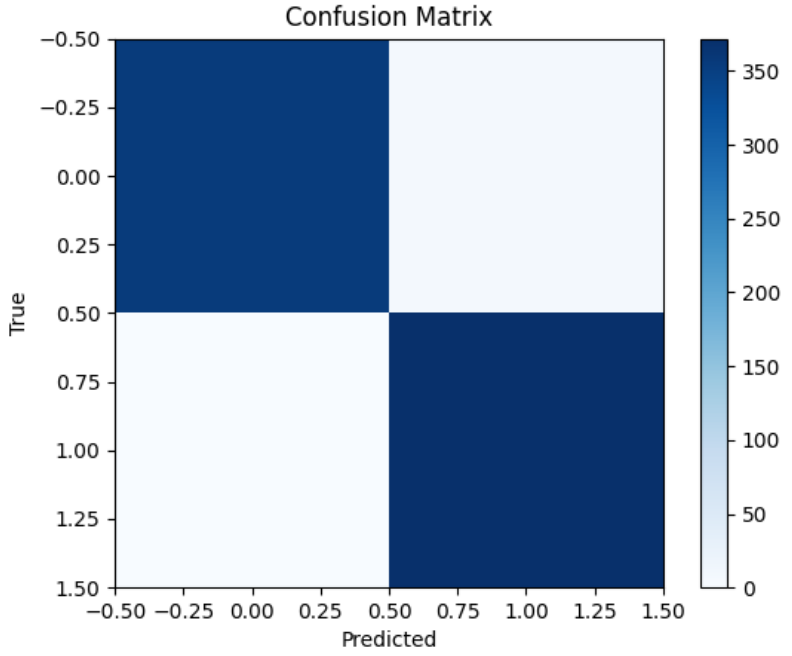

Figure 37 Swin Tiny Confusion Matrix for Binary Classification

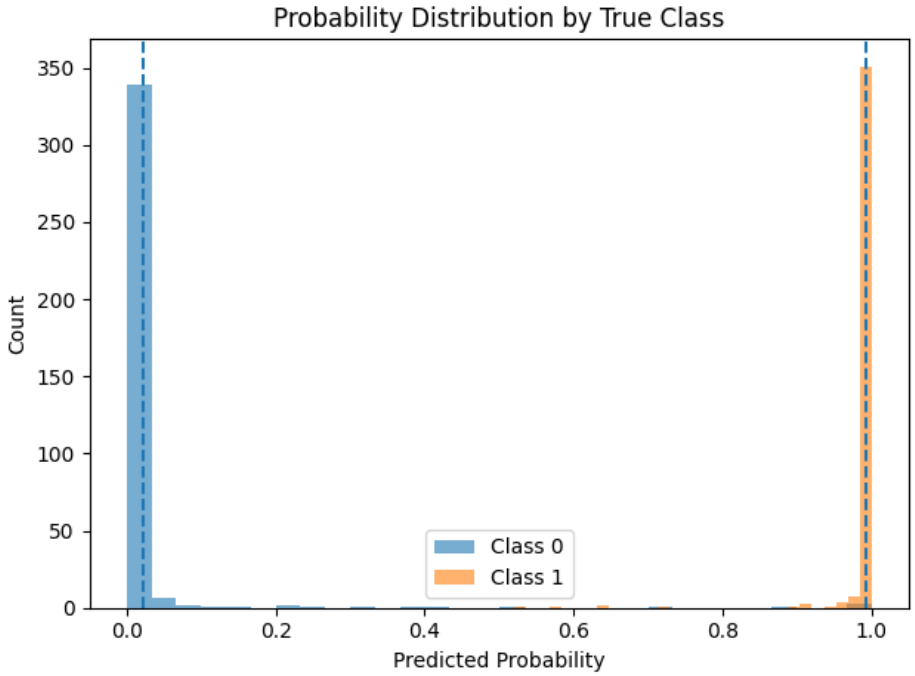

Figure 38 Swin Tiny Probability Histogram

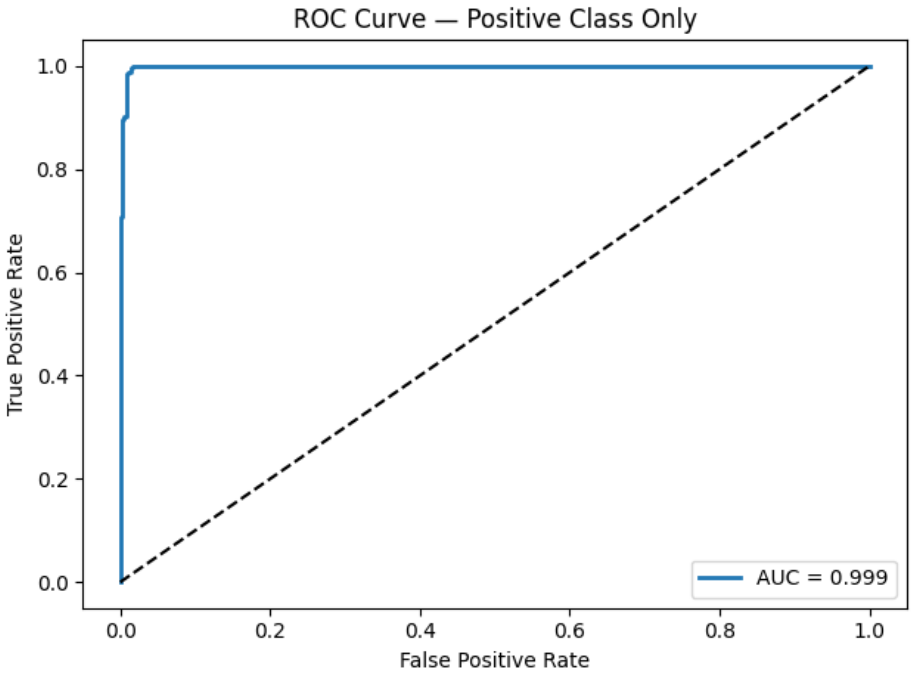

Figure 39 Swin Tiny ROC Curve for Binary Classification

4.2.1.2. ConvNeXt Tiny

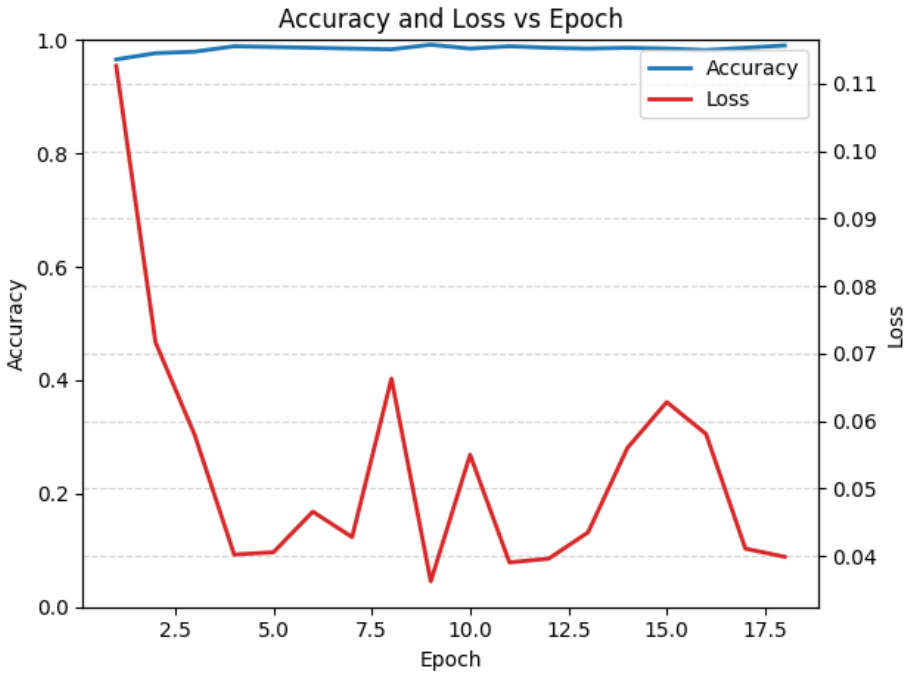

Figure 40 ConvNeXt Tiny Accuracy and Loss vs Epoch for Binary Classification

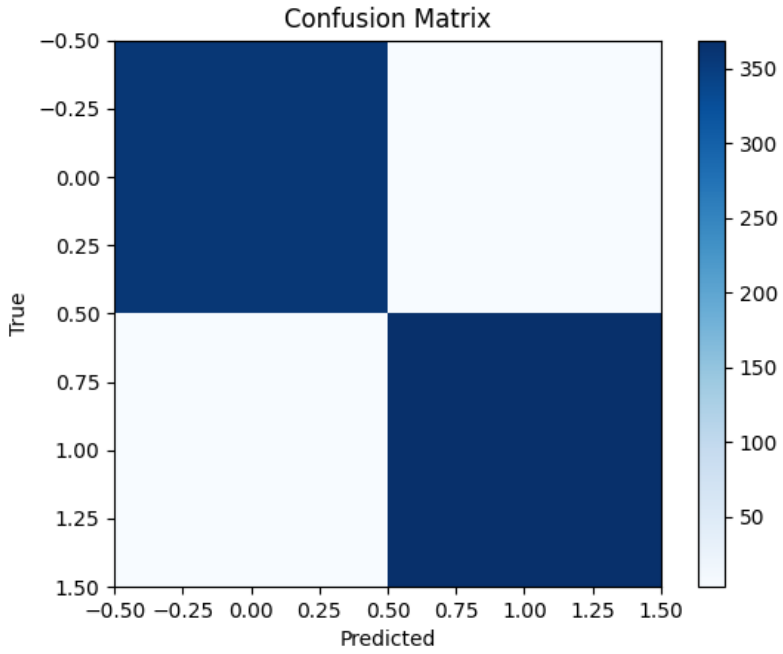

Figure 41 ConvNeXt Tiny Confusion Matrix for Binary Classification

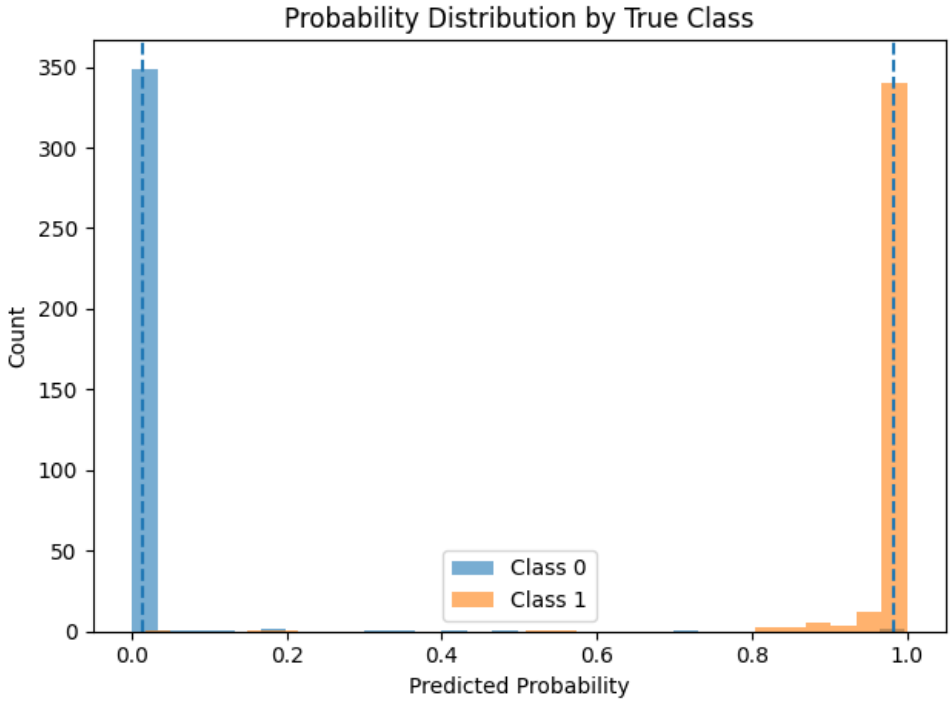

Figure 42 ConvNeXt Tiny Probability Histogram

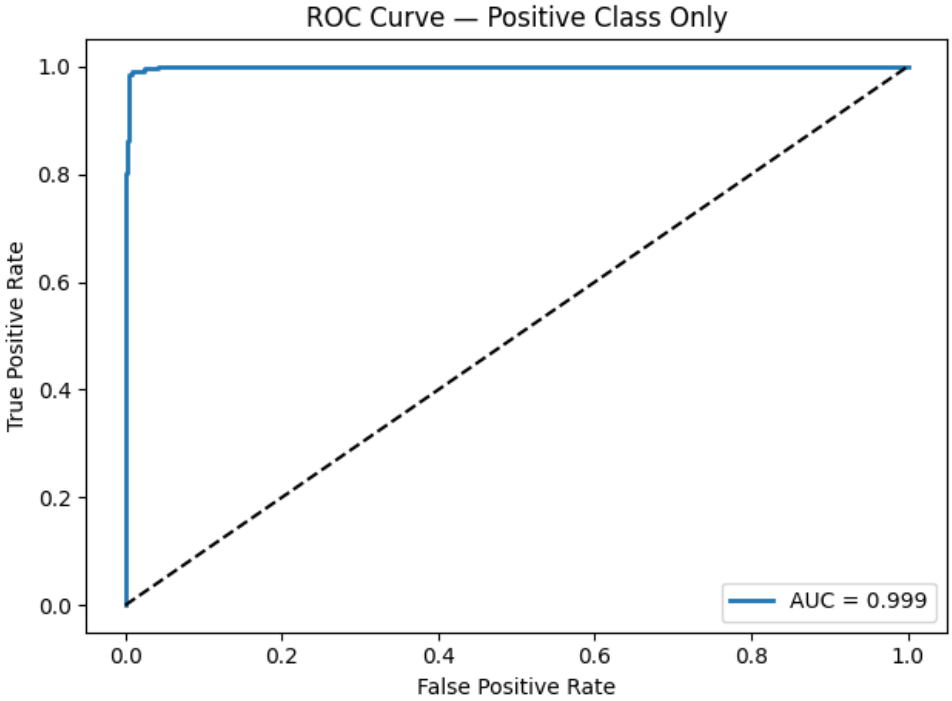

Figure 43 ConvNeXt Tiny ROC Curve for Binary Classification

4.2.1.3. Vision Transformer Tiny

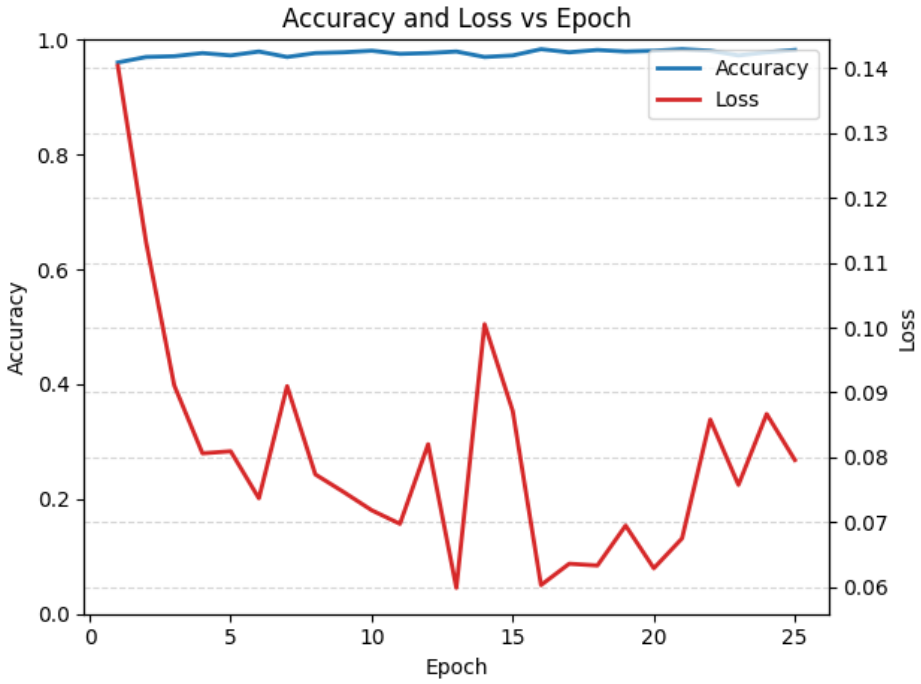

Figure 44 VIT Tiny Accuracy and Loss vs Epoch for Binary Classification

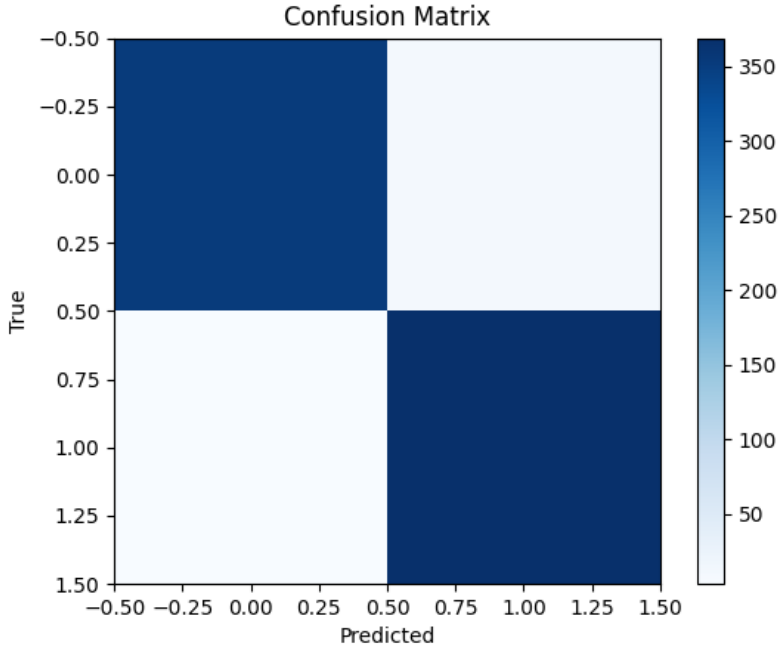

Figure 45 VIT Tiny Confusion Matrix for Binary Classification

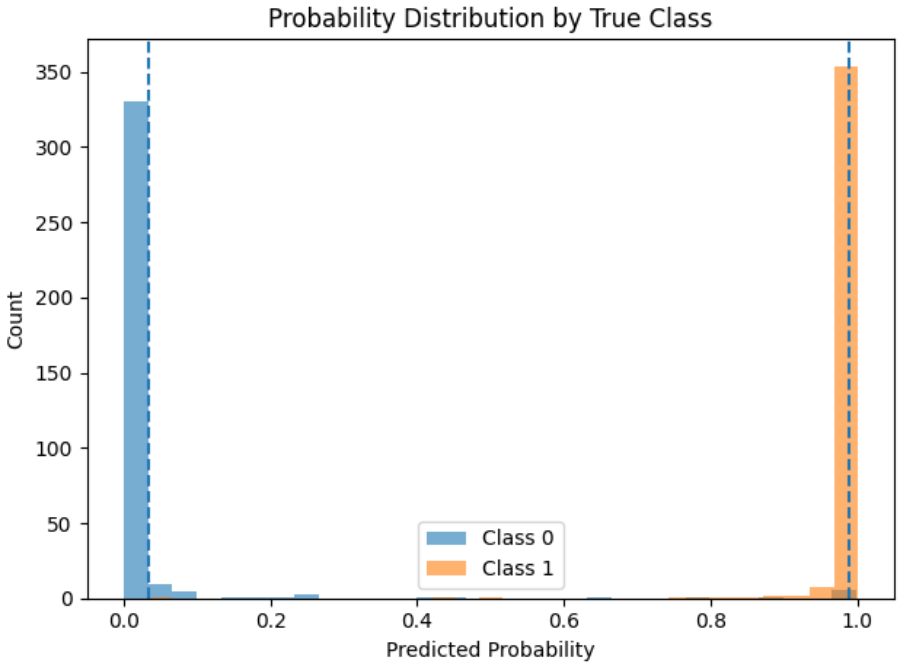

Figure 46 VIT Tiny Probability Histogram

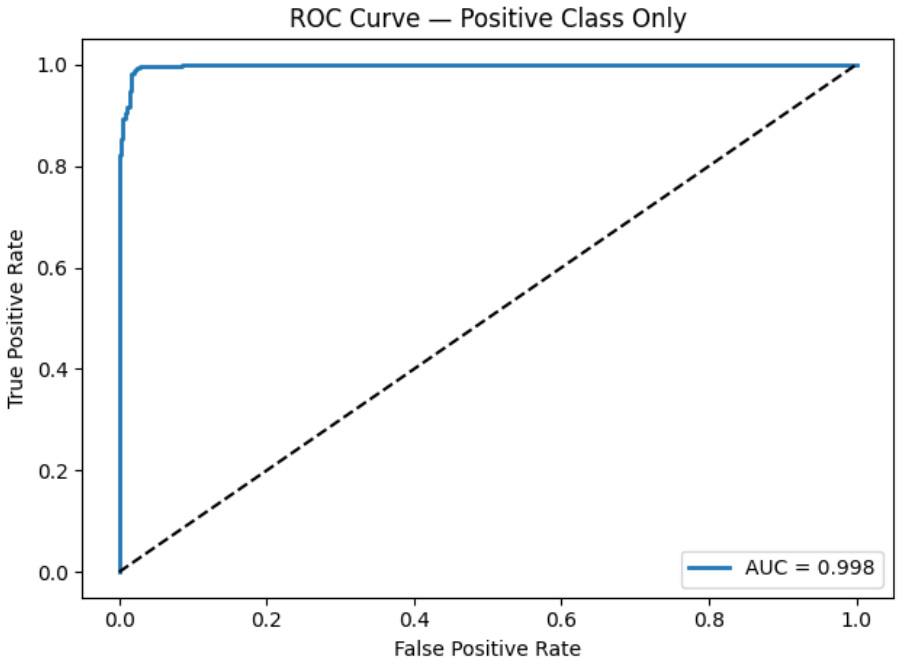

Figure 47 VIT Tiny ROC Curve for Binary Classification

4.2.1.4. MambaVision Tiny

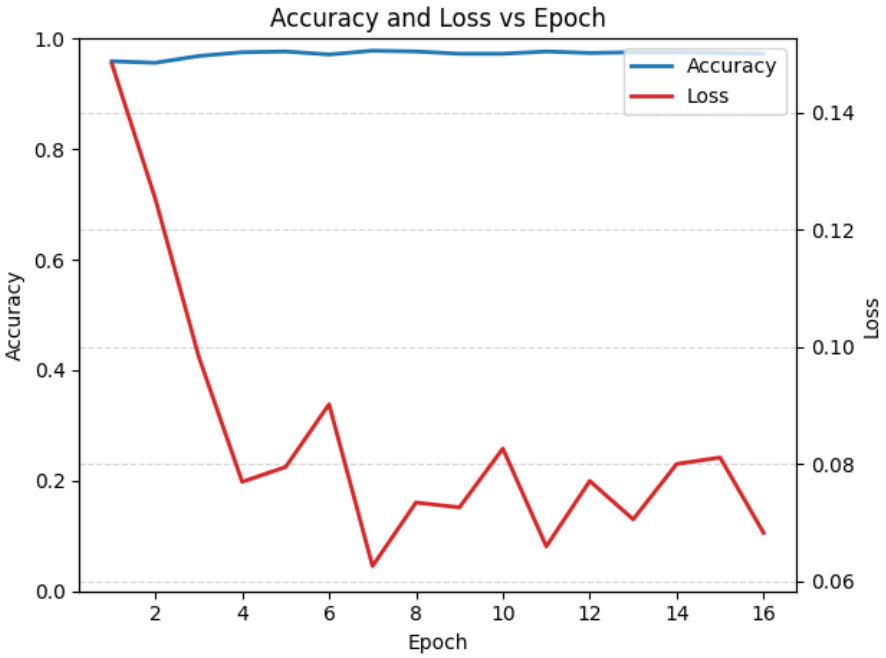

Figure 48 MambaVision Tiny Accuracy and Loss vs Epoch for Binary Classification

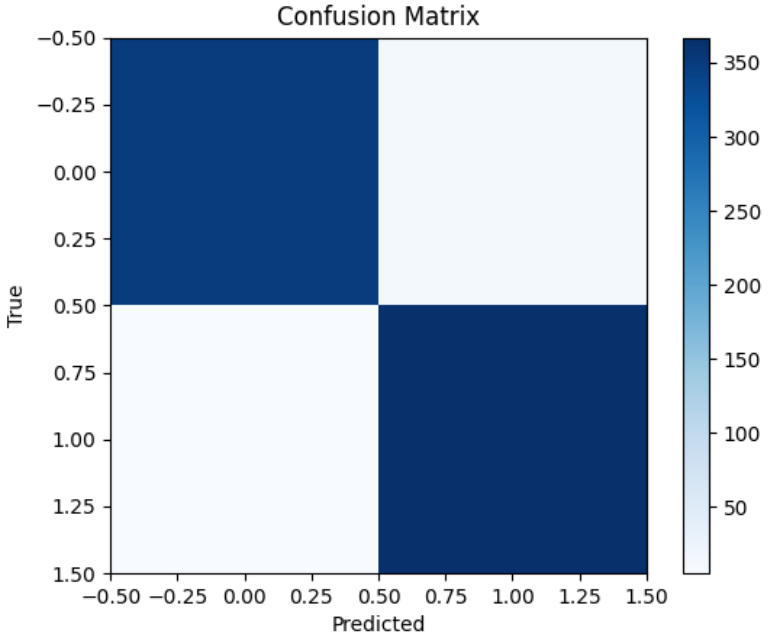

Figure 49 MambaVision Tiny Confusion Matrix for Binary Classification

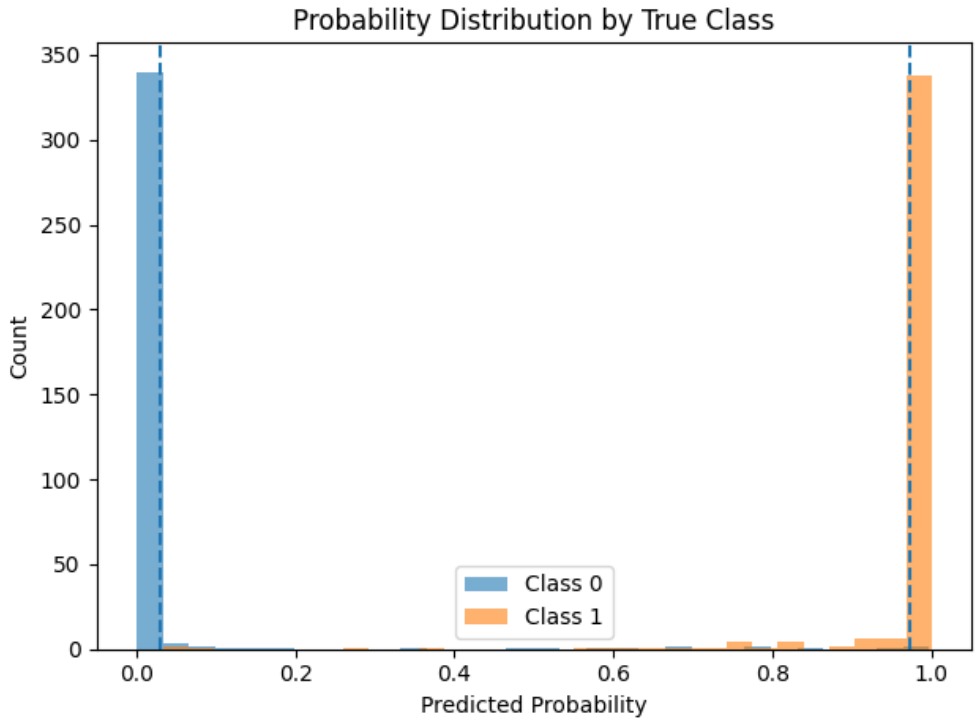

Figure 50 MambaVision Tiny Probability Histogram

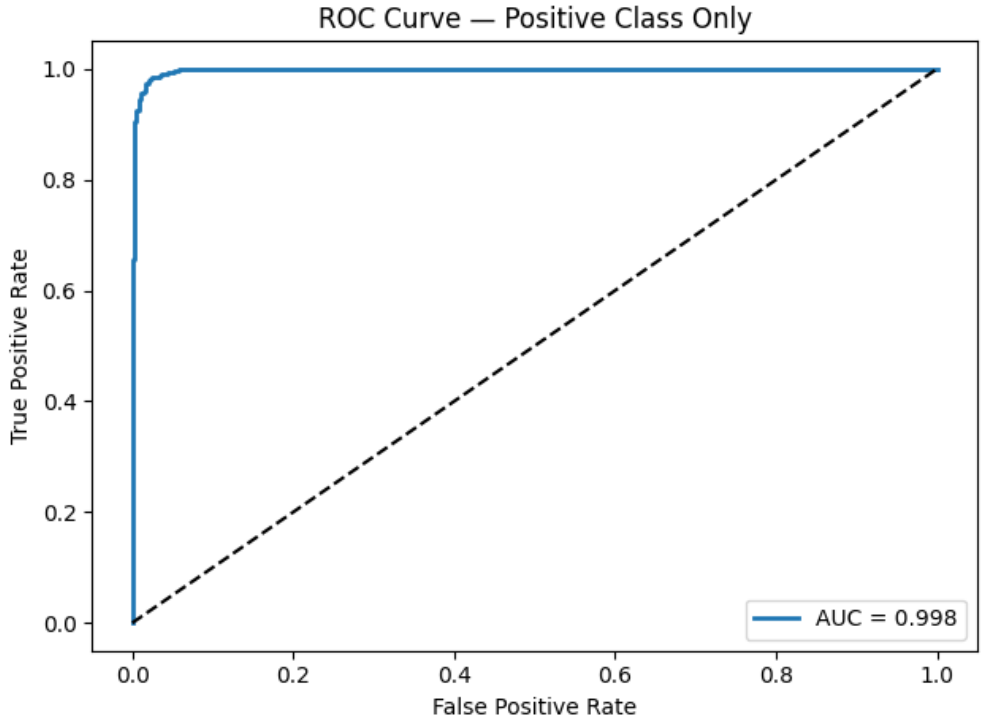

Figure 51 MambaVision Tiny ROC Curve for Binary Classification

4.2.2. Multiclass Classification

4.2.2.1. Swin Transformer Tiny

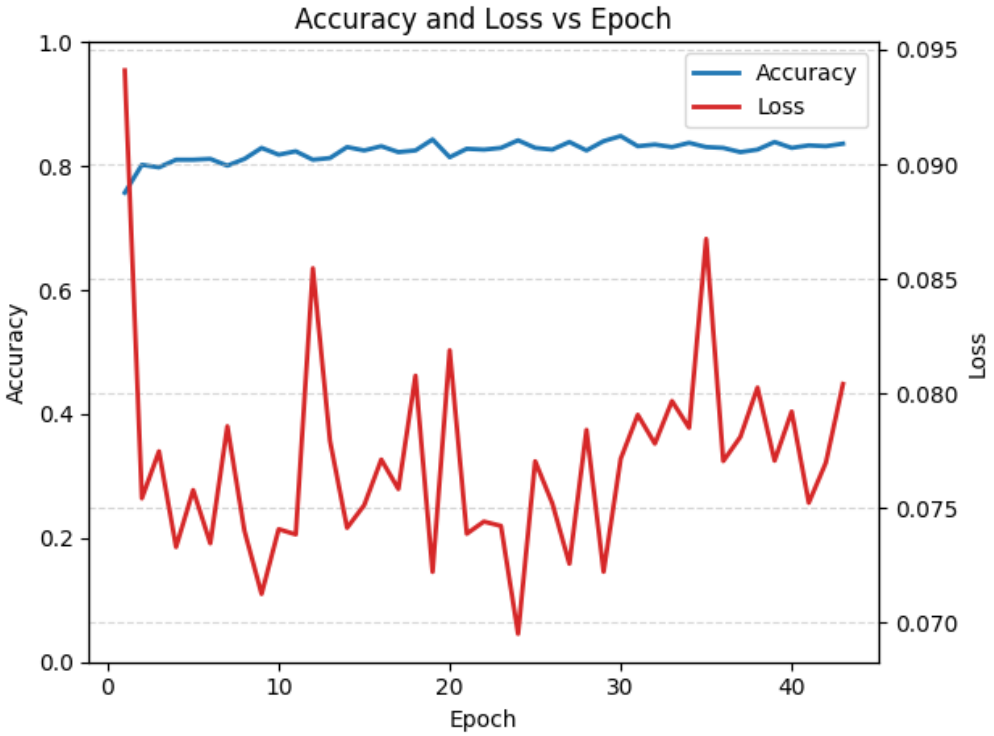

Figure 52 Swin Tiny Accuracy and Error vs Epoch for Multiclass Classification

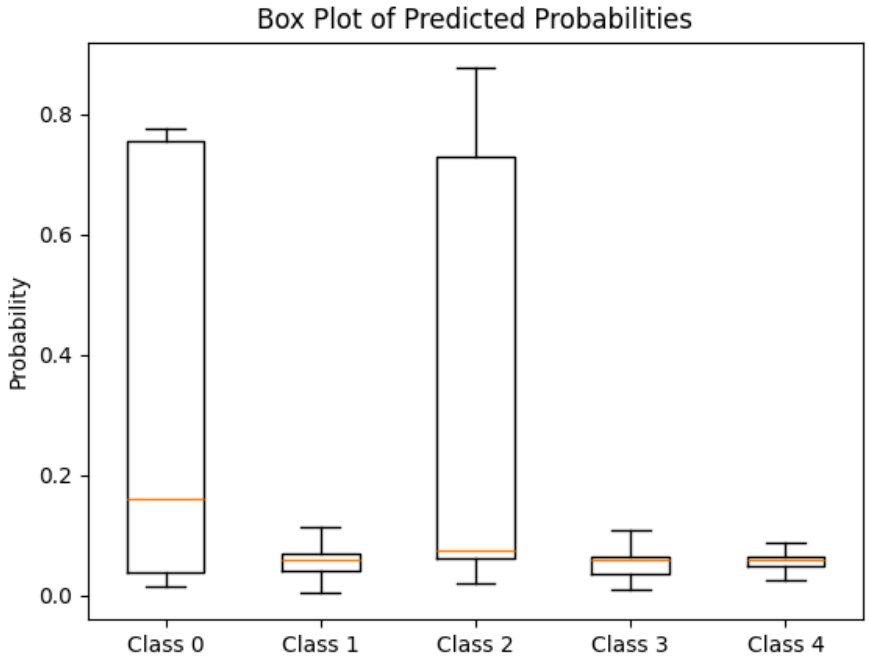

Figure 53 Swin Tiny Box Plot of Predicted Probabilities

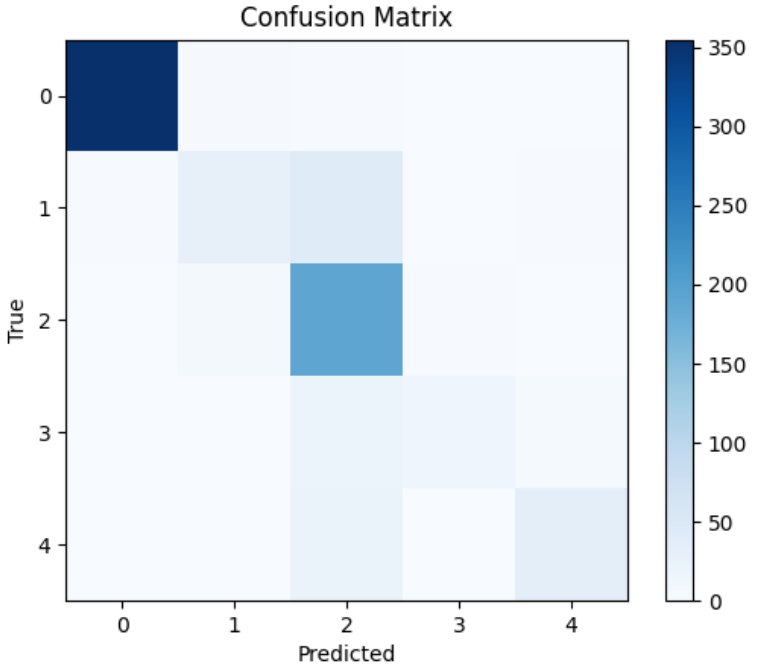

Figure 54 Swin Tiny Confusion Matrix for Multiclass Classification

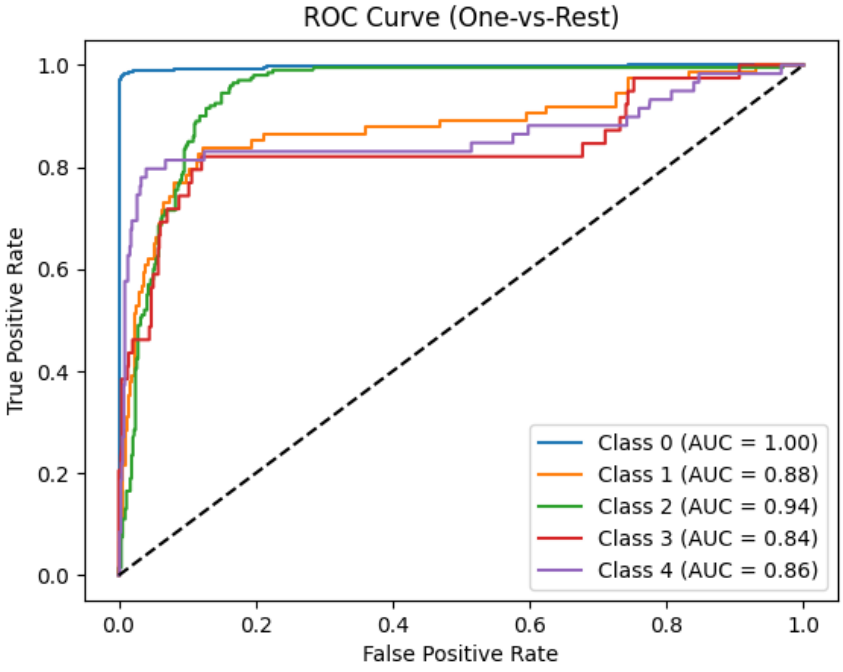

Figure 55 Swin Tiny ROC Curve for Multiclass Classification

4.2.2.2. ConvNeXt Tiny

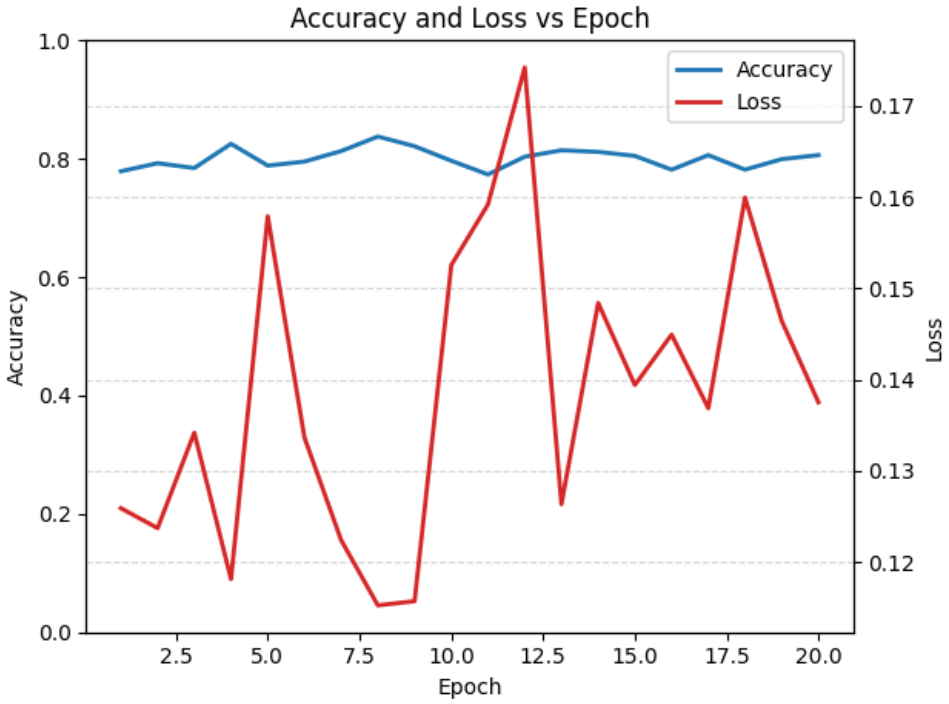

Figure 56 ConvNeXt Tiny Accuracy and Loss vs Epoch for Multiclass Classification

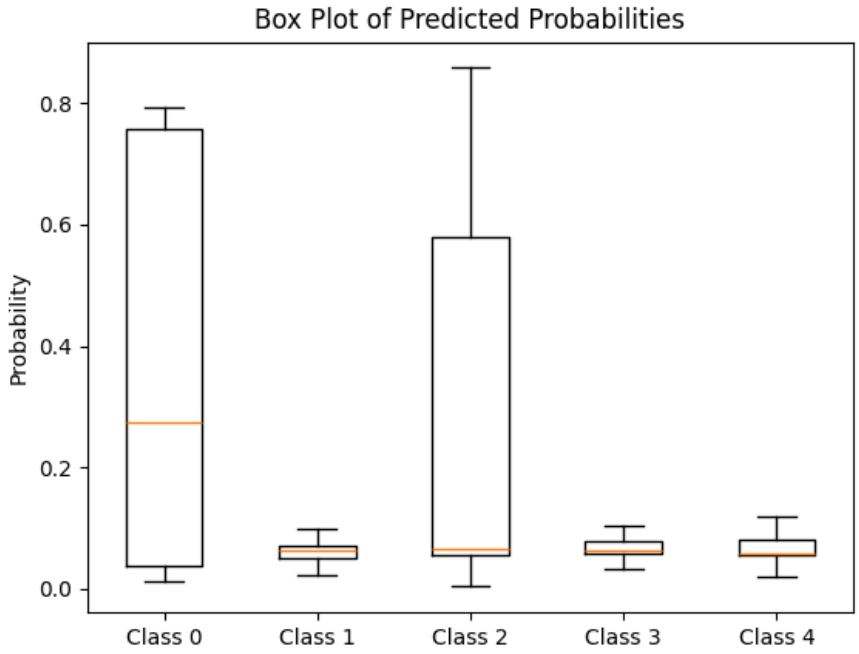

Figure 57 ConvNeXt Tiny Box Plot of Predicted Probabilities

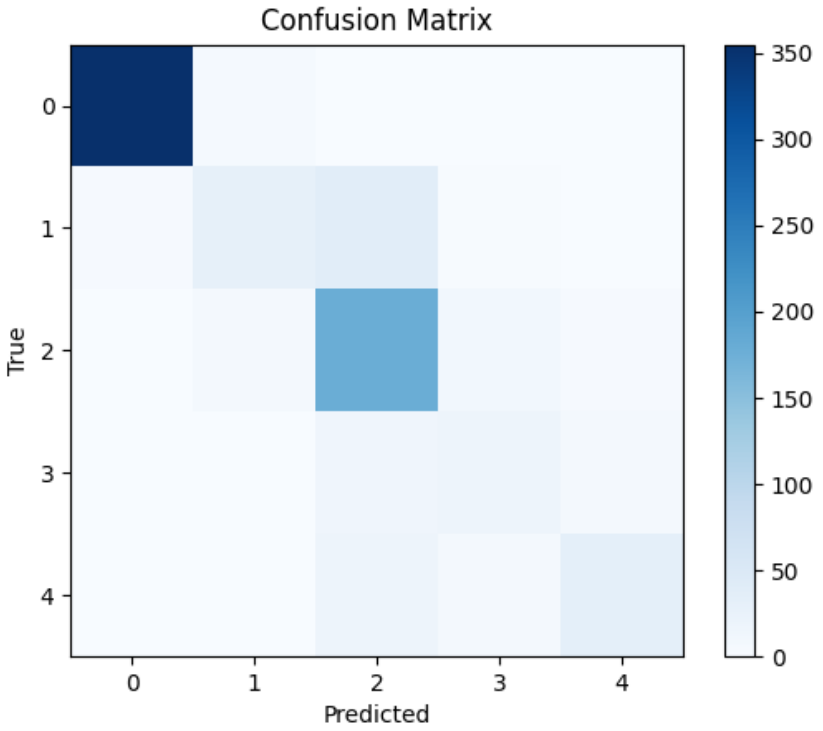

Figure 58 ConvNeXt Tiny Confusion Matrix for Multiclass Classification

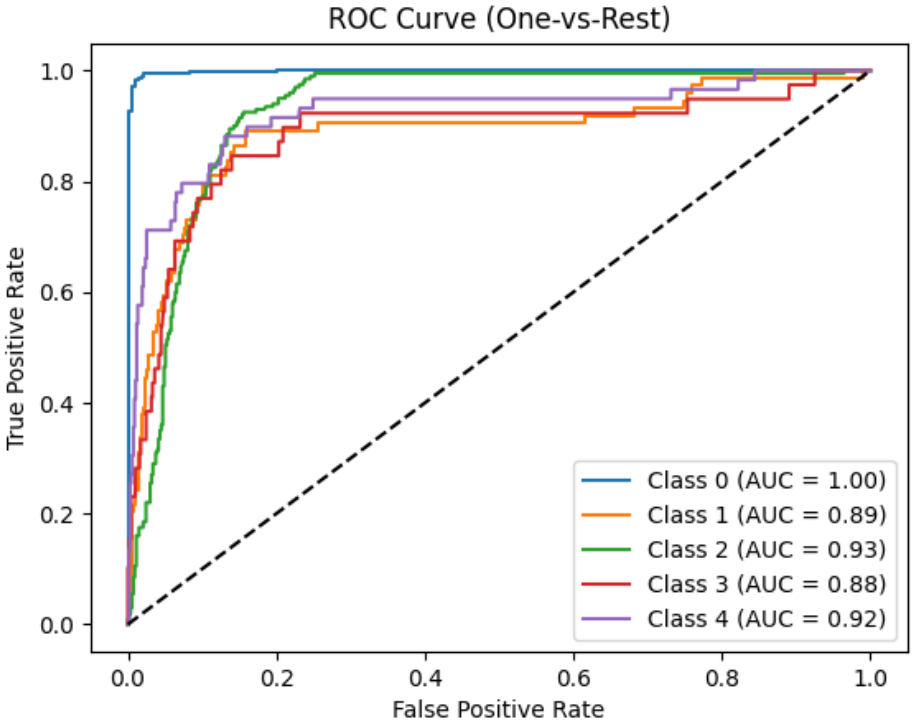

Figure 59 ConvNeXt Tiny ROC Curve for Multiclass Classification

4.2.2.3. Vision Transformer Tiny

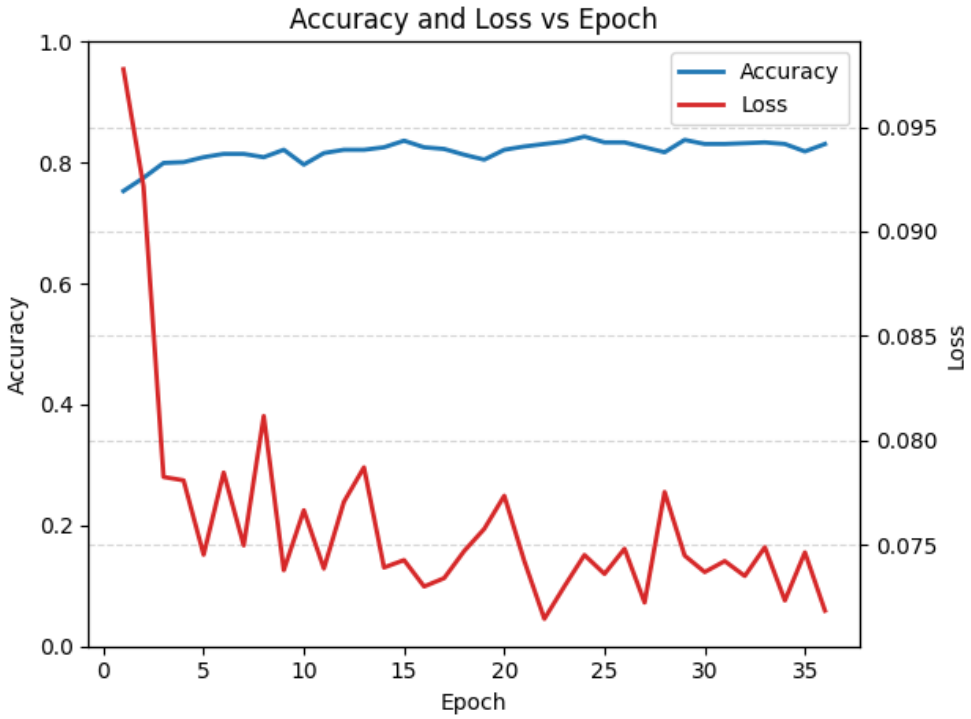

Figure 60 VIT Tiny Accuracy and Loss vs Epoch for Multiclass Classification

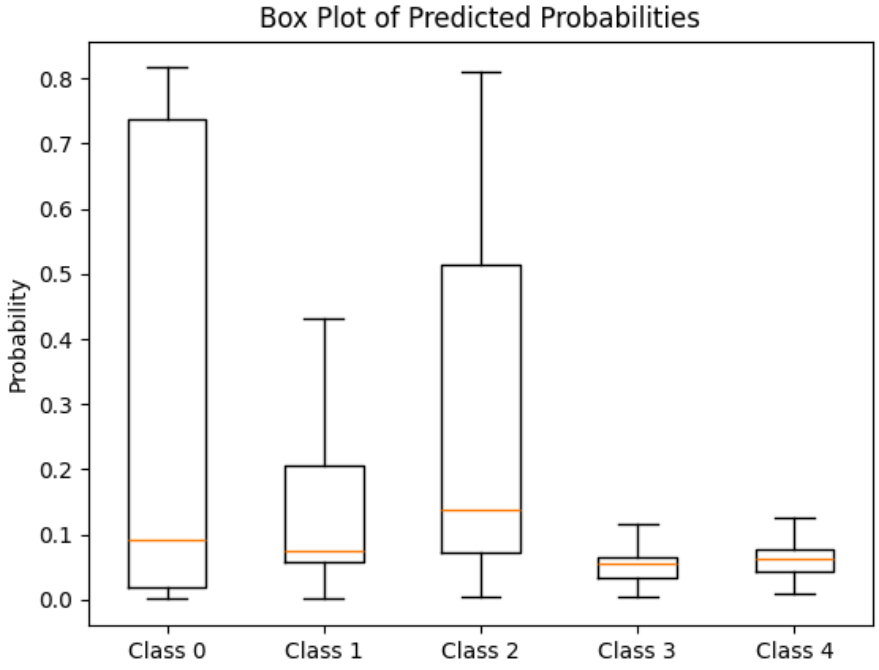

Figure 61 VIT Tiny Box Plot of Predicted Probabilities

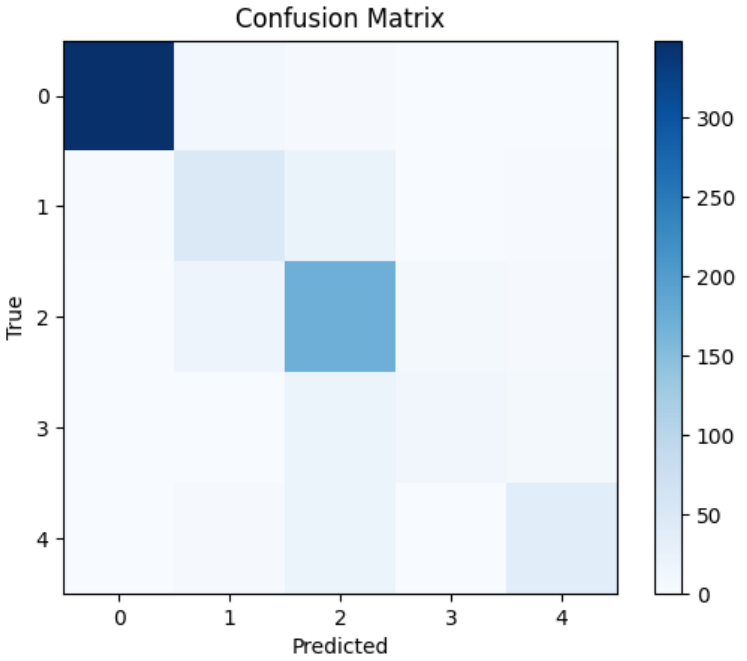

Figure 62 VIT Tiny Confusion Matrix for Multiclass Classification

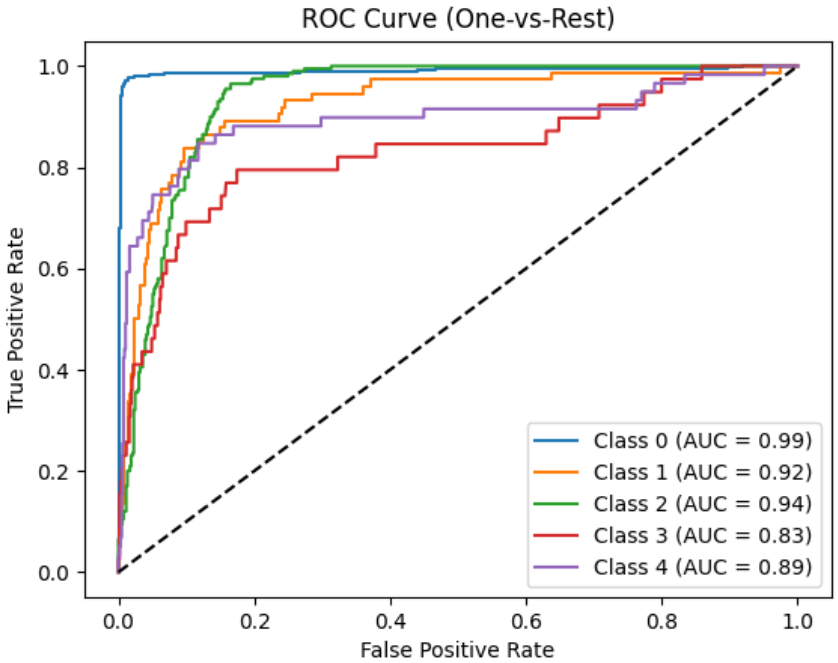

Figure 63 VIT Tiny ROC Curve for Multiclass Classification

4.2.2.4. MambaVision Tiny

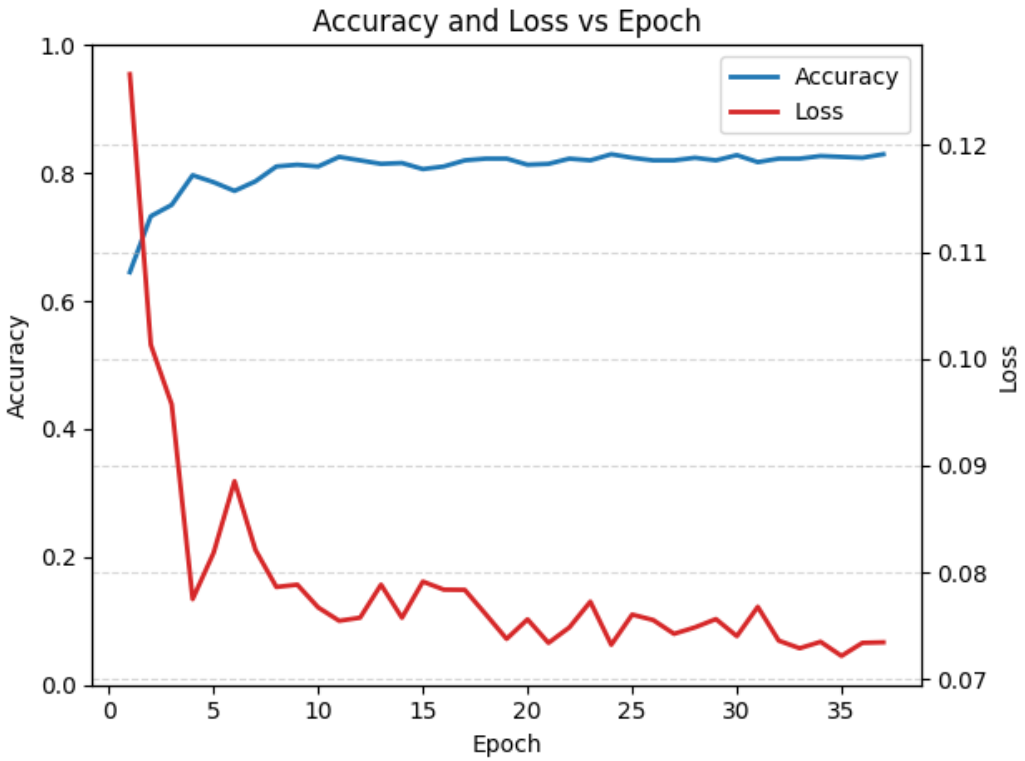

Figure 64 MambaVision Accuracy and Loss vs Epoch for Multiclass Classification

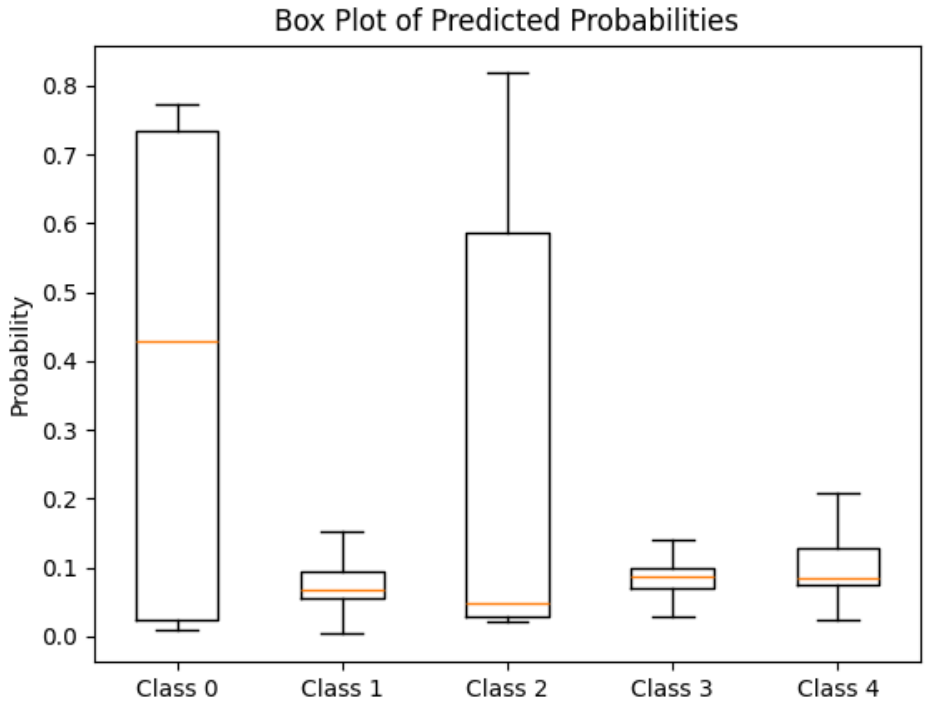

Figure 65 MambaVision Tiny Probability Box Plot

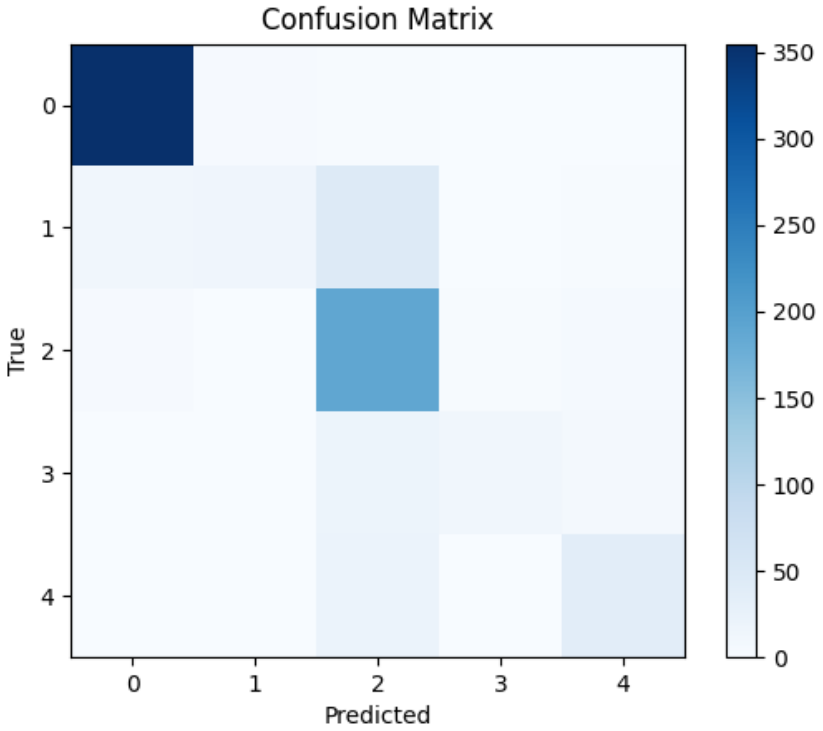

Figure 66 MambaVision Tiny Confusion Matrix for Multiclass Classification

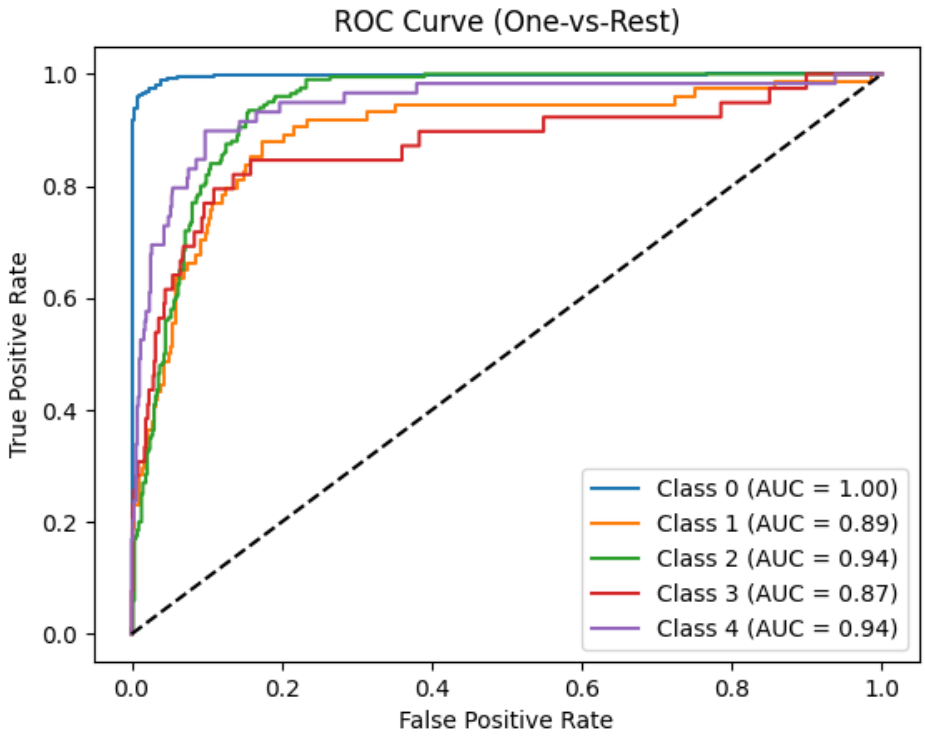

Figure 67 MambaVision Tiny ROC Curve for Multiclass Classification

Chapter 5: Result Discussion

5.1. Overview

The experiments done in this study evaluated the performance of 4 modern DL architectures-Swin Transformer, ConvNeXt, Vision Transformer and MambaVision- in both DR detection as well as grading. Overall, the models achieved high performance in both settings, especially in binary classification where evaluation metrics reached near ceiling level. The best results were achieved by the ConvNeXt algorithm which outperformed the 3 others in detection efficacy and overall generalization.

Regarding DR grading, the models experienced an expected drop in classification accuracy when moving from binary to 5 class grading. However, all algorithms achieved a competitive qwk score (>0.9) which supports their potential use in real world diagnosis and grading applications, with ConvNeXt once again excelling through a qwk of 91.2.

All in all, the results show robust algorithm performance and generalization for both DR screening and grading, once again proving the effectiveness of DL architectures in medical image classification. The dominance of ConvNeXt across both settings may suggest that transformer-inspired CNNs are best suited for generalization tasks with limited data.

5.2. Binary Result Interpretation

In binary classification, all models demonstrated high diagnostic performance, approaching the upper limits of their evaluation metrics. Across all tested algorithms, the confusion matrices showed almost no misclassified results, which is in line with the metrics obtained during quantitative evaluation. Similarly, all ROC curves showed a right angle in the upper left corner with an AUC greater than 0.99 in all cases, indicating that each of the algorithms tested was able to reach a near-perfect classification threshold.

Moreover, the networks exhibited similar accuracy and loss curves, with a slight increase in accuracy at the beginning of the training process followed by a plateau which is typical when applying large architectures to relatively small datasets such as the APTOS 2019. The initial rise in accuracy with epoch was accompanied by an appropriate sharp decline in validation loss in all cases. Furthermore, probability histograms were constructed using probabilities generated by the sigmoid classification function used in all models. Overall, the 4 histograms showed clustering near 0 and 1, which could be interpreted as a sign of overfitting.

Despite having very comparable results, the models still exhibited some variance in training behavior and their ability to generalize from limited datasets.

ConvNeXt achieved the strongest overall performance, tying with the Swin transformer for the highest accuracy (99.18), and scoring the highest specificity, precision and consequently F1-score (99.17%, 99.19%, and 99.19%). However, the model also required the highest total training time (36.74 minutes) as well as the highest average time per epoch (122.46 seconds) highlighting the tradeoff between performance and training length.

The Swin transformer also achieved remarkable results, most notably a sensitivity of 100% which although indicative of its efficacy, could be attributed to the limited size of the dataset used in training. VIT closely trailed behind in terms of accuracy (98.36%) and other relevant metrics. Notably, VIT required 15 epochs before reaching its best, suggesting that vision transformers may benefit from longer training patience. This is further supported by the validation and training loss at its best epoch (both 0.06), which although negligible relative to classification accuracy, are still the highest among all models.

Finally, MambaVision, despite being the fastest to converge with a total training time of 25.99 minutes, also scored the lowest across all evaluation metrics. This may suggest that hybrid state-space-transformer models are not the most optimal option for medical image classification. Alternatively, such models may require larger datasets to provide performance comparable to the other algorithms.

5.3. Multiclass Result Interpretation

All models suffered a considerable drop in performance when moving to 5 class DR grading. Although training settings attempted to account for dataset imbalance, its effect was clearly present in both qualitative and quantitative evaluation. Most notably, the average classification accuracy across the 4 models dropped from 98.36% in binary classification to 83.96% in multiclass classification. Similarly, the average F1-score dropped from 98.66% to 67.86%.

Images labeled as “Moderate DR” comprised 27.28% of the total dataset, which is greater than the number of images of the other 3 DR positive classes combined. Consequently, the confusion matrices of the networks show that for each model, when DR is detected, its severity is skewed towards the moderate class (class 2). However, the matrices show that misclassification often happened at a distance of one class, which is in line with the qwk scores

achieved by the algorithms. This follows from the fact that qwk heavily penalizes misclassifications at higher distances. Moreover, the models rarely classified a DR positive case as negative which matches their performance in the binary setting.

Accuracy and loss trends differed significantly in the multiclass setting as compared to binary. The most obvious difference was the oscillations present in both validation accuracy and loss per epoch curves, which reflects the model's uncertainty due to the increased number of classes. This is further supported by the variance in the SoftMax probabilities in the box plots of each model. It follows that both training and validation loss at each model's best epoch were significantly greater in multiclass classification. This becomes apparent when comparing the average binary training loss at best epoch across the 4 algorithms to its multiclass counterpart ($0.037<0.272$). Moreover, the networks required more epochs to converge due to the complexity of the task. On average, models performing binary classification required about half the number of epochs required by their multiclass counterparts to reach their best performance (10 and 21.5 respectively).

A clear outlier in training behavior was none other than ConvNeXt, which despite achieving the highest qwk score, had its best epoch shockingly early at epoch number 8. To put this into perspective, the next closest were VIT and MambaVision tied at 24 epochs. Additionally, its training loss at its best epoch was significantly greater than the other algorithms (0.45 with the next closest being 0.25). This reflects the ability of CNNs to converge faster than transformer-based models, especially in tasks involving smaller datasets. Although there was potential to minimize the loss even further, this might have negatively impacted the qwk score and led to overfitting. This peculiarity in training outcomes highlights the importance of choosing appropriate evaluation metrics for the classification task.

Furthermore, ROC curves were constructed to study the class specific behavior of each network. Although the curves themselves may seem messy and convoluted, a valuable metric is the AUC score for each class. All algorithms achieved a near perfect AUC score for class 0, which comprised half of the dataset. The models struggled most notably with the underrepresented classes 3 and 4. The highest AUC score for class 3 was achieved by ConvNeXt (0.88), as well as a formidable 0.92 for class 4, which correspond directly to its high qwk. Interestingly, MambaVision had the highest AUC for class 4 (0.94) and closely trailed behind ConvNeXt in its class 3 AUC (0.87). This suggests that Mamba's architecture may be better suited for more challenging classification tasks.

Overall, although there is no denying that the models struggled after transferring to the task of DR grading, the qwk scores reflect the competitive performance demonstrated by each of the models and supports their use in medical grade applications. ConvNeXt showed that modern CNNs are better at generalizing than transformers, especially when finetuned on smaller datasets. This is why ConvNeXt was selected as the algorithm to be used in our DR grading framework.

5.4. Comparison to Literature

Research studies in DR detection have reported a wide spectrum of performance metrics, depending on the dataset type and complexity. Many systems in the literature that rely on classical handcrafted features combined with machine learning classifiers such as SVM or Random Forest have achieved very high reported accuracy in binary DR detection, in some cases up to 99% and even 99.7% [31], [38]. However, these studies were mainly trained on very small datasets such as DIARETDB1 and STARE, which include fewer images, less noise, and lower variability. This makes the task fundamentally easier and often inflates sensitivity, specificity, and overall accuracy due to limited generalization challenges.

In contrast, our models have been trained on more realistic screening conditions in the APTOS dataset, including brightness variations, image blur, and occlusions. Despite that, our results in the binary setting were still highly competitive and, in various dimensions, better than benchmarks in literature. Across accuracy, sensitivity, specificity, precision and F1 score, all our models scored higher than 96%, placing them above or in line with the highest reported in similar binary classification studies. Most notably, both the Swin transformer and ConvNeXt achieved an accuracy of 99.18% with the Swin transformer scoring a sensitivity of 100%, meaning that across over 700 in the testing set, the algorithm did not miss a single DR case.

When extending to full multiclass grading, which is significantly more clinically meaningful but far more challenging, our results continued to reflect realistic performance trends described in literature. While some advanced methods have reported 95-99% accuracy [38] for severity classification, these studies often collapse the number of classes or use well-balanced datasets with simplified lesion differentiation. In our case, with five classes (1 normal, 4 DR) and a highly imbalanced dataset, our models achieved a very competitive top accuracy of 84.85%, with most misclassifications occurring on minority classes. These results confirm that

subtle lesions, especially microaneurysms in early stages, remain the primary source of misclassification, and dataset imbalance has a major effect on performance degradation. Thus, our performance fully agrees with reported challenges in real DR severity grading scenarios and demonstrates strong capability given the dataset complexity. Our results fall into the realistic and scientifically accepted performance range for true five-stage classification systems in multiclass grading. That means our models are not only powerful but also trained under conditions that reflect real-world DR screening variability better compared to several studies that report higher results under more constrained or less clinically challenging conditions.

To account for minor misclassifications, the standard in evaluating ML performance in medical multiclassification tasks has become the QWK. This metric, which is meant to describe the agreement between 2 raters quantitatively, punishes misclassifications according to their ordinal distance. Using the same dataset as the one used in our training, Kaggle hosted an open blindness detection challenge in 2019 where users could submit their solutions for a chance to win a cash prize. The goal was the creation of a framework capable of high performance on DR grading tasks, and for that QWK was used as the defining metric. Although the top performance managed to score a QWK of 0.93, their framework was pre-trained on multiple datasets before being tested on APTOS and their algorithm was an ensemble [52]. In contrast, our framework managed to achieve a QWK of 0.92 solely relying on one model (ConvNeXt) and without any datasets besides APTOS.

When comparing our results to the 2023 systematic review and meta-analysis on AI performance in diabetic retinopathy screening published in Frontiers in Endocrinology [53], it is apparent that our developed models achieve performance levels superior to the pooled estimates reported across multiple prospective studies. The meta-analysis, which synthesized evidence from diverse clinical environments and imaging systems, reported a pooled sensitivity of approximately 0.88 and pooled specificity of around 0.91 for AI-based DR screening systems. In contrast, our best-performing architecture, the Swin Transformer Tiny, achieved 100% sensitivity and 98.34% specificity, with an overall binary accuracy of 99.18%, while ConvNeXt similarly showed very strong performance with high precision and F1-scores. Even the lower-ranking models in our study-maintained sensitivity values above 98% and specificity above 96%, clearly outperforming the average levels reported in the meta-analysis. The results remain competitive even when comparing them to scores achieved by subgroups of the meta-analysis. The CNN subgroup achieved a pooled sensitivity and specificity of 0.95, and 0.92

respectively and the high-quality study subgroup scored 0.93 and 0.9, which are scores clearly lower than our metrics. That suggests that, within the controlled setting of the APTOS dataset, our transformer-based models can detect DR with substantially fewer false negatives and false positives than the aggregated AI systems evaluated in clinical practice. However, the available results of the meta-analysis represent the pooling of performance across different patient populations, camera types, and test conditions, while our results are based on a single public dataset. Thus, while our models outperform the pooled benchmarks for diagnostic metrics, external validation on multi-center datasets would be required to confirm whether such superior performance generalizes to real-world screening conditions.

5.5. Limitations

Despite very promising results, some limitations need to be addressed to give a realistic estimate of the robustness of the system and its potential for clinical deployment. First, the most significant limitation is that the dataset size of APTOS which is relatively small compared to public datasets like EyePACS, which contain tens of thousands of labeled images. Since the severe and proliferative DR cases are under-represented, insufficient training exposure to those critical stages makes the model prone to misclassification during multiclass grading. Moreover, the dataset has been collected from a single source, introducing limited variations in imaging equipment, demographics, and clinical settings, which reduces its external validity and increases the risk of distribution shift when applying it to real hospital settings.

Another limitation arises from the one-time training protocol adopted in this work due to restricted training time and computational resources. Ideally, DL models should undergo multiple training trials with extensive hyperparameter tuning, cross-validation, and LR optimization to ensure performance is stable, consistent, and not contingent on a single training run. Although training parameters were varied for each algorithm tested in the framework, and different pre-processing pipelines were employed, optimizing the results from each of the algorithms requires more extensive probing and testing. In other words, although our metrics are already strong, further improvement and optimization are likely possible.

Additionally, the system relies exclusively on CFP without making use of important clinical biomarkers like fasting glucose levels, disease duration, or symptoms of the patients that may provide better disease staging. Including any type of additional information such as

the patient's age, gender, or type of diabetes may help considerably boost classification accuracy. Furthermore, CFP is not universally used for DR diagnosis. Local clinics may rely on OCT or OCTA images in order to assess DR severity and track its progression.

Finally, though sensitivity values can be excellent in binary classification, particularly for Swin Transformer, real-world medical screening requires external testing and prospective studies in primary care or ophthalmology settings to establish clinical safety and reliability. In other words the system needs to be tested extensively in a real world setting before ensuring that it's safe for medical teams to rely on.

Despite these limitations, the framework demonstrates a highly competitive performance foundation, showing that advanced transformer-based architectures can support reliable and accurate DR screening systems when trained on suitable retinal datasets.

5.6. GUI Integration

In order to extend the accessibility of our findings to the medical staff, a GUI was designed under the name "eyeCare" to go along with our DL classifier after concluding training. Developed in MATLAB and interfaced with python, the GUI provides an intuitive workflow for non-technicals to perform automated DR screening using our best DR grading model, ConvNeXt Tiny.

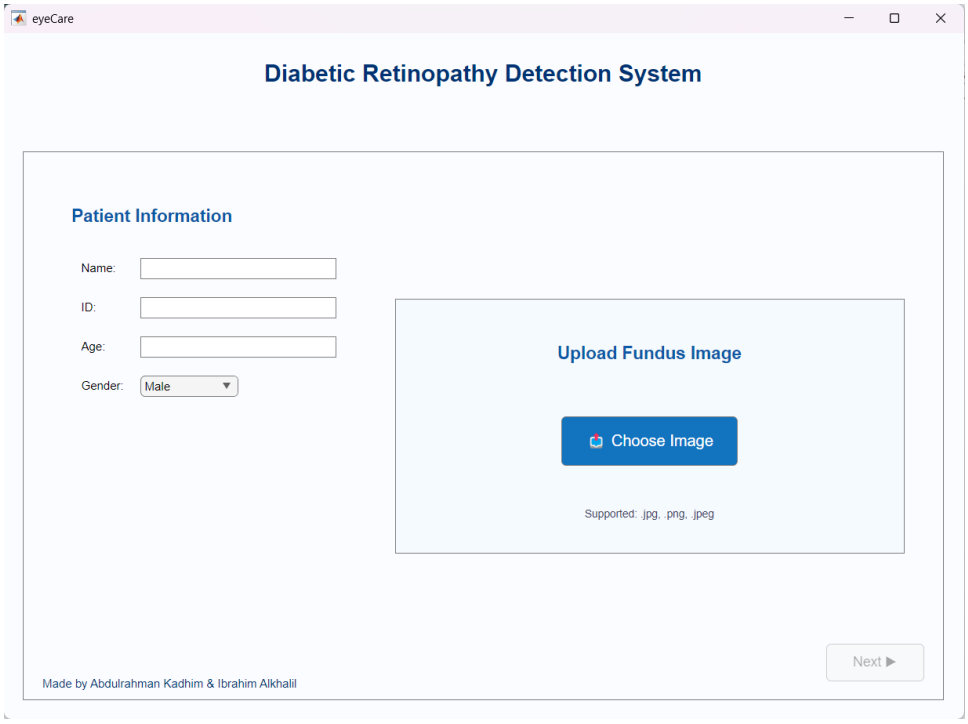

Figure 68 GUI first page

Consisting of 3 pages, its first page collects input about basic patient information (Name, Age, ID, Gender) from the user and offers a clear upload box for the user to select a local color fundus image for analysis. The system accepts a multitude of image formats including jpg, png, jpeg. After uploading an image, a context sensitive “Next” button is used to advance to the second page.

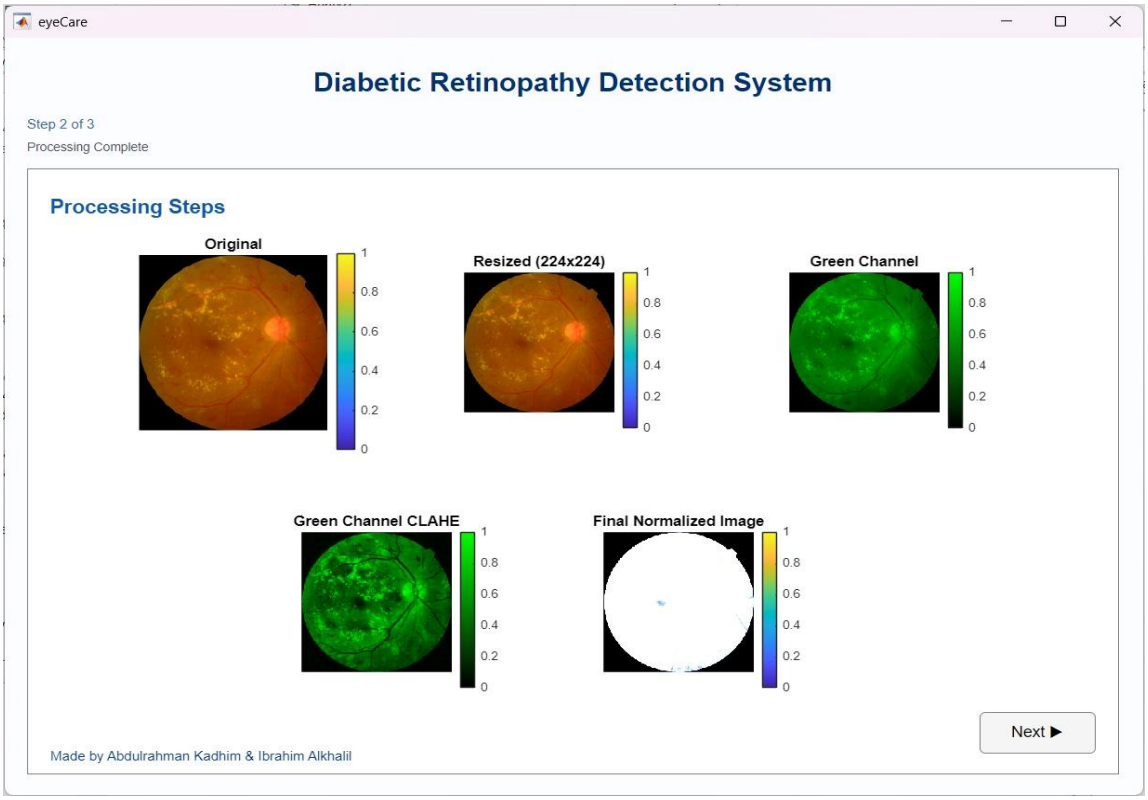

Figure 69 GUI second page

The second page visualizes the pre-processing pipeline that is applied to the image uploaded. The processing figures: Original Image, Resized Image to (224x224), Green channel and Green-channel CLAHE and Final Normalized Image, are shown side by side with the color bars so the operator can inspect intermediate results. Resizing changes the dimensions of the uploaded image to those expected by the algorithm. Extraction of the green channel emphasizes vasculature and hemorrhagic lesions, because the green channel typically offers the highest contrast for blood-related features in RGB fundus photography. CLAHE on the green channel improves local contrast and enhances small lesions like (microaneurysms and exudates), increasing their visibility to the classifier. The final normalization step includes normalizing each channel of the image to a specified standard mean and deviation. This helps scale intensities so

the model receives a uniform input distribution; this step also reduces the influence of illumination differences and camera variability, improving model robustness.

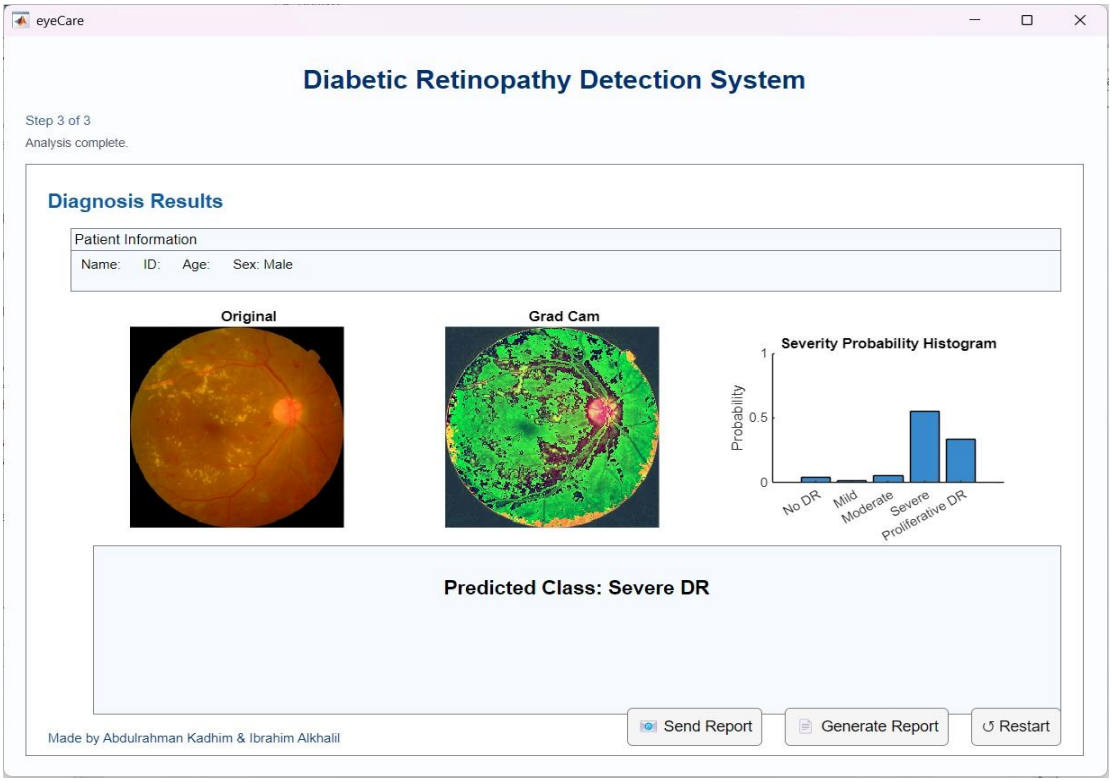

Figure 70 GUI third page

The last page shows diagnosis outputs and interpretation tools. First, the original fundus image is displayed for reference. Besides this image, there's a Grad-CAM Heat Map that highlights the regions that influenced the model's gradient the most during classification, allowing the user to understand its decision processing and boosting trust between machine and operator. Furthermore, the right side of the results page displays a Severity Probability Histogram. This histogram illustrates probability values from 0-1 assigned to each of the 5 severity classes by the SoftMax activation function of the classifier. These classes include No DR, and then sequentially Moderate, Severe, and Proliferative DR. This also aids in getting the practitioner to trust classification results. A text box occupies space on the display area indicating the predicted result with an accompanying message, for instance 'Predicted Class: Severe DR.' The GUI also contains interactive and actionable buttons at the end. These buttons include 'Restart' and 'Generate Report.' 'Restart' prompts a button click for diagnosis on a new patient. 'Generate Report' allows the user to automatically generate a report that is saved locally with the click of a button. Lastly, there's 'Send Report' which after automatically generating a patient

report, prompts the user to enter an email address for a recipient. Upon entering an email address, the GUI makes use of a custom python module to automatically send the report to the specified address, and a message at the top left of the page provides confirmation that the email's been sent.

Finally, with respect to implementation and deployment, GUI functionality maintains a good level of transparency, medical usefulness, and simplicity. While displaying preprocessing steps and Grad-CAM outputs enhances transparency and debugging; viewing the probability histogram requires careful consideration in indeterminate instances; and report generation/sending functionality streamlines medical record-keeping and tele consults. To deploy our model for medical uses, we propose incorporating an image-quality check flagging blurs and under/over exposures, cross-validation or ensemble model selection, and a simple logging file documenting predictions, preprocessing variables, and timestamps. Moreover, before applying our GUI model for medical purposes, it should be shown prospectively with healthcare professionals for intervening and ensuring patient confidentiality provisions for report transmission and storage.

Chapter 6: Conclusion

This study effectively illustrates the use of cutting-edge ML and DL techniques to create a reliable and intelligent system for the early detection and classification of DR. The proposed system accurately identifies several phases of DR by using state-of-the-art models such as ConvNeXt, Vision Mamba, Vision Transformers, and Swin Transformers with thorough image pre-processing and enhancement procedures. The system's usefulness is further enhanced by its intuitive graphical user interface and telemedicine module, which make it feasible for real-world deployment, particularly in environments with low resources and remote locations. The suggested models' dependability and effectiveness were validated by a comparative study of algorithm performance. By enhancing early detection, reducing the risk of blindness, and assisting medical practitioners in prompt intervention, this initiative advances AI-assisted ophthalmic diagnostics.

DR remains one of the leading causes of blindness worldwide; early detection is imperative to prevent irreversible vision loss. The need for the development of accessible, scalable, and accurate screening tools has grown, so the present work is focused on the development and evaluation of an automated system for DR detection and its classification in terms of severity by means of advanced DL architectures. Using the APTOS 2019 Blindness Detection dataset and the implementation of state-of-the-art algorithms, this report has proved that modern transformer-based architectures have the capability to enhance the detection of DR using only color fundus images.

The results obtained in this project illustrate that transformer-based models can achieve extremely high performance in binary DR classification. The best performances were by the Swin Transformer Tiny and ConvNeXt Tiny, which had an accuracy of 99.18%, very high precision and F1-scores, and sensitivity of 100% for Swin Transformer. Even models with somewhat poorer performance, like ViT Tiny and Vision Mamba Tiny, reached accuracy values higher than 97%, again proving their strong capability of distinguishing between healthy and diseased retinas. These results confirm that DL models, especially transformer-based ones, achieve high levels of reliability in medical classification tasks, supporting the argument for their use in DR diagnosis systems, especially in the preliminary screening phase.

With the increase in difficulty and significant class imbalance, a natural consequence of the multiclass classification task of the system (which needs to distinguish between five severity levels of DR), performance obviously decreased. However, algorithms were able to remain competitive, with Swin Transformer reaching an accuracy of 84.85%, followed closely by ViT, ConvNeXt, and Vision Mamba. These results are expected, as it is normally difficult to distinguish between Mild and Moderate DR cases, especially due to the very subtle differences and limited representation of Severe and Proliferative DR grades. Therefore, considering the complexity of the dataset used and the adoption of a single-training strategy within this project, the performance obtained for multi-class grading was still very good.

Further comparison with the literature emphasizes the strength of the developed system. Benchmarked against previous studies based on APTOS and the results of the APTOS 2019 competition, models developed in this project performed competitively or even better, especially on binary classification tasks. Compared to the results from large meta-analyses, which pooled sensitivities of around 0.88 and specificities around 0.91 across a wide range of real-world settings, our models outperformed these averages: higher sensitivity, higher specificity, and higher overall accuracy. This suggests that the proposed system can operate at least at the same level as, or even much better than, many existing AI-based DR screening tools reported in academic research. However, it is also important to consider that the benchmarks of meta-analyses are based on a wide range of datasets, equipment types, and patient populations, while our evaluation here is bound within a single dataset. Despite these promising results, there are some limitations to this study. The APTOS dataset is relatively small, and class imbalance exists, especially in the advanced stages of DR. Because of the computational resources available, the models have been trained only once without any cross-validation or extensive hyperparameter tuning. Besides, the system has been developed and tested on only one dataset and was not externally validated on datasets such as Messidor, EyePACS, or IDRiD. Therefore, even though the performance is remarkable under controlled conditions, further testing is still needed to confirm real-world generalizability.

Finally, the models rely only on the fundus images and do not take into consideration clinical data that could increase diagnostic accuracy in complicated cases. The project successfully designed and evaluated four state-of-the-art deep learning models for diabetic retinopathy detection and grading, with evidence that transformer-based architectures are among the most powerful and reliable solutions for medical image analysis. The strong binary classification

performance, competitive multiclass results, and consistency with state-of-the-art benchmarks underline the probable impact of these models in supporting automated DR screening applications. With future iteration involving external validation, repeated training, diverse datasets, explainability methods, and integration into telemedicine environments, the system designed and proposed in this work could contribute even more to early detection workflows and help decrease unnecessary vision loss from the complication of DR. Therefore, this work has been an important step towards the wider implementation of AI-assisted ophthalmic screening and grading both in clinical and in remote healthcare.

References

[1] "Diabetes-Related Retinopathy," 26 2 2024. [Online]. Available: https://my.clevelandclinic.org/health/diseases/8591-diabetic-retinopathy.

[2] M. Sugue, "Diabetic Retinopathy Statistics," 23 7 2024. [Online]. Available: https://www.visioncenter.org/resources/diabetic-retinopathy-statistics.

[3] R. L. Thomas, "IDF Diabetes Atlas: A review of studies utilising retinal photography on the global prevalence of diabetes related retinopathy between 2015 and 2018," 14 November 2019. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/31733978/.

[4] N. Waked, "Epidemiology of diabetic retinopathy in Lebanon," March 2006. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/16557173/.

[5] B. VanderBeek, "Twenty year trends in prevalence and incidence of diabetic retinal disease," 29 January 2025. [Online]. Available: https://www.aaojournal.org/article/S0161-6420%2825%2900076-4/fulltext.

[6] "Eye Anatomy," [Online]. Available: https://www.themedicaleyecenter.com/eye-conditions-manchester/eye-anatomy/.

[7] S. Standring, "The Eye," in *Gray's Anatomy: The Anatomical Basis of Clinical Practice, 42nd ed.* , Philadelphia, Elsevier, 2021.

[8] "What a Typical Retina Looks Like vs. Someone with Diabetic Retinopathy," 26 September 2023. [Online]. Available: https://www.healthline.com/health/diabetes/normal-retina-vs-diabetic-retinopathy.

[9] M. Christina J. Flaxel, "Diabetic Retinopathy Preferred Practice Pattern," 1 2020. [Online]. Available: https://www.aaojournal.org/article/S0161-6420(19)32092-5/fulltext.

[10] G. B. Usharani bhimavarapu, "PUBMED central," 28 december 2022. [Online]. Available: https://doi.org/10.3390/healthcare11010097.

[11] A. J. w. David A salz, "National library of medicine," 2015. [Online]. Available: https://doi.org/10.4103/0974-9233.151887.

[12] M. Karmel, "Retinal Imaging: Choosing the Right Method," 1 July 2014. [Online]. Available: https://www.aao.org/eyenet/article/retinal-imaging-choosing-right-method.

[13] J. S. R. T. B. W. D. Y. X. L. X. T. D. S. W. S. D. A. M. G. G. &. S. Chua, "Pubmed central," national library of medicine , 2020. [Online]. Available: https://doi.org/10.3390/jcm9061723.

[14] "Fundus Photography," 23 July 2024. [Online]. Available: https://my.clevelandclinic.org/health/diagnostics/fundus-photography.

[15] S. NA, "Retinal blood vessel segmentation using a deep learning method based on modified U-NET model," September 2021. [Online]. Available: https://www.researchgate.net/publication/354789945_Retinal_blood_vessel_segmentation_using_a_deep_learning_method_based_on_modified_U-NET_model.

[16] E. Odeh, "Diabetic Retinopathy Detection using Ensemble Machine Learning," June 2021. [Online]. Available: https://www.researchgate.net/publication/353069572_Diabetic_Retinopathy_Detection_using_Ensemble_Machine_Learning/figures?lo=1&utm_source=google&utm_medium=organic.

[17] "Chapter 4 - Fringe noise removal of retinal fundus images using trimming regions," 2015. [Online]. Available:

https://www.sciencedirect.com/science/article/abs/pii/B9780128020456000041.

[18] "Detection of hemorrhage in retinal images using linear classifiers and iterative thresholding approaches based on firefly and particle swarm optimization algorithms," January 2019. [Online]. Available: https://www.researchgate.net/publication/330560522_Detection_of_hemorrhage_in_retinal_images_using_linear_classifiers_and_iterative_thresholding_approaches_based_on_firefly_and_particle_swarm_optimization_algorithms.

[19] "haralick texture , computer vision," word press, 22 july 2020. [Online]. Available: https://cvexplained.wordpress.com/2020/07/22/10-6-haralick-texture/.

[20] "What is artificial intelligence (AI)?," 9 August 2024. [Online]. Available: https://www.ibm.com/think/topics/artificial-intelligence.

[21] "Navigating the Ebb and Flow: Understanding AI Winters and Their Impact on Innovation," 8 August 2024. [Online]. Available: https://www.cognitech.systems/blog/artificial-intelligence/entry/ai-winter-periods.

[22] S. Alowais, "Revolutionizing healthcare: the role of artificial intelligence in clinical practice," 22 September 2023. [Online]. Available: https://bmcmededuc.biomedcentral.com/articles/10.1186/s12909-023-04698-z.

[23] "A Brief Overview of Support Vector Machines (SVM)," [Online]. Available: https://www.iunera.com/kraken/fabric/support-vector-machines-svm/.

[24] "What is Random Forest?," [Online]. Available: https://dida.do/what-is-random-forest.

[25] "What makes a good prediction? Feature importance and beginning to open the black box of machine learning in genetics," [Online]. Available: https://www.researchgate.net/publication/356781515_What_makes_a_good_prediction_Feature_importance_and_beginning_to_open_the_black_box_of_machine_learning_in_genetics.

[26] Y. Kokkinos, "Simulating parallel scalable Probabilistic Neural Networks via Exemplar Selection and EM in a Ring Pipeline," March 2018. [Online]. Available: https://www.researchgate.net/publication/318441940_Simulating_parallel_scalable_Probabilistic_Neural_Networks_via_Exemplar_Selection_and_EM_in_a_Ring_Pipeline.

[27] X. Knag, "A Deep Similarity Metric Method Based on Incomplete Data for Traffic Anomaly Detection in IoT," January 2019. [Online]. Available: https://www.researchgate.net/publication/330106889_A_Deep_Similarity_Metric_Method_Based_on_Incomplete_Data_for_Traffic_Anomaly_Detection_in_IoT.

[28] "Understanding the Mechanism and Types of Recurrent Neural Networks," 15 December 2020. [Online]. Available: https://opendatascience.com/understanding-the-mechanism-and-types-of-recurring-neural-networks/.

[29] "An Overview of Deep Belief Network (DBN) in Deep Learning," 27 May 2024. [Online]. Available: https://www.analyticsvidhya.com/blog/2022/03/an-overview-of-deep-belief-network-dbn-in-deep-learning/.

[30] A. Zaylaa and S. Kourtian, "From Pixels to Diagnosis: Early Detection of Diabetic Retinopathy Using Optical Images and Deep Neural

Networks," 3 March 2025. [Online]. Available: https://www.mdpi.com/2076-3417/15/5/2684.

[31] S. Gayarthi, "Automated binary and multiclassclassification of Diabetic Retinopathyusing Haralick and Multiresolution Features," 10 March 2020. [Online]. Available: https://ieeexplore.ieee.org/document/9031365.

[32] T. Ratanapakorn, "Digital image processing software for diagnosing diabetic retinopathy from fundus photograph," 17 April 2019. [Online]. Available: https://www.dovepress.com/digital-image-processing-software-for-diagnosing-diabetic-retinopathy--peer-reviewed-fulltext-article-OPTH.

[33] C. Suedumrong, "Diabetic Retinopathy Detection Using Convolutional Neural Networks with Background Removal, and Data Augmentation," 30 September 2024. [Online]. Available: https://www.mdpi.com/2076-3417/14/19/8823.

[34] A. Senapati, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2352914824000017?via%3Dihub.

[35] R. Kim, "The use of teleconsultation and technology by the Aravind Eye Care System, India," 7 June 2022. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC9412088/.

[36] S. Pal, "Morphological operations with iterative rotation of structuring elements for segmentation of retinal vessel structures," 5 March 2018. [Online]. Available: https://link.springer.com/article/10.1007/s11045-018-0561-9.

[37] C. Zhou, "A new robust method for blood vessel segmentation in retinal fundus images based on weighted line detector and hidden Markov model," April 2020. [Online]. Available:

https://www.sciencedirect.com/science/article/abs/pii/S0169260719306169.

[38] S. Akhtar, "A deep learning based model for diabetic retinopathy grading," 30 January 2025. [Online]. Available: https://doi.org/10.1038/s41598-025-87171-9.

[39] G. Tabacaru, "A Robust Machine Learning Model for Diabetic Retinopathy Classification," 28 December 2023. [Online]. Available: https://www.mdpi.com/2313-433X/10/1/8.

[40] X. Qian, "The effectiveness of artificial intelligence-based automated grading and training system in education of manual detection of diabetic retinopathy," 7 November 2022. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/36419999/.

[41] N. Gharaibeh, "An effective image processing method for detection of diabetic retinopathy diseases from retinal fundus images," January 2018. [Online]. Available: https://www.researchgate.net/publication/326658937_An_effective_image_processing_method_for_detection_of_diabetic_retinopathy_diseases_from_retinal_fundus_images.

[42] A. Mutawa, "Randomization-Driven Hybrid Deep Learning for Diabetic Retinopathy Detection," 27 Febraury 2025. [Online]. Available: https://ieeexplore.ieee.org/document/10906576.

[43] Z. Liu, "A ConvNet for the 2020s," 2 March 2022. [Online]. Available: https://arxiv.org/pdf/2201.03545.

[44] X. Liu, "Vision Mamba: A Comprehensive Survey and Taxonomy," 7 May 2024. [Online]. Available: https://arxiv.org/pdf/2405.04404.

[45] L. Zhu, "Vision Mamba: Efficient Visual Representation Learning with Bidirectional," 14 November 2024. [Online]. Available: https://arxiv.org/pdf/2401.09417.

[46] A. Hatamizadeh, "MambaVision: A Hybrid Mamba-Transformer Vision Backbone," 25 March 2025. [Online]. Available: https://arxiv.org/pdf/2407.08083.

[47] A. Dosovitskiy, "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," 22 October 2020. [Online]. Available: https://arxiv.org/abs/2010.11929.

[48] [Online]. Available: https://github.com/microsoft/Swin-Transformer.

[49] "GeeksforGeeks," 20 january 2025. [Online]. Available: https://www.geeksforgeeks.org/swin-transformer/.

[50] Z. Liu, "Swin Transformer: Hierarchical Vision Transformer using Shifted Windows," 25 March 2021. [Online]. Available: https://arxiv.org/abs/2103.14030.

[51] "GeeksforGeeks," 3 Feb 2025. [Online]. Available: https://www.geeksforgeeks.org/understanding-logistic-regression/.

[52] J. Gildenblat, "PyTorch library for CAM methods," 2021. [Online]. Available: https://github.com/jacobgil/pytorch-grad-cam.

[53] "Aptos 2019 competition," [Online]. Available: https://www.kaggle.com/competitions/aptos2019-blindness-detection/leaderboard.

[54] "frotiers in Endocrinology," [Online]. Available: https://www.frontiersin.org/journals/endocrinology/articles/10.3389/fendo.2023.1197783/full.

[55] 16 December 2021. [Online]. Available: https://www.nhs.uk/conditions/diabetic-retinopathy/.