Engineering Applications of Artificial Intelligence 160 (2025) 111938

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

journal homepage: www.elsevier.com/locate/engappai

Automated detection of stages, zones, and plus diseases of Retinopathy of Prematurity using quantum convolutional networks in neonatal fundus images

Raja Sankari VM [a], Snekhalatha Umapathy [a,*], Ashok Chandrasekaran [b], Prabhu Baskaran [c], Varun Dhanraj [d]

[a] Department of Biomedical Engineering, College of Engineering & Technology, SRM Institute of Science and Technology, Kattankulathur, 603203, Tamil Nadu, India
[b] Department of Neonatology, SRM Medical College Hospital & Research Centre, Kattankulathur, 603203, Tamil Nadu, India
[c] Vitreo-Retina Surgeon, Aravind Eye Hospital, Chennai, Tamil Nadu, India
[d] Department of Physics and Astronomy, University of Waterloo, Waterloo, ON, Canada

ARTICLE INFO

Keywords:
Quantum computing
Deep learning
Retinal blood vessel segmentation
Quantum mobile vision transformer
Multiclassification
Retinopathy of prematurity

ABSTRACT

The proposed work develops Quantum Mobile Vision Transformer (QMViT) designed with the principles of quantum mechanics, vision transformer and pretrained Convolutional Neural Network (CNN) network to classify stages, zones and severity of Retinopathy of Prematurity (ROP). The proposed work aims to i) design and validate a light-weight quantum network QMViT to predict the characteristic stages, zones and plus disease of ROP; ii) classify the images into three multiclass classifications using state-of-the-art networks like Vision Transformer (ViT), Swin Transformer, Residual Vision Transformer (ResViT) and Mobile ViT and machine learning (ML) classifiers. The study utilises the HVDROPDB-BV (ROP Blood vessel) dataset in training the segmentation networks, and a real-time collected dataset with 2400 images (2000 ROP and 400 Normal) in training the classification models. The retinal blood vessels are segmented using transformer-based SwinUNet and MultiResUNet, from which 1000 Scale Invariant Feature Transform (SIFT), Speeded Up Robust Features (SURF), and Oriented Fast and Rotated Brief (ORB) features are extracted, fused, and dimensionally reduced to 25 components using Principal Component Analysis (PCA). Three ML classifiers classify 25 components into different ROP stages, zones, and severity. However, transformer-based networks such as ViT, Swin Transformer, ResViT, MobileViT, and Quantum-based QMViT classified retinal images directly into characteristic stages, zones, and Plus disease of ROP. The novel QMViT surpasses all other state-of-the-art deep learning and ML classifiers, with an accuracy of 95.5 %, 96.88 %, and 96.67 % in classifying stages, zones, and plus disease of ROP, respectively. QMViT is a hardware-efficient network that leverages quantum-inspired features for medical image classification.

1. Introduction

Retinopathy of prematurity (ROP) is a retinal proliferative vascular disorder which is the leading cause of preventable blindness in preterm infants. It affects premature infants with delayed retinal vasculature development, progressing to abnormal angiogenesis and resulting in total retinal detachment and irreversible blindness (Jang, 2024). The incidence of ROP increases with the degree of immaturity and intra-uterine growth retardation (Good, 2004). Early gestational age, lower birth weight, and higher concentration of oxygen therapy are crucial risk factors for ROP (Hellström et al., 2013). Factors such as anaemia, sepsis, postnatal weight gain, serum insulin-like growth factor-1 (IGF-1) levels, thrombocytopenia, bilirubin levels, gender, multiple gestations, intraventricular haemorrhage, and blood transfusion influence the incidence and course of ROP (Kim et al., 2018).

Improved survival rates in neonatal critical care units and inadequate facilities for monitoring oxygen therapy result in an increasing population of neonates suffering from ROP, notably in developing countries (Freitas et al., 2018). Smith et al. describe ROP as the leading preventable cause of blindness Worldwide (Smith, 2004). About 50,000 neonates lose their sight due to ROP each year, with the highest incidence in Latin America, Southeast Asia, and Eastern Europe (Gilbert,

* Corresponding author.
E-mail address: snehalau@srmist.edu.in (S. Umapathy).

https://doi.org/10.1016/j.engappai.2025.111938
Received 13 December 2024; Received in revised form 27 June 2025; Accepted 31 July 2025
Available online 5 August 2025

2008). The prevalence of ROP varies from 7 % to 37 % globally (Fares et al., 2024). In India, 38 %–47 % of babies were diagnosed with ROP across various locations (Bowe et al., 2019). In a study conducted in South India, the prevalence of ROP was estimated to be 32.6 %, of which 13.2 % of infants required treatment for severe ROP (Ahuja et al., 2018).

According to the third edition of the International Classification of Retinopathy of Prematurity (ICROP) (Chiang et al., 2021), three types of classification are used to describe the characteristics of ROP in an eye. The first type of classification is based on the stage of the disease, ranging from 1 to 5. Stage 1 is defined by the presence of a thin, flat, and white structure called a demarcation line at the junction of the vascular and avascular retina. This demarcation line progresses into a ridge with the width and height of colours varying from white to pink in stage 2. Popcorn, little tufts of neovascular tissue on the retina, can be found posterior to this ridge in stage 2. As the proliferation becomes significant in stage 3, the extraretinal neovascular proliferation spreads from the ridge into the vitreous and is continuous with the posterior aspect of the ridge. Partial retinal detachment occurs in stage 4 and excludes or affects the fovea. It is characterised by the loss of delicate choroidal vasculature or granular pigment epithelium and a glass appearance relative to the adjacent attached retina. Total retinal detachment occurs in stage 5 and leads to permanent blindness. Fig. 1 depicts the examples of neonatal fundus images in ROP stages 1–4.

The second type of classification is based on the location of vascularisation as a concentric circle with the optic disc as the centre, ranging from zones I to III, as in Fig. 2. The indication of the zone in ROP describes the maturity and risk of developing ROP in the infant. A circle with a radius twice the distance between the optic disc and the foveal centre represents Zone I. Zone II is a circular region that extends nasally from the outer boundary of Zone I to the nasal ora serrata and at similar temporal, superior, and inferior distances. The peripheral retinal crescent beyond Zone II constitutes Zone III. The third type of ROP classification is based on the severity of the disease and is divided into plus and preplus diseases. Plus disease is characterised by the presence of dilation and tortuosity of the posterior retinal vessels. Preplus disease indicates the stage that could, over time, progress into plus disease. It is aberrant retinal vessel dilation and tortuosity but not as severe as plus disease.

Although many treatments with better outcomes are available to treat ROP, many preterm suffer from permanent blindness due to delayed screening, notably in developing countries. Severe cases of ROP require surgical treatment, and only 20 %–50 % of surgeries provide anatomical success. Even after successful anatomical results, the visual outcome may be slow and limited, with retinal detachment recurrence of 5 % and 22 % in Stages 4 and 5 (Sen et al., 2018). The typical ophthalmoscopic diagnosis of ROP in infants is complex and requires expertise. Sen et al. emphasize that delayed presentation of preterm infants with severe retinal detachment and lack of sufficiently trained paediatric retinal surgeons and anaesthetists leads to an enormous burden of infants becoming blind, or severely visually impaired (Sen et al., 2020). Due to high variability and inter-observer inconsistency in the ROP diagnosis (Chiang et al., 2007; Kalpathy-Cramer et al., 2016) and shortage of skilled personnel, it is essential to identify and diagnose different stages of ROP to prevent progression to severe sight-threatening disease.

2. Related works

In recent years, there has been an elevated demand for artificial intelligence (AI) to transform healthcare across various areas of specialisation (Bajwa et al., 2021). Developments in deep learning using Convolutional Neural Network (CNN) enable image-based AI systems to resemble manual clinical implementation by experts. Applications of AI in the analysis of retinal images have improved the diagnosis of various ophthalmologic diseases, such as Diabetic Retinopathy (Grzybowski et al., 2020), Glaucoma (Zheng et al., 2019), and Age-related Macular Degeneration (Crincoli et al., 2024).

Coyner et al. detected plus disease using a pre-trained CNN trained

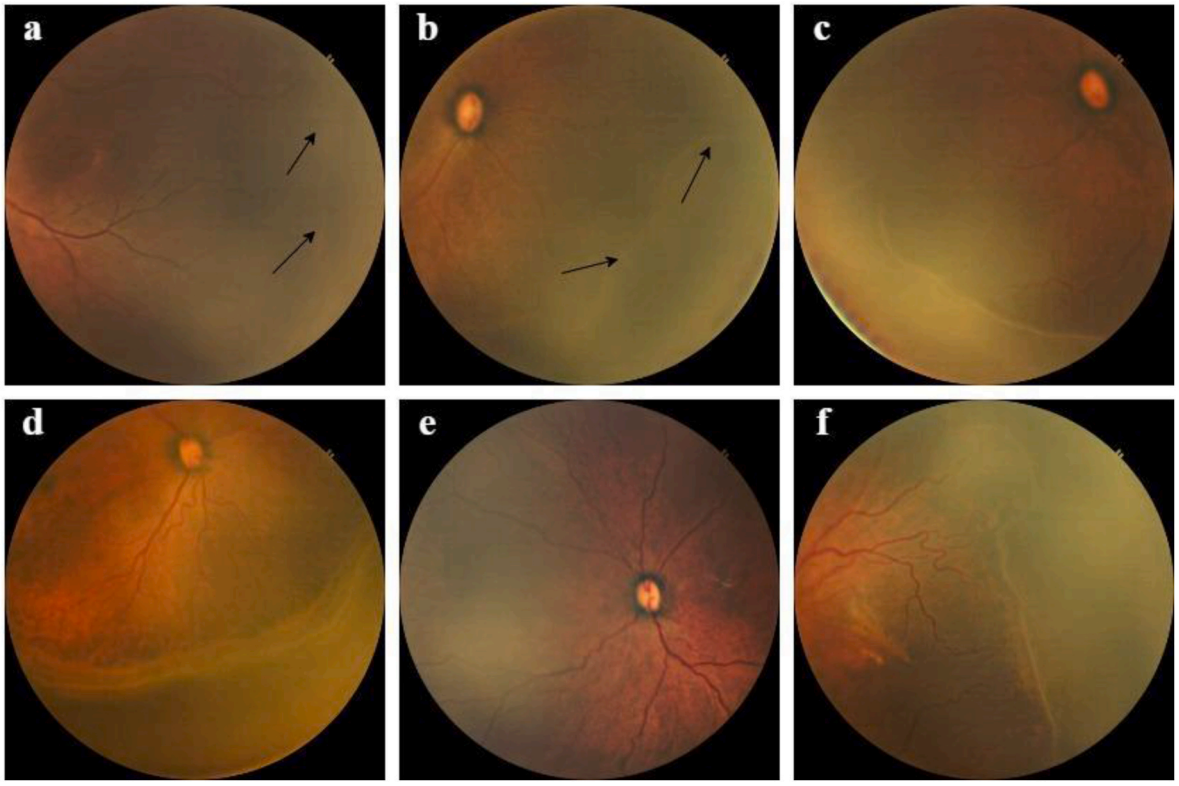

Fig. 1. Fundus images demonstrating examples of ROP a) Stage 1 (black arrow facing the demarcation line); b) Stage 2 (black arrow facing the ridge); c) Stage 3; d) Stage 4; e) Preplus disease; f) Plus disease.

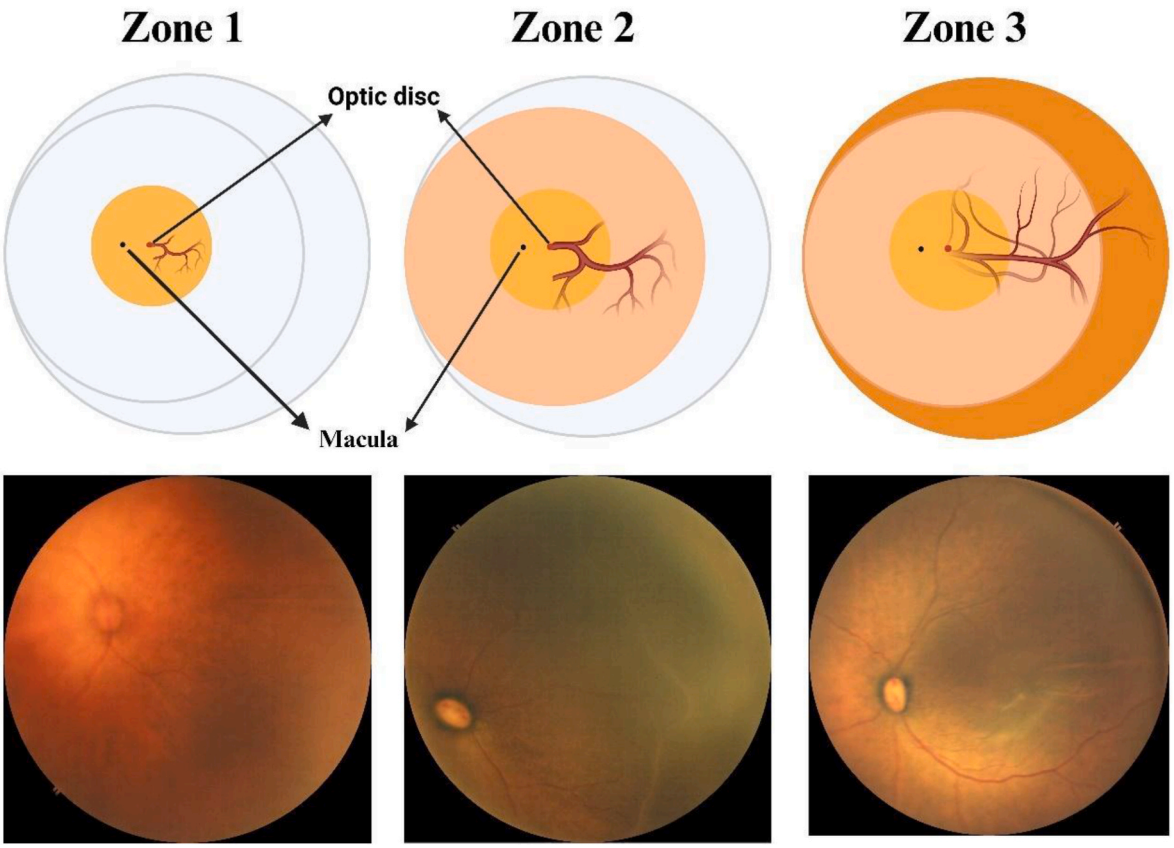

Fig. 2. Schematic illustration and examples of fundus images of ROP in Zone I, II and III.

by synthetic data (Coyner et al., 2022). Retinal vessel maps (RVM) were segmented using U-Net and were used to train progressively growing generative adversarial networks (PGAN) to generate synthetic RVMs. Finally, they classified using ResNet 18 CNN to predict plus and preplus disease from normal images with an Area Under the Curve (AUC) of 0.971 and 0.934 in synthetic and raw datasets, respectively. The study focused on classifying the retinal images into plus and preplus disease but could not specify the stage or zone of the disease. Deng et al. investigated the morphological characteristics of ROP using the segmented retinal vessels in fundus images (Deng et al., 2023). The blood vessels are segmented from the retinal images using U-Net, from which the morphological characteristics such as avascular area, angle of the vessel, vessel density, fractal dimension (FD), and tortuosity are estimated. Although the system achieved a 72 % sensitivity and 99 % specificity in neonatal blood vessel segmentation, the images only with the optic disc centre located within the specified circle were selected for segmentation based on a U-Net.

Agrawal et al. segmented blood vessels and the optic disc using an ensemble network of U-Net and Circle Hough Transform to detect three zones of ROP in a private dataset (Agrawal et al., 2021). The system achieved an accuracy of 98 % in predicting zones of ROP, but it does not focus on the stages and severity of the disease. Rao et al. designed a binary classification system to detect ROP from neonatal retinal images (Rao et al., 2023). The authors classified 227,326 early-stage ROP images obtained from the KIDROP tele-ROP screening program using EfficientNet-B0 CNN. The classification accuracy is 91.29 %, with 91.46 % sensitivity and 91.22 % specificity. The binary classification achieved in this study could further be improved to detect ROP's characteristic stage, zone and severity.

Jemshi et al. designed an Artificial Neural Network (ANN) using Wavelet and Curvelet transform-based feature selection to classify the preterm fundus images into Plus disease and Normal (Jemshi et al., 2024). The authors achieved a maximum accuracy of 96 % with 93 % Specificity and 100 % Sensitivity using Curvelet transform features combined with vascular features. However, the developed system was not trained to identify the preplus disease or other characteristics of ROP. Young et al. designed a smartphone-based fundus imaging method to classify the infant retinal images into normal and preplus or plus disease and to decide whether referral is warranted (RW) and treatment required (TR)-ROP (Young et al., 2023). The binary classification was carried out using ResNet 18 CNN with a sensitivity of 80.0 %, a specificity of 59.3 % for RW, 100 % sensitivity and 58.6 % specificity for TR ROP.

Salih et al. classified the three zones of ROP using three pre-trained networks such as Visual Geometry Group 19 (VGG 19), ResNet 50, and EfficientNetB5, from 1365 fundus images (Salih et al., 2023). The authors ensembled all the three models and classified them using a voting classifier technique to achieve a maximum accuracy of 88.82 %. However, ensembling three pre-trained CNNs elevates the computational complexity. Liu et al. designed a binary classification system to identify ROP, determine the need for treatment, and determine the treatment modalities (Liu et al., 2023). The authors used DenseNet 121 and ResNet 18 CNN to classify 24495 RetCam images with an accuracy of 92.0 % in all the three tasks. Subramaniam et al. developed an automated network to perform binary classification of plus disease and normal images using pre-trained GoogLeNet (Subramaniam et al., 2023). The retinal blood vessels are enhanced, harmonised, and then classified to achieve an Area Under the Receiver Operating Characteristic Curve (AUROC) of 0.97.

Table 1 lists the identified research gaps from the previous works of ROP prediction. Most of the related works (Coyner et al., 2022; Deng et al., 2023; Agrawal et al., 2021; Rao et al., 2023; Jemshi et al., 2024; Young et al., 2023; Salih et al., 2023; Liu et al., 2023; Subramaniam et al., 2023) are focused on either the binary classification of ROP or the classification of zone, stage or plus disease separately. However, a single

Table 1
Identified Research gaps from previous related works.

Year	Article	Segmentation Network	Classification Network	Performance	Lagged Work
2022	Coyner et al. (Coyner et al., 2022)	U-Net	ResNet 18 CNN	AUC of 0.971 and 0.934 in synthetic and raw datasets in predicting plus and preplus disease	Could not specify the stage or zone of the disease
2023	Deng et al. (Deng et al., 2023)	U-Net	–	72 % sensitivity and 99 % specificity in neonatal blood vessel segmentation	Images only with the optic disc centre located within the specified circle were utilised for retinal vascular segmentation
2021	Agrawal et al. (Agrawal et al., 2021)	U-Net and Circle Hough Transform	Self-Collected Images	Accuracy of 98 % in predicting zones of ROP	Could not specify the stages and severity of the disease
2023	Rao et al. (Rao et al., 2023)	–	EfficientNet-B0 CNN	Binary classification accuracy is 91.29 %, with 91.46 % sensitivity and 91.22 % specificity	Could not specify the stages, zones and severity of the disease
2023	Jemshi et al. (Jemshi et al., 2024)	–	ANN based on Wavelet and Curvelet transform-based feature selection	Accuracy of 96 % with 93 % Specificity and 100 % Sensitivity	Could not specify the stages, zones and preplus condition of the disease
2023	Young et al. (Young et al., 2023)	–	ResNet 18 CNN	Sensitivity of 80.0 %, a specificity of 59.3 % for RW, and 100 % sensitivity and 58.6 % specificity for TR ROP.	Could not specify the stages, zones and severity of the disease
2023	Salih et al. (Salih et al., 2023)	–	VGG 19, ResNet 50, and EfficientNetB5 with a voting classifier	Classified three zones of ROP with a maximum accuracy of 88.82 %.	Ensembling three pre-trained CNNs increases the complexity of the computation and could not specify the stages and severity of the disease
2023	Liu et al. (Liu et al., 2023)	–	DenseNet 121 and ResNet 18 CNN	92.0 % accuracy in identifying ROP, determining the need for treatment and determining the treatment modalities	Could not specify the stages, zones and severity of the disease
2023	Subramanium et al. (Subramaniam et al., 2023)	–	GoogLeNet	AUROC of 0.97	Could not specify the stages, zones and severity of the disease

automated system designed to detect the characteristic stage, zone and plus disease of ROP can be significant for relevant clinical diagnosis. Also, all the previous research is based on conventional pre-trained CNNs. The applications of state-of-the-art networks like transformers in ROP diagnosis are limited. Further, the computational complexity of utilising these classical CNNs could be reduced by quantum computing. Quantum Convolutional Neural Networks (QCNN) ensembles the principles of quantum superposition and entanglement with classical CNN in order to improve the computational capability of the network (Sharma, 2022). To the best of our knowledge, this is the first study to introduce a hybrid model named Quantum Mobile Vision Transformer (QMViT) designed with the principles of quantum mechanics, vision transformer and pretrained CNN network to classify stages, zones and severity of ROP.

The novelty of the proposed work are as follows: characterise all the stages, zones, and severity of ROP using a hybrid quantum-based network and state-of-the-art transformer-based pre-trained networks. The current study proposes a novel lightweight QMViT to classify the stages (1–4), zones (I-III) and plus (plus and preplus) disease of ROP. The network utilises the quantum principles of computing the MobileNet pre-trained CNN by converting the images into patches using linear embeddings and recognising the features using added positional embeddings and a self-attention mechanism. Also, the proposed work segments the retinal blood vessels of neonatal fundus images using novel shifted windows-based transformer network like Swin UNet and classifies them using machine learning (ML) classifiers.

The proposed QMViT model was explicitly designed with the principle of lightweight computation and architectural efficiency using MobileNet blocks. This makes it inherently conducive to migration, deployment on low-resource medical devices, and interoperability with diverse imaging modalities. Unlike many existing quantum image processing models that rely on complex image encoding schemes such as the Flexible Representation of Quantum Images (FRQI) and the Novel Enhanced Quantum Representation (NEQR) (Le et al., 2011; Zhang et al., 2013), deep quantum circuits, or large numbers of qubits for tasks such as edge detection (Syed et al., 2024) and compression (Majumder et al., 2023), the proposed network is intentionally lightweight. It uses a minimal quantum footprint—just four qubits and a shallow, fixed-depth random quantum circuit—to perform localized nonlinear feature transformations. Prior methods such as Quantum Hadamard Edge Detection and QCNNs (Syed et al., 2024; Srivastava and Dwivedi, 2021) often require significantly more quantum resources, including deeper circuits and qubit counts exceeding 16, which increases hardware demands and limits scalability. In contrast although the number of parameters and inference time in the classical approach of MobileViT (2.3 M parameters with inference time of 7.8 ms) is less than its quantum version, our model integrates this minimal quantum enhancement within a hybrid classical architecture, combining the spatial inductive biases of CNNs with the global reasoning capabilities of transformers. This results in a network that is computationally efficient, hardware-feasible, and still capable of leveraging quantum-induced representational advantages for medical image classification.

The proposed work is aimed to i) extract high-level features from Swin UNet segmented neonatal blood vessels and classify them using three ML classifiers such as Support Vector Machine (SVM), Random Forest (RF) and *k*-Nearest Neighbour (*k*-NN) to detect stages (1–4), Zones (I-III) and Plus (Plus and preplus) disease of ROP; iv) classify the raw images into three types of multiclass classifications using state-of-the-art networks such as Vision Transformer (ViT), Swin Transformer, ResViT and MobileViT; iii)design Quantum based QMViT network to detect ROP with characteristic stages, zones and plus disease.

The contributions of the proposed work are summarised below

1. The neonatal retinal blood vessels are segmented from the fundus images using state-of-the-art transformer-based Swin UNet and compared with MultiResUNet;
2. Thousand high-level feature descriptors such as Scale Invariant Feature Transform (SIFT), Speeded Up Robust Features (SURF), and Oriented Fast and Rotated Brief (ORB) are extracted from the segmented vessels and are then fused together. Principal Component Analysis (PCA) is used to reduce the dimensionality of the extracted features to 25.

3. The dimensionally reduced features are classified using three ML classifiers such as SVM, RF and k-NN, to predict the stage, zone and severity of the disease.
4. The stages (1–4), zones (I-III) and plus (Plus and preplus) classification is carried out by stand-alone transformer-based networks such as ViT, Swin Transformer, ResViT and MobileViT.
5. Compared the classical CNN and QCNN in ROP multiclass classification.
6. Designed and validated light-weight quantum CNN QMViT to predict the characteristic stages, zones and plus disease.

3. Methodology

The overall block diagram for ROP prediction is illustrated in Fig. 3. The fundus images from preterm individuals are segmented to extract the retinal blood vessels using two CNNs, namely transformer-based Swin UNet and MultiResUNet. Thousand SIFT, SURF, and ORB high-level feature descriptors are extracted from the segmented retinal vessels and are fused to result in 3000 features. To reduce the number of features, PCA is applied, bringing the feature dimensions down to 25. Twenty-five dimensionally reduced features are used as an input to the three ML classifiers such as SVM, RF, and k-NN, to predict the characteristic stages, zones, and severity of the input retinal image. However, deep learning classification bypasses complex segmentation, feature extraction, dimensionality reduction, and ML classification. Transformer-based networks such as ViT, Swin Transformer, ResViT, and MobileViT were utilised to process the raw input images and predict their respective stages, zones, and severity. Quantum MobileViT, a lightweight quantum-based deep learning transformer model, predicts stages, zones, and severity of ROP using quantum preprocessed images at a reduced computational complexity.

3.1. Retinal image collection

Preterm fundus images were collected retrospectively from July 2019 to July 2023 from SRM Medical College, Hospital and Research Centre, Chengalpattu, in collaboration with Aravind Hospital, Chennai. The study has adhered to the tenets of the Declaration of Helsinki and was ethically approved by the Institutional review board of the hospital with clearance number 8241/IEC/2022. Informed consent was obtained from the parents for essential investigations and treatments. Standard ROP screening guidelines were used to select the preterm fundus images for the study. All preterm infants were initially screened at 4 weeks postnatally. Neonates were examined by dilating the pupils with 0.5 % tropicamide and 2.5 % phenylephrine eye drops. A paediatric eye speculum was employed by qualified specialists to keep the inspected eye open.

The neonatal retinal images were obtained using a digital wide-field fundus camera 3NethraNeo (Forus Health, Bangalore) with a FOV of 120°. About 2400 images (with 2000 ROP and 400 Normal) were selected for the study. The characteristic stages, zones and severity of the disease are annotated by paediatric ophthalmologists. The unbalanced dataset in hand is augmented using geometric and optical techniques in order to generalise the designed CNNs to predict the images in all positions with and without noise. Table 2 describes the dataset in each stage, zone and plus disease, and the type of augmentations utilised. In order to make a balanced dataset of all classes, the images are augmented and selected based on the quality of the collected images. The number of images in each stage, zone and plus is augmented and selected to 400, 400, and 200, respectively. The dataset HVDROPDB-BV (Agrawal et al., 2023) is used in training the segmentation networks, which has publicly available fundus images of neonates of 26–36 weeks gestation and weighing 3000 g or less. The dataset has 50 fundus images and 50 ground truth images annotated by ROP experts, obtained using a Neo fundus camera.

3.2. Segmentation

Segmentation of regions of interest, such as blood vessels, optic disc and ridges, is crucial in the automatic medical image analysis of the retina in ROP diagnosis. However, the segmentation of blood vessels has the potential to explain all the stages, zones and severity of ROP with the least computations. The input retinal images collected were resized to $128 \times 128 \times 3$ and provided to the proposed segmentation algorithm called Swin UNet and is validated against MultiResUNet, the network evolved from the widely utilised UNet.

Medical image segmentation algorithms typically employ U-shaped CNNs called U-Net, which have an encoder, a symmetric decoder, and skip connections (Ronneberger et al., 2015). The encoder downsamples the image retaining the deep features, and the decoder upsamples the feature map fused with the features from the skip connections to provide the semantic prediction of the segmentation. Thus, U-Net succeeded in medical image segmentation in various specialities, including cardiology, neurology, ophthalmology, etc. The evolution of U-Net resulted in

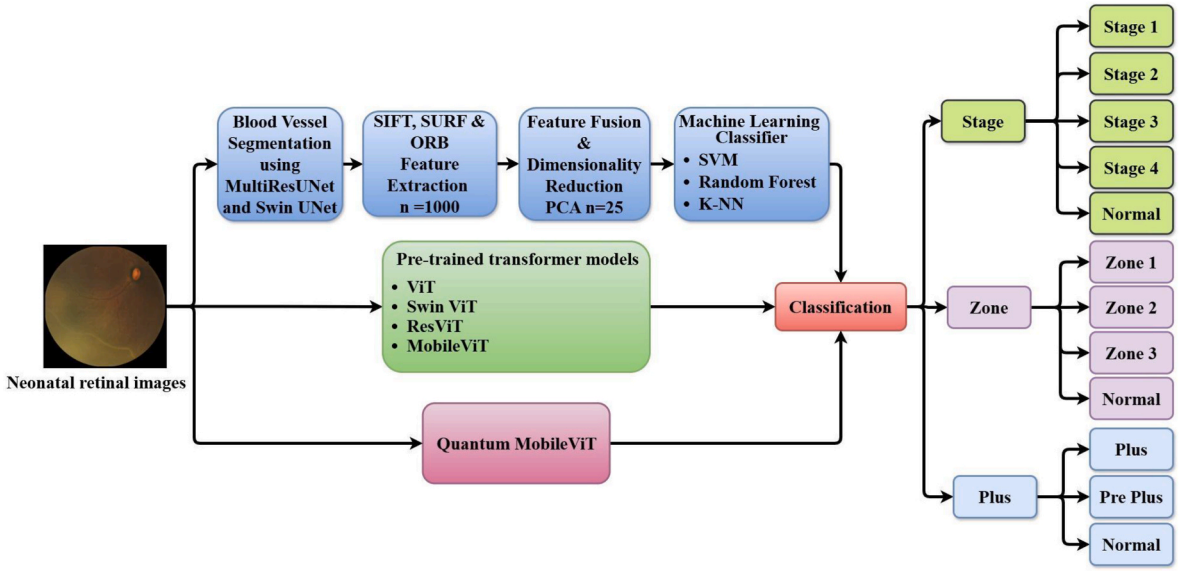

Fig. 3. Overall Schematic diagram of the proposed study in ROP prediction.

Table 2
Description of the collected images for ROP multiclass classification.

Attributes		Number of Images Available	Types of Augmentation	Augmented and Selected Images	Average GA (in weeks)
Stages	Stage 1	202	Rotate 90°	400	29.23 ± 2.4
	Stage 2	368	Rotate 90°	400	28.52 ± 2.39
	Stage 3	229	Rotate 90°	400	29.65 ± 3.54
	Stage 4	44	Rotate 90°, Horizontal and vertical flip, Transpose, Elastic transform, Motion blur, median and Gaussian blur, and Gauss noise	400	30.12 ± 1.24
Zones	Zone 1	25	Rotate 90°, Horizontal and vertical flip, Transpose, Elastic transform, Motion blur, median and Gaussian blur, Gauss noise, Grid distortion, Random brightness and contrast, Random gamma, Hue saturation value and channel shuffle.	400	28.84 ± 1.37
	Zone 2	389	Rotate 90	400	27.98 ± 3.21
	Zone 3	1515	–	400	30.96 ± 3.24
Severity	Plus	13	Rotate 90°, Horizontal and vertical flip, Transpose, Elastic transform, Motion blur, median and Gaussian blur, Gauss noise, Grid distortion, Random brightness and contrast, Random gamma, Hue saturation value and channel shuffle.	200	26 ± 3
	Preplus	63	Rotate 90°, Horizontal and vertical flip	200	31.18 ± 2.67
Normal		600	–	400	30.53 ± 3.39

many advanced networks such as U-Net++ (Zhou et al., 2018), Res-UNet (Xiao et al., 2018), MultiResUNet (Ibtehaz and Rahman, 2020), etc. MultiResUNet utilises multiple Res pathways to process encoded features through a series of convolution layers with residual connections to reduce the semantic gap between encoder and decoder features.

However, the convolution-based U-Net and its evolutions could not extract long-range semantic information from the medical images (Chen et al., 2021). CNN-based networks face challenges in obtaining global and long-range semantic information due to the intrinsic locality of the convolutions. Researchers have tried introducing the transformer architecture to vision applications after its success in natural language processing (NLP) (Vaswani et al., 2017). ViT recognises the image by converting it into image patches with positional embeddings and a self-attention mechanism (Dosovitskiy et al., 2010). The hierarchical Swin Transformer is designed using shifted windows to improve computational efficiency by non-overlapping local and cross-window connections (Liu et al., 2021).

Swin UNet is built using the Swin transformer as the backbone of the U-shaped encoder-decoder architecture, leveraging its applications in segmenting blood vessels in neonatal retinal images. The Swin UNet has linear computational complexity with respect to image size.

The architecture of Swin UNet is depicted in Fig. 4. The Swin UNet has an encoder, bottleneck, decoder, and interconnecting skip connections with the Swin transformer as the basic unit. In the encoder block, each retinal image provided as input is split into n non-overlapping patches. Let us consider the dimensions of input raw images be height H and width W and channel C. The input feature map of dimensions H × W × C are converted into n patches of size x × y. Then, the total number of patches extracted from one image as $n = \frac{H}{x} \times \frac{W}{y} \times C$. The input image of size 128 × 128 × 3 is provided to extract 1024 patches of dimension 4 × 4 × 3 = 48. The extracted patches are fed to the patch embedding layer, which converts them into tokens of dimension 128. These tokens are passed to three Swin Transformer blocks, each with four Swin transformers and patch merging layers. A Swin transformer block consists of a layer normalisation, a multi-head self-attention module, residual connections, and a two-layered multi-layer perceptron (MLP) with Gaussian Error Linear Unit (GELU) non-linearity. The Window-based Multi-head Self-Attention (W-MSA) and the Shifted Window-based Multi-head Self-Attention (SW-MSA) are utilised in the two successive Swin transformer blocks. The Swin transformers extract the deep features from the tokens, while the patch merging layer downsamples the feature map to reduce the total number of tokens to half and doubles its dimensions.

The bottleneck of Swin UNet learns the extracted feature representation using two Swin transformer blocks without altering the feature resolution and dimension. The decoder consists of three Swin Transformer blocks, each with four Swin Transformers and patch-expanding layers. The context features extracted from the encoder are up-sampled at the decoder using a patch-expanding layer. The patch-expanding layer retains the resolution of the feature maps by doubling it. These up-sampled features are then combined with the multi-scale features from the encoder using skip connections at each stage in order to retain the spatial resolution of the feature maps. In the end, the linear projection layer produces the pixel-level segmented blood vessels from the up-sampled feature maps.

The dataset HVDROPDB-BV, which includes 50 neonatal fundus images and their respective ground-truth images of blood vessels, is utilised to train and test the Swin UNet and MultiResUNet architectures. The images were split in the ratio of 80:10:10 to train (40 images), validate (5 images) and test (5 images) the networks. Both Swin UNet and MultiResUNet are trained using the loss function set as Categorical Cross-entropy with adam as the optimizer at a learning rate of 1 × e−4 for 300 epochs. After successful training and testing of both networks, all the 2400 collected images (with 2000 ROP and 400 Normal) are provided to the above-mentioned networks to extract the neonatal retinal blood vessels.

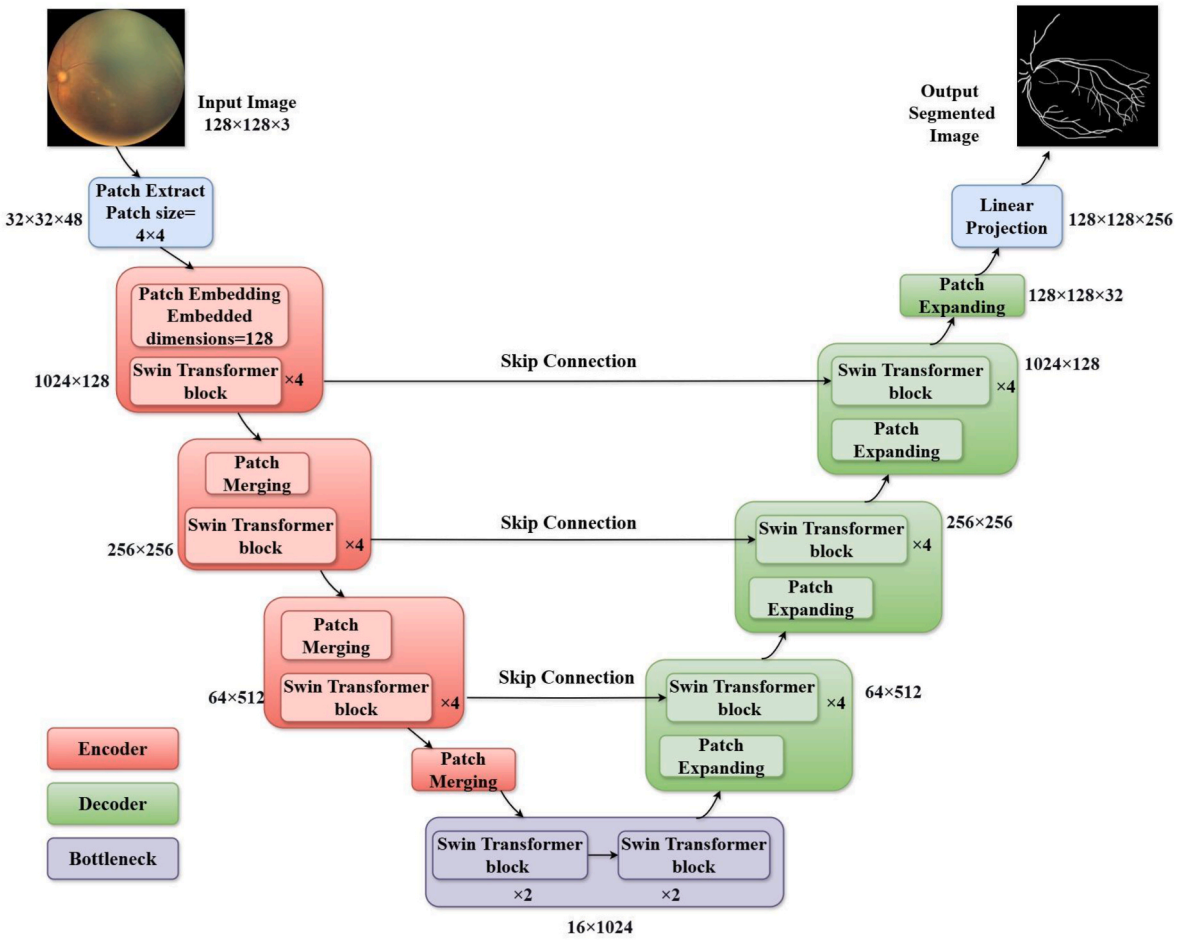

Fig. 4. Illustration demonstrating the architecture of Swin UNet for ROP prediction.

3.3. Feature engineering

The acquired 2400 images segmented using the Swin UNet and MultiResUNet were given as input to the feature engineering block, where three descriptor-based high-level features, such as SIFT, SURF, and ORB, are extracted. The SIFT algorithm extracts robust high-level features from the segmented retinal vessels that are invariant to changing scales and rotation (Lowe, 2004). The SIFT algorithm follows four major stages of computation such as Scale-space extrema detection, Keypoint localisation, Orientation assignment and keypoint descriptor. The segmented blood vessels are transformed into scale space images of varying scales. The difference-of-Gaussian (DoG) is estimated to detect the scale- and orientation-invariant potential points of interest. The optimal keypoints are then localised by a stability-based cascade filtering approach and selected based on local maxima and minima of the blood vessels. Keypoints localised near edges and poor contrast are rejected. About, 1000 SIFT descriptors that are highly invariant to scale and orientation are selected from the segmented blood vessels.

The SURF algorithm effectively extracts the scale and orientation invariant keypoints from the blood vessels at higher computational speed without compromising the robustness of the descriptors (Bay et al., 2008). The algorithm identifies the interest points using a Hessian matrix for the scale-space analysis. This is carried out using integral images with rapid convolutions, which significantly increases the speed of computation. The reproducible orientation is assigned to the identified interest points by estimating Haar-wavelet responses in both the x and y directions. The orientation of the descriptor is estimated on the basis of the dominant wavelet response, ensuring that the descriptor remains invariant to rotation. The SURF descriptor extracts the distribution of the wavelet responses and provides a concise representation of the local image structure. The performance of the SURF algorithm is improved by effective matching techniques, such as similarity thresholding and nearest neighbour ratio methods.

ORB provides a fast and efficient method for image feature recognition and description. The ORB keypoints are detected using the FAST (Features from Accelerated Segment Test) algorithm, which recognises the image corners by comparing the intensity between the central pixel and its surroundings (Calonder et al., 2010). ORB employs a Harris corner measure in order to rank the detected keypoints by corners, ensuring that the most significant features are selected. ORB computes the descriptors for all identified keypoints using the BRIEF (Binary Robust Invariant Scalable Keypoints) algorithm. ORB includes the orientation component to BRIEF by calculating the intensity centroid, making the descriptors invariant to rotations. ORB combines the strengths of FAST and BRIEF, providing computationally efficient descriptors that maintain robustness against noise, rotation, and scale variations.

Finally, 3000 distinctive keypoints (1000 SIFT, 1000 SURF, and 1000 ORB) containing high-level information about the localisation of blood vessels are obtained from each preterm retinal image and fused images. When these 3000 fused features are extracted from all 2400 images and processed for classification using neural networks, the computational complexity is extremely high. PCA is used to address the elevated complexity due to the 'curse of dimensionality'. PCA statistically reduces the dimensions of the features while preserving as much variance as possible in a dataset (Lever et al., 2017). PCA transforms a set of correlated features into a smaller set of uncorrelated features called the principal components using a linear transformation. The process of

reducing the dimensions involves standardisation of the input data, computation of the covariance matrix, eigenvalues and eigenvectors, and finally, selecting the best k eigenvectors related to the largest eigenvalues. Thus, PCA represents data more efficiently with minimal information loss. In the proposed work, the extracted 3000 features are reduced to 25 principal components using PCA, reducing the computational complexity in classification in ML classification.

3.4. Machine learning classification

ML classification is the process of categorizing data instances into predefined classes based on the input features by training and validating the classifier. About, 25 principal components are extracted from the 2400 collected images (with 2000 ROP and 400 Normal) and provided as input to the classifiers, and three types of multiclass classifications, such as stage, zone, and plus, are carried out. In all the types of classifications, 80 % of images are utilised in training, and the rest, 20 %, are used to test the classifier. The ML classification is carried out using three classifiers such as SVM, RF and *k*-NN. SVM classifies the data points by determining the optimal hyperplane in order to maximise the margin between different classes using kernel functions (Hearst et al., 1998). RF is an ensemble learning classifier, which includes several decision trees to generate an accurate result by reducing data overfitting (Cutler et al., 2012). *k*-NN is a non-parametric classifier that predicts the class of the input instance based on the distance metrics selected by the members of each class (Cunningham and Delany, 2021).

The SVM classifier is tuned using hyperparameters such as the Radial Basis Function (RBF) as the kernel function, the regularization parameter as 1.0, squared l2 as the penalty, and the gamma value is set to vary based on the variance between the input features. The hyperparameters used in the RF classifier are the Gini index for the impurity criteria, with 10 trees in each forest, and a minimum of two samples would be required to split an internal node. In the *k*-NN classifier, the Euclidean distance is used as the distance metric with the number of nearest neighbours to consider when making a prediction tuned to 10.

3.5. Deep learning classification

The state-of-the-art deep learning network like Vision Transformer excels in image classification tasks, outperforming convolution-based CNNs (Dosovitskiy et al., 2010). The proposed work involves classical preprocessing techniques such as image resizing (to a size of 128×128) and normalizing to classify the input neonatal retinal images using ViT, Swin ViT, ResVit, and MobileViT networks. The vanilla ViT converts the images into flattened patches and treats them as a sequence of token embeddings. It utilises a standard transformer structure with self-attention mechanisms to extract the relationships between patches in order to classify the images effectively while leveraging scalable NLP techniques for efficiency. The hierarchical Swin ViT employs shifted windows in the self-attention mechanism, which enables efficient local computations while maintaining linear complexity (Liu et al., 2021). The ResViT is a hybrid architecture ensembling convolution-based CNNs and transformers (Dalmaz et al., 2022). It has Aggregated Residual Transformer (ART) blocks with residual connections, skip connections, and transformer architectures to obtain the local and global features, which improve the contextual understanding of the images. The MobileViT network ensembles the strengths of lightweight CNNs and self-attention-based ViTs in classification tasks (Mehta and Rastegari, 2021).

In the current study, the input neonatal fundus images are classified using a novel lightweight Quantum-based transformer, MobileNet model QMViT, to predict the characteristic stages, zones, and severity of ROP. Quantum Computing (QC) is a rapidly advancing technology that leverages the principles of quantum mechanics to perform computations that can be infeasible for classical computers. Unlike classical computers, which use bits as the fundamental unit of data that exist in a binary state (either 0 or 1), quantum computers use qubits. Qubits can exist in a superposition of states, meaning they can represent both 0 and 1 simultaneously due to the quantum superposition principle. This capability allows quantum computers to perform certain complex calculations much more efficiently than classical computers (Vallero et al., 2024).

Additionally, quantum entanglement—a phenomenon where qubits become interconnected and the state of one qubit can instantaneously affect the state of another, no matter the distance—further enhances the computational power of quantum systems. This entanglement enables more complex data processing and contributes to the potential speedup of quantum algorithms compared to their classical counterparts (Henderson et al., 2020). In the context of Quanvolutional Neural Networks (QNNs), these quantum properties are utilised to enhance the feature extraction process in neural networks, particularly for tasks like image classification. The quantum layers (quanvolutional layers) process image data through quantum circuits, potentially discovering patterns and features that classical layers might miss (Gordienko et al., 2024).

Fig. 5 depicts the architecture of QMViT utilised in classification of stages, zones and plus disease of ROP. The input retinal images are resized to size $128 \times 128 \times 3$ and converted from Cartesian to polar coordinates. The system processes each channel of the image individually, transforms it with respect to the centre, and resizes it to maintain the original dimensions. This enables better analysis of the image with respect to the circular or radial features.

The polar images are given to the quanvolutional layer, which is made up of a 4-qubit quantum device. Polar coordinates emphasize radial and circular structures, aligning better with the geometry of retinal features like the optic disc and blood vessel radiations. This improves spatial inductive bias in early processing layers of the model. Additionally, many of the images have empty space near the corners of the image, and align better with using polar representations since we can filter out data beyond the radius of the eye. The polar input data is encoded into its quantum states through Y-axis rotations on each qubit. A randomly parameterized quantum circuit is applied to the generated data and finally measures the expectation values of the Pauli-Z operator for each qubit, returning feature maps through locally transforming input data. The processed feature maps are provided to a standard convolution layer of stride 3x3 encoding the local spatial information of the input image. This layer is followed by a point-wise convolutional layer, which projects the tensor to a high-dimensional space by capturing the linear combinations of the input channels. The convolutional layers are followed by MobileNetv2 blocks, which are designed to down-sample the feature maps and maintain low computational costs. The feature maps are further taken to the core MobileViT blocks, which integrate the transformer layers into the architecture. These blocks apply global processing instead of local processing present in ordinary convolutions. This lets the model learn local and global representations efficiently. This model utilises the Swish activation function, which has been shown to improve performance in deep learning models. MobileViT network leverages the spatial inductive biases of CNNs and the global processing capabilities of transformers. The study utilises 4 qubits to balance expressiveness and tractability. More qubits increase feature richness but also simulation cost exponentially. On real hardware, increasing qubits might improve results, but introduces decoherence and scalability concerns.

Classical pixel values $p \in [0, 1]$ are normalized to rotation angles using the transformation $\phi = \frac{\pi}{2}p$. Each value ϕ_j corresponding to a pixel is encoded onto qubit j via a single-qubit rotation gate $R_\alpha\left(\phi_j\right)$, where the axis $\alpha \in \{X, Y\}$. This places the qubit into a specific location on the Bloch sphere depending on the chosen axis, resulting in a high-dimensional, nonlinear feature mapping from classical image patches to quantum state space:

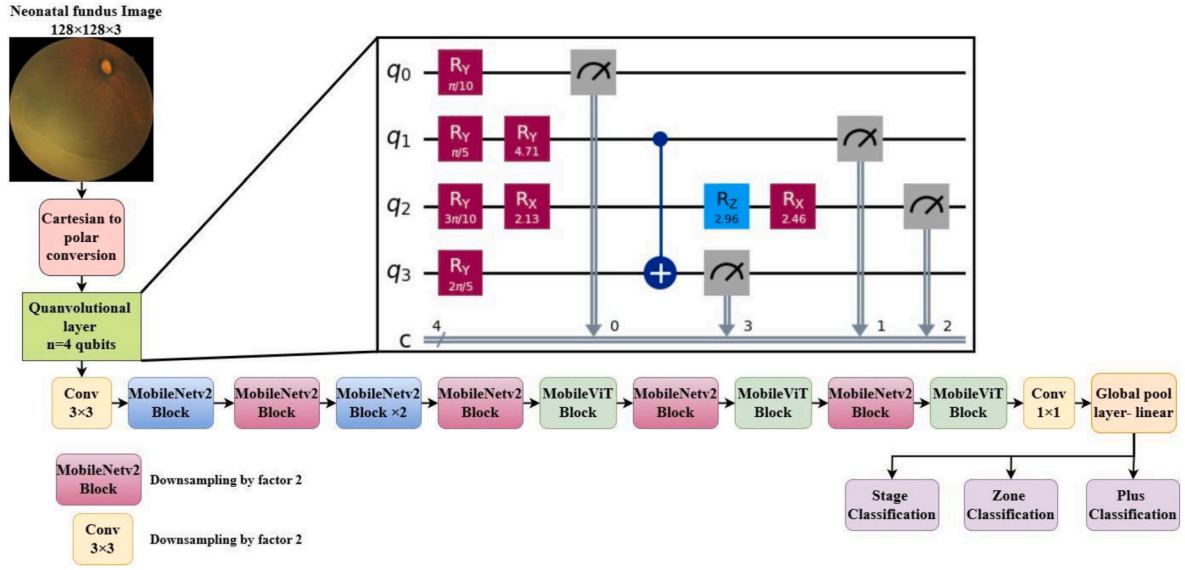

Fig. 5. Architecture of QMViT Network utilised in classification of stages, zones and plus disease of ROP. The quanvolutional layer uses placeholder input values of [0.1, 0.2, 0.3, 0.4] as the pixel intensities, which are converted to Ry rotation gates.

$$\mathscr{T}_{\alpha} : R^4 \rightarrow C^{2^4}. \tag{1}$$

The rotation operations are defined as:

$$R_X(\phi) = e^{-i\phi\sigma_X/2}, \tag{2}$$

$$R_Y(\phi) = e^{-\phi\sigma Y/2} \tag{3}$$

where σ_X and σ_Y are the Pauli X and Y matrices, respectively.

After encoding, a fixed random quantum circuit U_{rand} is applied to entangle and transform the qubits. This circuit is implemented using the *RandomLayers* template from PennyLane (The PennyLane Team, 2023), which generates a hardware-efficient random quantum feature map. For each layer, the circuit applies randomly sampled single-qubit rotations (combinations of R_X, R_Y, and R_Z) followed by randomly selected entangling operations (such as CNOT or CZ) between qubit pairs. The rotation angles are drawn from a uniform distribution over $[0, 2\pi]$ and are fixed once at initialization. This creates a shallow unitary transformation that introduces nonlinear mixing and entanglement over the encoded classical inputs, without requiring trainable quantum parameters.

Finally, the expectation values of the Pauli-Z operator on each qubit are measured, yielding four classical outputs corresponding to:

$$\langle Z_i \rangle = \langle 0 \left| R_{\alpha}^{\dagger} U_{\text{rand}_i Z}^{\dagger} U_{\text{rand}} R_{\alpha} \right| 0 \rangle, \tag{4}$$

Rotations around the X or Y axis modify the amplitude population between the $|0\rangle$ and $|1\rangle$ states and are therefore observable in $\langle Z\rangle$-basis measurements. In contrast, R_Z encodings affect only the phase and do not alter the measurement probabilities in the Z-basis, making them undetectable in this context. Therefore, we restrict our encoder rotation gates to $\{R_X, R_Y\}$.

The use of quanvolution is motivated by prior work (Henderson et al., 2020), which demonstrated that even shallow, random quantum circuits can act as expressive, non-classical kernels. These quantum kernels have the ability to separate complex data distributions more effectively than purely classical transformations, particularly in low-data regimes. We integrated quanvolution, MobileNet, and transformer blocks to combine local nonlinear quantum preprocessing, lightweight CNN feature extraction, and global attention capabilities of transformers. A fully quantum approach is impractical due to the large number of qubits which would be required; a purely classical one underutilises quantum advantages. The QMViT model is executed using the Pennylane library in the Google Colab platform.

The computational complexity of the quanvolutional preprocessing step can be analyzed as a function of the number of image patches, qubits, and circuit depth. Let the input image be of size $N \times N$ with C channels (or spectral bands), and let q denote the number of qubits used in the quantum circuit. The image is convolved with a fixed quantum circuit applied to every non-overlapping 2×2 patch, yielding a total of $\frac{N^2}{4}$ patches per channel. For each patch, the circuit performs a statevector simulation of a q-qubit quantum system with a gate depth D. The PennyLane *RandomLayers* template used in our implementation applies $\mathscr{O}(Dq)$ gates per circuit. The cost of classically simulating a q-qubit quantum circuit with G gates scales as $\mathscr{O}(G \cdot 2^q)$, since the underlying state vector has 2^q complex amplitudes. Therefore, the total runtime of the quanvolutional preprocessing step is:

$$T(N, C, q, D) = \mathscr{O}\left(C \cdot \frac{N^2}{4} \cdot Dq \cdot 2^q\right) \tag{5}$$

In our implementation, we use $q = 4$ qubits and $D = 4$ layers (from a (4×4) random parameter tensor), which results in a tractable constant factor of $2^4 = 16$ per circuit evaluation. This ensures that the simulation remains efficient and scalable for moderately sized images (e.g., 128×128) and channel counts. Whereas the computational complexity of MobileViT is given as

$$T(N, D) = \mathscr{O}(N^2 D) \tag{6}$$

Note that when simulated on a classical computer, the runtime increases exponentially with the number of qubits, which may limit future scalability without access to real quantum hardware. On an actual quantum computer, the exponential cost of simulating a q-qubit circuit would be eliminated, and the runtime would scale linearly with the number of qubits. In that case, the overall complexity of the quanvolutional preprocessing step would reduce to:

$$T(N, C, q, D) = \mathscr{O}\left(C \cdot \frac{N^2}{4} \cdot Dq\right) \tag{7}$$

making the method more practical for larger images and deeper circuits under quantum execution.

All the deep learning networks utilised in the current study are trained and validated with 80 % of the total images and are tested with 20 % of the total images. The split of train, validation and test images is

selected in the current study based on existing literature on the networks used and a trial-and-error basis. The stage classification is performed using 2000 images (400 images for four stages and 400 Normal images), with 1280 images for training, 320 images for validating and 400 images for testing the networks. The classification of zones involves 1600 images (400 images for each zone and 400 Normal images), with 1024, 256 and 320 images used for training, validating, and testing the networks. 600 images are used in the severity classification (200 plus, 200 preplus, and 200 Normal images), which are split into 384, 96 and 120 images for training, validating, and testing the networks. The patch size of the images is set to 2 with Sparse Categorical Cross entropy as the loss function, $1 \times e^{-4}$ as the learning rate, Adam as the optimizer and a batch size of 32.

The proposed work designs two networks, a simple convolution-based classical CNN and a quantum-based CNN, with a similar number of layers and characteristics, to study the nature of quantum-based CNNs in the classification of ROP. The architecture of custom classical CNN and quantum-based CNN is depicted in Fig. 6. The custom classical CNN is made up of four convolutional layers with 1,8,16 and 32 kernels of size 3×3. The four convolutional layers are followed by a dropout (20 %) and max-pooling layers of stride 2 to downsample the feature map. Finally, the flattened layer is used to convert the convolved feature map to a 1-dimensional vector, which is sent to a dense layer of neurons 128. Finally, the features are provided to the final dense classification layers with several classes as the neurons perform three types of ROP classifications: stages, zones, and plus disease. In the case of Quantum CNN, the initial quanvolutional layer with 4 qubits is used as the QMViT network in Fig. 3. The rest of the network is similar to a classical CNN. Both the CNNs are trained for 200 epochs with a batch size of 32, with adam as the optimizer and Sparse Categorical Cross entropy as the loss function.

4. Results and discussions

The proposed study involves the segmentation of the retinal blood vessels from neonatal fundus images, which is trained and tested in a publicly available dataset, HVDPDROP. The results of segmentation using MultiResUNet and Swin UNet are depicted in Fig. 7. The first column of images is the resized original retinal images, and the second column shows the annotations of blood vessels from the original images. The third and fourth column images represent the vessels that are segmented by MultiResUNet and Swin UNet, respectively. It is qualitatively observed that Swin UNet segments the finer vessels much better than MultiResUNet.

The quantitative analysis of retinal vessel segmentation using MultiResUNet and Swin UNet is given in Table 3. The performance metrics appropriate for quantification of segmentation, such as accuracy, precision, sensitivity, dice coefficient and Intersection over Union (IoU) are employed to compare the performance of MultiResUNet and Swin UNet. MultiResUNet segments neonatal retinal vessels with an accuracy of 94.1 %, precision of 96 %, sensitivity of 83.3 %, dice coefficient of 0.89 and IOU of 80.6 %. Whereas Swin UNet outperforms MultiResUNet with an accuracy of 98.4 %, precision of 97.7 %, sensitivity of 97 %, dice coefficient of 0.97 and IOU of 94.8 %.

Tables 4–6 illustrates the results of three ML classifiers in the multiclass classification of ROP of different stages, zones, and plus diseases, respectively. They describe the impact of different high-level features such as SIFT, SURF, ORB, fused SIFT-SURF-ORB features, and PCA reduced principal components on the classification of neonatal retinal vessels. Tables A1–A6 in the appendix gives a detailed description of the results in Tables 4–6 with different combinations of the features such as SIFT-SURF, SIFT-ORB, SURF-ORB and different numbers of PCA components with individual class performance metrics and their macro average values.

It is observed that in all the classification outcomes, SVM outperforms the other two classifiers in both MultiResUNet and Swin UNet segmentations. Fig. 8 illustrates the plot of cumulative variance captured by the PCA components in classifying stages, zones, and severity of ROP using SIFT, SURF, and ORB combined features. It is noted from Fig. 8 that the classification of ROP stages and plus disease, the SIFT-SURF-ORB fused features with dimensions reduced to 25, achieve maximum classification accuracy. But in ROP zone classification, the SIFT-SURF-ORB fused features with dimensions reduced to 20 achieve maximum classification accuracy. This discrepancy is caused by the variations in maximum cumulative variance in the features extracted

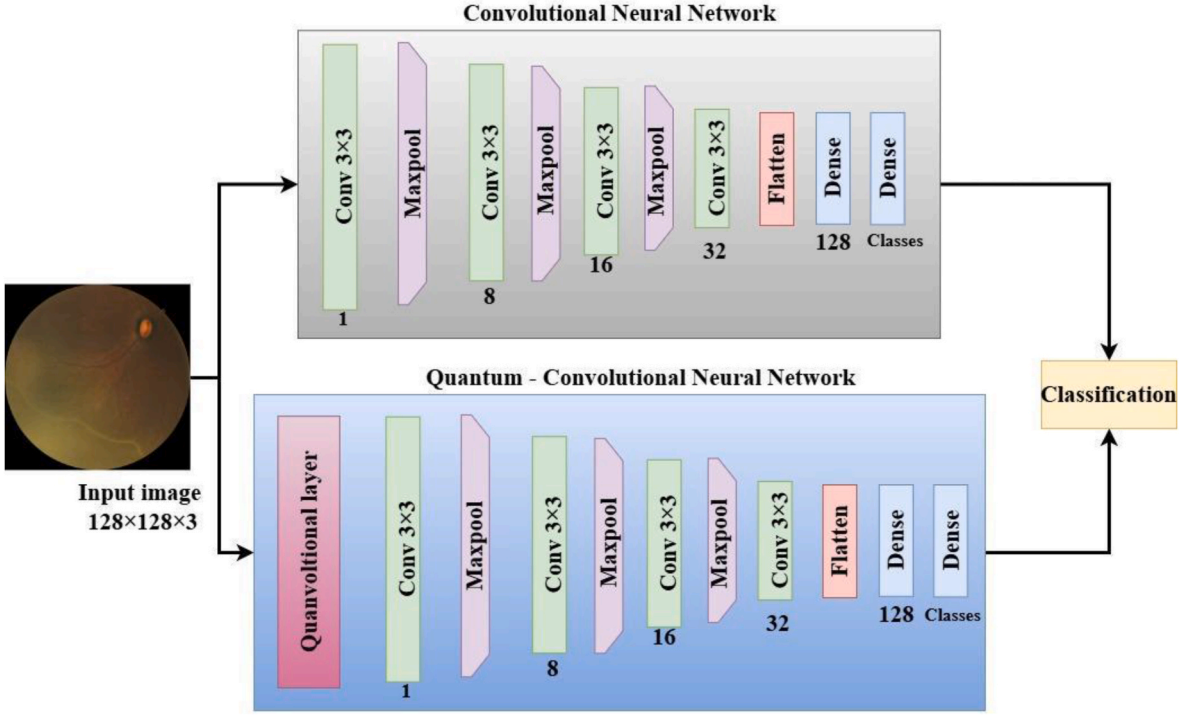

Fig. 6. Architecture of designed custom Classical CNN and Quantum CNN.

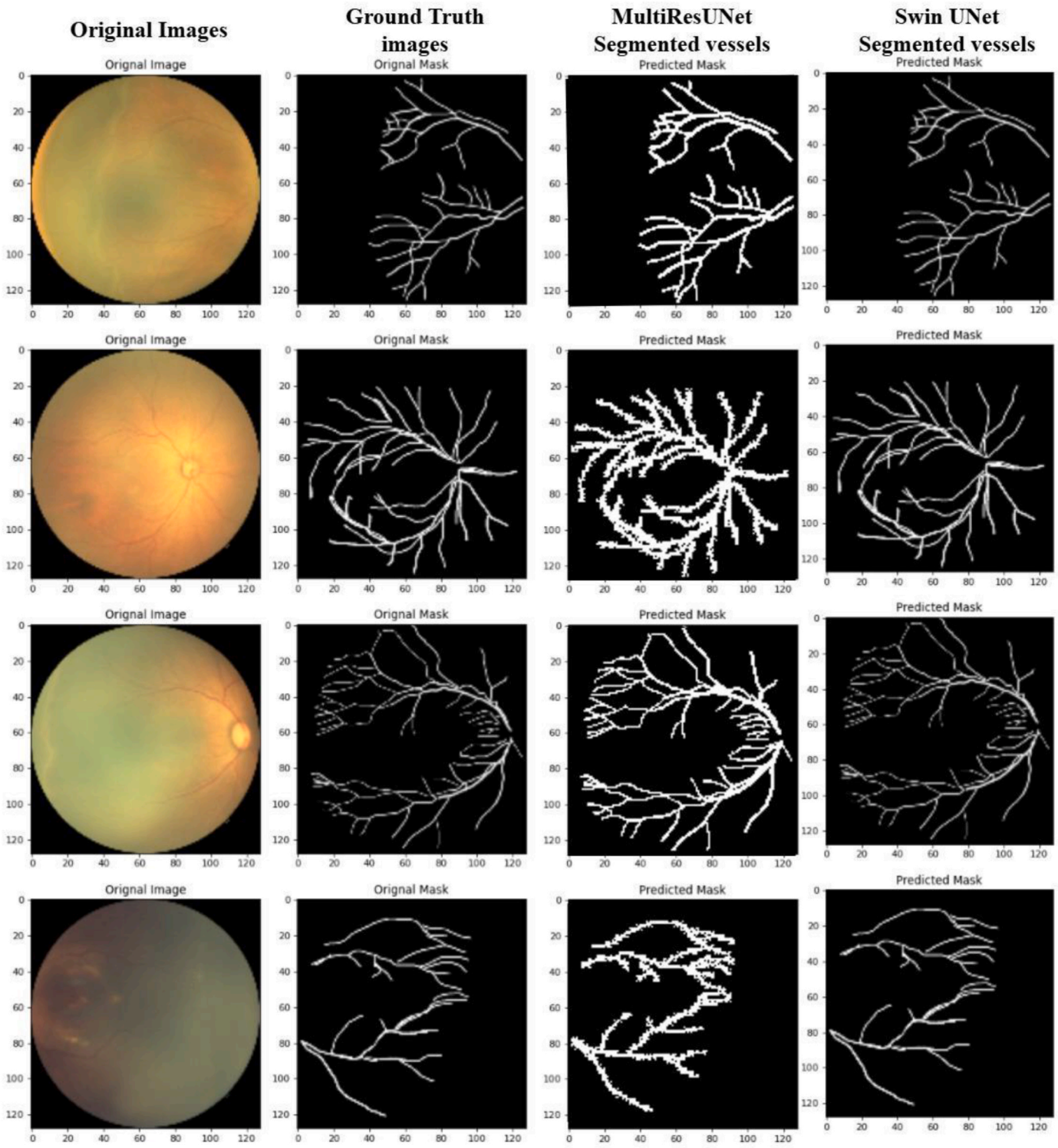

Fig. 7. Segmented blood vessels from the neonatal images using MultiResUNet and Swin UNet: comparison with ground truth images.

from stage, zone and plus diseases of ROP. However, it is noted that in all three classifications, the SIFT-SURF-ORB fused features classify better than individual features, which is further improved by dimensionality reduction by PCA. Also, in all the cases, Swin UNet segmented vessels are classified with maximum accuracy compared to the vessels segmented by MultiResUNet. The qualitative and quantitative segmentation results, along with the ML classification results, reveal that Swin UNet achieves maximum accuracy in the segmentation of retinal blood vessels, which aids in the classification of ROP.

Table 7 displays the performance of different deep learning networks, such as ViT, Swin ViT, ResVit, MobileViT, and Quantum-based QMViT, in classifying different stages, zones, and stages of ROP. Tables A7–A9 in the appendix provide a brief description of Table 7 with estimated individual performance metrics and their macro averages. The QMViT network outperforms other networks in classifying different ROP stages with 95.5 % accuracy, 95.49 % sensitivity, 98.83 % specificity, 95.54 % precision, 98.84 % NPV, and an AUC of 0.97. In classifying different ROP zones, QMViT achieves a maximum accuracy of 96.88 %,

Table 3
Comparison of quantitative analysis of MultiResUNet and Swin UNet in segmentation of blood vessels.

Metrics	MultiResUNet	Swin UNet
Accuracy (%)	94.1	98.4
Precision (%)	96	97.7
Sensitivity (%)	83.3	97.0
Dice coefficient	0.89	97.3
IoU (%)	80.6	94.8

a sensitivity of 96.91 %, a specificity of 98.94 %, a precision of 96.92 %, NPV of 98.93 %, and an AUC of 0.98. The Plus disease of ROP is classified using ViT, Swin ViT, ResVit, MobileViT, and Quantum-based QMViT with an accuracy of 88.33 %, 80 %, 96.67 %, 94.17 %, and 96.67 %, respectively. It is observed that combinations of convolutional-based CNNs and transformers, such as ResVit and MobileViT, perform better than transformer-based vanilla ViT and Swin ViT in all types of ROP classifications. The QMViT outperforms all the other state-of-the-art deep learning networks and ML classifiers with an accuracy of 95.5 %, 96.88 %, and 96.67 % in the classification of stages, zones and plus disease of ROP, respectively.

Fig. 9 illustrates the Receiver Operating Characteristic (ROC) curve plotted for the classification of Plus, zone and stages using QMViT network. Compared to other networks, QMViT achieves maximum AUC of 0.97, 0.98, and 0.97 in the classification of stages, zones and plus disease of ROP, respectively. Table 8 estimates the mean accuracy and loss in validating the QMViT network in Stage, zone and severity classification of ROP using a five-fold cross-validation. Table 9 displays the result of hyper-tuning the better axis of encoding rotation gates (Rx and Ry). It is observed that for all the multiclass classification of ROP, Ry encoding performs better than Rx rotation gate with maximum accuracy and least loss function.

Swish was chosen over other activation functions in QMViT based on literature survey due to its unique combination of smoothness, non-monotonicity, and gradient-preserving behavior, all of which contribute to better expressiveness, stability, and performance in deep hybrid architectures. In quantum-enhanced vision models where subtle spatial and intensity patterns must be retained throughout the network Swish acts as a crucial component in maintaining both computational tractability and clinical accuracy. This ultimately supports improved classification of ROP stages, zones, and severity, even in the presence of limited training data and complex visual cues. Table 10 depicts the analysis of QMViT in classifying Stage, Zone and severity of ROP using different activation functions such as Swish, ReLU and GeLU. It is observed that for all the multiclass classifications of ROP, Swish outperforms other activation functions with improved performance.

Fig. 10 illustrates the training loss obtained in training and validating the proposed network in ROP classification. The analysis of loss curves indicates that the network is trained and validated well in zone classification, having good generalization without any overfitting. Whereas in stage and plus classifications, moderate overfitting occurs, which could be avoided in the future by improving the dataset utilised in the study.

Table 11 compares the results of different types of ROP classification using the designed custom classical CNN and quantum CNN. The quantum CNN achieved an accuracy of 85 %, 77.5 % and 82.5 % in classifying stages, zones and plus disease of ROP in neonatal retinal images. Although the classification by these custom CNNs results in reduced accuracies for all the types of classifications when compared with QMViT, it provides a good insight into the impact of the Quantum CNN in the classification of ROP. The designed custom Quantum CNN improves the accuracy of a classical CNN by about 30.27 %, 2.47 % and 6.45 % in classifications of stages, zones and plus disease of ROP, respectively.

Fig. 11 displays the validation accuracy and loss curves obtained while validating the custom classical and Quantum CNNs in the classification of plus, zone and stages of ROP. The green curve depicts the classical CNN, and the blue curve depicts the Quantum CNN. It is observed that in both accuracy and loss plots, the Quantum CNN

Table 4
Performance of ML classifiers on the classification of different stages of ROP.

Features	Classifier	MultiResUNet						Swin UNet					
		Sensitivity (%)	Specificity (%)	Precision (%)	NPV (%)	Accuracy (%)	AUC	Sensitivity (%)	Specificity (%)	Precision (%)	NPV (%)	Accuracy (%)	AUC
SIFT	SVM	59.74	85.49	59.12	85.89	59.50	0.75	68.76	90.25	68.57	90.58	68.25	0.80
	RF	56.16	83.78	56.60	83.85	56.00	0.73	60.25	86.17	60.96	86.07	59.50	0.75
	KNN	40.85	75.13	47.54	73.62	41.00	0.63	42.15	77.07	57.27	74.80	41.75	0.64
SURF	SVM	54.15	82.34	53.04	82.86	53.00	0.71	66.48	89.73	67.09	89.54	66.75	0.79
	RF	42.25	75.38	44.02	74.77	42.25	0.64	64.70	88.74	65.73	88.43	65.00	0.78
	KNN	46.48	78.85	47.95	78.45	46.75	0.67	53.35	83.31	57.60	82.31	52.75	0.71
ORB	SVM	55.97	83.15	55.53	83.48	54.25	0.72	69.22	90.49	68.57	90.71	69.00	0.81
	RF	34.53	67.85	35.30	67.48	33.75	0.59	51.41	81.11	51.50	81.37	50.75	0.70
	KNN	49.32	79.05	48.23	79.51	48.00	0.68	60.73	87.35	61.71	87.04	60.75	0.76
SIFT-SURF-ORB	SVM	64.18	88.10	63.81	88.49	64.50	0.78	72.64	91.74	71.90	91.98	72.50	0.83
	RF	44.89	77.05	45.34	76.94	45.25	0.66	57.28	85.13	56.83	85.57	57.75	0.73
	KNN	53.44	83.47	55.07	83.04	53.75	0.71	66.53	89.51	67.02	89.34	66.00	0.79
SIFT-SURF-ORB PCA-20	SVM	58.20	84.85	57.60	85.03	57.25	0.74	72.70	91.58	72.84	91.55	71.75	0.83
	RF	49.86	80.44	51.89	79.83	49.00	0.69	71.27	90.89	70.95	90.98	70.50	0.82
	KNN	52.94	81.80	51.66	82.38	51.25	0.70	66.58	89.32	66.42	89.54	65.50	0.79
SIFT-SURF-ORB PCA-25	**SVM**	**66.31**	**88.75**	**66.36**	**88.83**	**65.50**	**0.79**	**75.22**	**93.37**	**75.44**	**93.34**	**76.25**	**0.85**
	RF	53.23	82.06	52.96	82.22	52.50	0.71	69.99	91.38	70.10	91.44	71.00	0.81
	KNN	57.66	84.89	59.50	84.94	56.75	0.73	70.93	91.88	71.07	91.91	72.00	0.82

ML classification is carried out using three classifiers: SVM, RF, and *k*-NN. The SIFT-SURF-ORB fused features extracted from Swin UNet segmented retinal vessels, with PCA reduced 25 components classified using the SVM classifier, achieve a maximum accuracy of 76.25 %, a sensitivity of 75.22 %, specificity of 93.37 %, precision of 75.44 %, NPV of 93.34 %, and an AUC of 0.85 in classification of different stages of ROP. In ROP-zone classification, SIFT-SURF-ORB fused features, reduced to 20 principal components from Swin UNet segmentation, achieved a maximum accuracy of 77.81 %, a sensitivity of 77.53 %, specificity of 91.59 %, precision of 77.82 %, Negative Predictive Value (NPV) of 91.84 %, and an AUC of 0.85 when classified using SVM classifier. In plus disease classification, SVM classifies the SIFT-SURF-ORB fused features reduced to 25 principal components from Swin UNet segmented vessels with a maximum of 98.33 % accuracy, 98.21 % sensitivity, 99.19 % specificity, 98.21 % precision, 99.19 % NPV and an AUC of 0.99.

Table 5
Performance of ML classifiers on the classification of different zones of ROP.

Block	Classifier	MultiResUNet						SwinUNet					
		Sensitivity (%)	Specificity (%)	Precision (%)	NPV (%)	Accuracy (%)	AUC	Sensitivity (%)	Specificity (%)	Precision (%)	NPV (%)	Accuracy (%)	AUC
SIFT	SVM	70.79	87.68	70.44	88.37	70.31	0.80	70.90	88.14	69.42	89.29	71.25	0.81
	RF	61.78	82.56	60.82	83.10	61.25	0.74	58.09	80.57	57.21	82.18	57.81	0.72
	KNN	62.01	83.19	63.00	83.38	61.25	0.75	44.40	73.72	52.97	71.16	45.00	0.63
SURF	SVM	69.53	87.12	69.90	88.68	69.69	0.80	68.71	87.42	68.37	88.83	70.31	0.80
	RF	56.82	79.59	54.57	81.18	56.88	0.71	64.59	85.49	63.52	86.56	65.94	0.77
	KNN	59.01	81.73	59.16	83.08	59.06	0.73	57.73	81.68	55.96	82.98	59.69	0.72
ORB	SVM	60.21	82.69	58.85	83.62	60.94	0.74	73.03	89.12	72.85	90.04	73.44	0.82
	RF	44.31	70.74	43.20	71.54	45.00	0.63	47.89	73.23	46.09	74.31	47.81	0.65
	KNN	58.18	80.82	57.18	83.65	59.38	0.72	64.78	84.70	65.05	86.18	64.69	0.77
SIFT-SURF-ORB	SVM	74.89	89.67	74.67	90.31	74.38	0.83	75.39	90.80	75.63	91.62	76.88	0.84
	RF	47.52	72.96	46.96	73.47	46.88	0.65	51.73	77.04	50.97	77.72	53.13	0.68
	KNN	61.34	81.66	59.78	84.61	60.00	0.74	62.85	84.49	64.70	86.84	65.31	0.76
SIFT-SURF-ORB	**SVM**	**75.78**	**90.43**	**75.56**	**90.49**	**75.31**	**0.84**	**77.53**	**91.59**	**77.82**	**91.84**	**77.81**	**0.85**
PCA-20	RF	63.26	83.63	62.12	84.17	62.50	0.75	71.54	89.05	71.15	89.34	72.19	0.81
	KNN	74.49	89.57	73.28	90.13	73.75	0.83	74.22	90.29	74.17	90.83	75.00	0.83
SIFT-SURF-ORB	SVM	72.00	88.22	70.29	89.08	70.94	0.81	75.07	90.37	74.90	90.85	75.31	0.83
PCA-25	RF	62.91	83.02	61.48	83.88	61.88	0.75	69.08	87.32	68.71	88.07	69.69	0.79
	KNN	74.73	89.55	73.42	90.15	73.75	0.83	76.56	90.94	75.90	91.66	76.88	0.84

Table 6
Performance of ML classifiers on the classification of severity of ROP.

Block	Classifier	MultiResUNet						Swin UNet					
		Sensitivity (%)	Specificity (%)	Precision (%)	NPV (%)	Accuracy (%)	AUC	Sensitivity (%)	Specificity (%)	Precision (%)	NPV (%)	Accuracy (%)	AUC
SIFT	SVM	85.00	92.57	85.06	92.04	85.00	0.85	83.15	90.63	82.89	90.48	82.50	0.87
	RF	73.77	85.86	73.75	85.78	75.00	0.81	70.76	82.41	70.46	83.05	70.00	0.78
	KNN	74.86	87.52	78.20	89.59	78.33	0.84	74.73	84.96	75.33	86.82	74.17	0.81
SURF	SVM	86.95	92.85	86.69	92.99	86.67	0.90	81.08	90.24	81.08	90.15	81.67	0.86
	RF	75.66	85.88	75.50	86.47	75.00	0.82	73.04	85.23	72.97	85.64	74.17	0.80
	KNN	74.48	85.57	80.09	87.58	73.33	0.81	78.11	88.70	80.60	88.96	79.17	0.84
ORB	SVM	83.35	91.61	83.40	91.47	84.17	0.88	85.23	92.52	85.18	92.55	85.83	0.89
	RF	58.95	75.63	58.95	75.63	60.83	0.71	59.58	75.27	59.75	76.03	60.83	0.70
	KNN	75.00	86.93	75.80	87.21	76.67	0.82	81.99	90.76	82.89	91.49	83.33	0.87
SIFT-SURF-ORB	SVM	89.16	94.20	89.32	94.27	89.17	0.92	91.69	95.62	91.82	95.85	91.67	0.94
	RF	72.24	83.43	71.31	83.98	71.67	0.79	76.81	86.64	76.89	87.89	76.67	0.83
	KNN	81.87	90.30	87.04	91.55	82.50	0.86	86.66	92.95	87.87	93.17	86.67	0.90
SIFT-SURF-ORB	SVM	93.55	96.37	92.93	95.96	92.50	0.95	94.56	97.23	94.04	96.98	94.17	0.96
PCA-20	RF	81.97	90.02	81.35	89.40	80.83	0.86	89.19	94.69	89.67	94.99	90.00	0.92
	KNN	89.35	94.18	88.12	93.67	88.33	0.92	93.33	96.67	93.27	96.56	93.33	0.95
SIFT-SURF-ORB	**SVM**	**96.75**	**98.39**	**96.75**	**98.30**	**96.67**	**0.98**	**98.21**	**99.19**	**98.21**	**99.19**	**98.33**	**0.99**
PCA-25	RF	86.09	92.57	85.79	92.39	85.83	0.90	90.92	95.25	90.67	95.16	90.83	0.93
	KNN	93.85	96.71	93.48	96.57	93.33	0.95	94.76	97.50	94.71	97.48	95.00	0.96

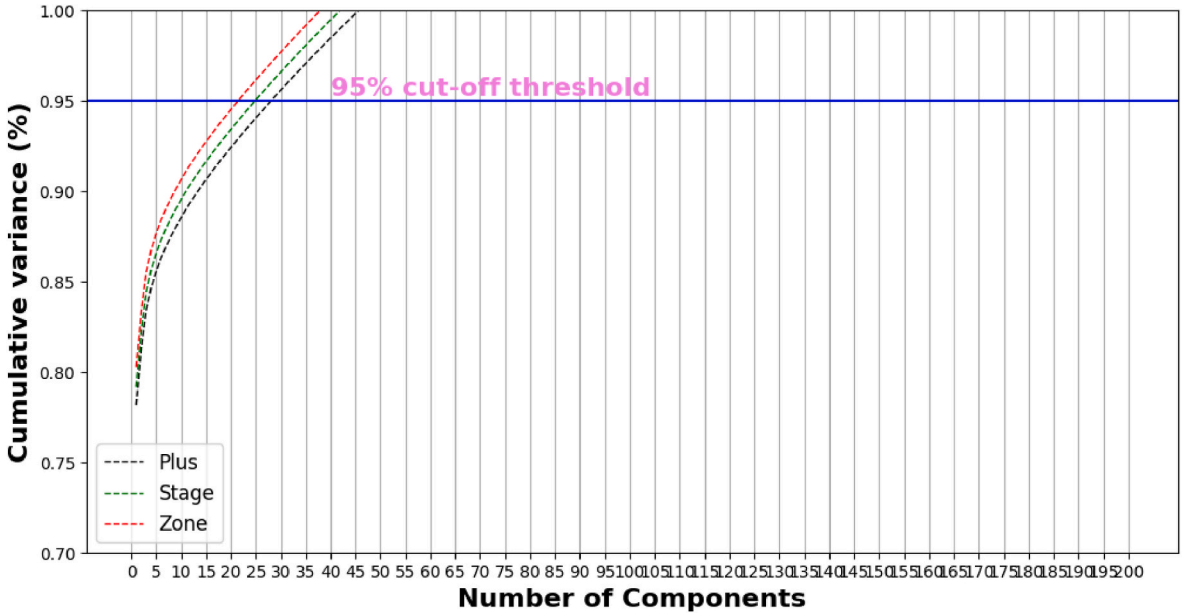

Fig. 8. Cumulative Variance plot for the PCA reduced components in classification of Stages, Zone and severity of ROP.

Table 7
Performance of Deep learning networks on the Classification of ROP into Characteristic Stages, Zones and Severity.

Classification	Network	Sensitivity (%)	Specificity (%)	Precision (%)	NPV (%)	Accuracy (%)	AUC
Stage	ViT	73.37	92.56	73.65	92.57	74.5	0.84
	SwinViT	77.25	93.27	77.59	93.42	77.25	0.96
	ResViT	72.62	92.64	70.92	93.08	75	0.79
	MobileViT	93.1	98.16	93.09	98.18	93	0.95
	QMViT	**95.49**	**98.83**	**95.54**	**98.84**	**95.5**	**0.97**
Zone	ViT	80.63	92.99	80.47	93.26	80.63	0.87
	SwinViT	84.69	94.75	85.19	94.74	84.69	0.96
	ResViT	85.88	94.73	85.93	94.74	85	0.9
	MobileViT	94.68	98.28	95.42	98.38	95	0.96
	QMViT	**96.91**	**98.94**	**96.92**	**98.93**	**96.88**	**0.98**
Plus	ViT	88.33	93.77	88.53	94.06	88.33	0.91
	SwinViT	81.81	90.1	81.81	90.46	80	0.98
	ResViT	97.22	98.35	96.23	98.25	96.67	0.98
	MobileViT	94.18	96.95	94.22	97.02	94.17	0.96
	QMViT	**96.48**	**98.30**	**96.85**	**98.33**	**96.67**	**0.97**

performs better than classical CNN with minimum fluctuations and stability in plus and stage classification. Surprisingly, in Zone classification, although the accuracy remains comparable for both networks, the loss is minimized better by the classical CNNs. Deeper evaluations of the networks with more qubits can improve the performance in zone classification.

Kumar et al. deployed U-Net to segment neonatal retinal vessels with a dice co-efficient of 0.64 and classify ROP with 88.23 % accuracy (Kumar et al., 2023). However, the Swin UNet employed in the current study segments the retinal blood vessels with a higher dice coefficient of 0.97, resulting in a classification accuracy of 98.33 %. The proposed study enhances plus and preplus disease prediction with an AUC of 0.97, compared to Coyner et al. (2022).

Agrawal et al. utilised the same dataset HVDROPDB-BV used in the current study to segment the retinal blood vessels using three networks: U Net, attention gates (AG) U Net, and squeeze and excitation (SE) U Net (Agrawal et al., 2022). The maximum dice coefficient obtained by Agrawal et al. is 0.687 using AG UNet, which is improved to 0.97 in this proposed work using Swin UNet. Although the dice coefficient is improved in the current work, stage classification using segmented vessels, demarcation lines and ridges is 97 %, which could not be achieved. Hence, the segmentation of demarcation lines and ridges along with the blood vessels and the utilisation of RetCam images can improve diagnostic accuracy in ML classification.

Rahim et al. classified three stages (1–3), zones (I-III) and plus disease using ResNet50 and InceptionResV2 CNN (Rahim et al., 2023). The RetCam images are pre-processed using Pixel Colour Amplification and Double Pass Fundus Reflection and combined with contrast-limited adaptive histogram equalisation (CLAHE). The authors achieved maximum accuracy of 92 %, 90 % and 97 % in classifying stages, zones and plus disease, respectively. The proposed work utilises state-of-the-art algorithms and improves the accuracy to 95.5 %, 96.88 % and 96.67 % in classifying stages, zones and plus disease, respectively.

The study by Sankari et al. involves extracting SIFT and SURF features from SegNet segmented retinal vessels and classifying them with quantum-based ML classifier like QSVM (Sankari et al., 2023). The method performed binary classification to predict the presence of ROP with an accuracy of 95.5 % and a sensitivity of 93 %. The proposed work performs multiclass classification using a quantum-based transformer network with better computational cost and performance. Yenice et al. predicted the development of ROP in twin preterm infants using variables such as gender, GA, postmenstrual age at examination, birth weight, discordance rate, ROP Stages and Zones (Yenice et al., 2023). Future work can involve predicting the development of ROP in preterm infants based on clinical features using artificial intelligence, which can aid in improved treatment outcomes.

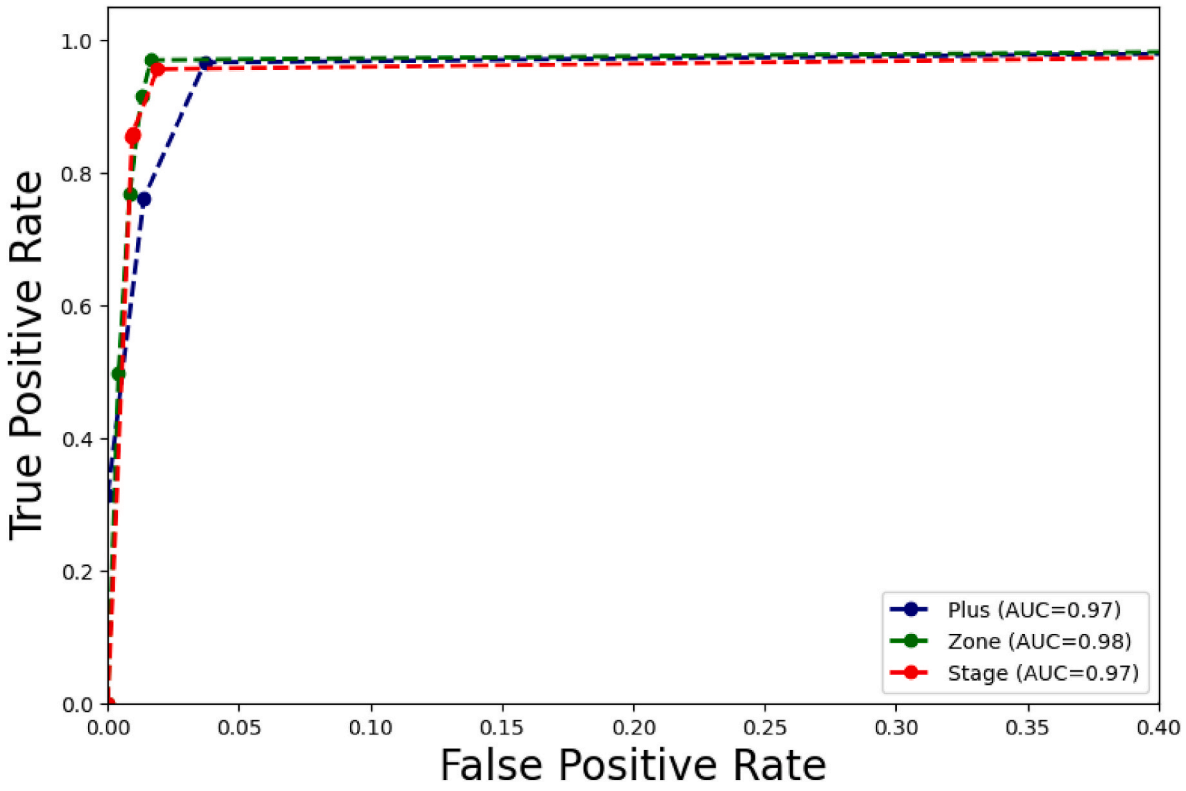

Fig. 9. ROC curves plotted for classification of different Stages, Zones and plus disease of ROP using QMViT Network.

Table 8
Performance of QMViT network in classifying Stage, Zone and severity of ROP using Five-Fold Cross-Validation.

Cross Validation	Stage		Zone		Plus	
	Accuracy	Loss	Accuracy	Loss	Accuracy	Loss
Fold 0	0.64	1.92	0.8	0.83	0.91	0.47
Fold 1	0.65	2.01	0.78	0.88	0.91	0.31
Fold 2	0.66	1.66	0.79	1.22	0.9	0.44
Fold 3	0.67	1.87	0.8	0.87	0.9	0.44
Fold 4	0.62	2.11	0.81	1.11	0.93	0.29
Mean	0.65	1.92	0.80	0.98	0.91	0.39

Table 9
Performance of QMViT network in classifying Stage, Zone and severity of ROP using two encoding rotation gates Rx and Ry.

Classification	Augment	Validation Accuracy	Validation Loss
Stage	Rx	0.63	2.06
	Ry	0.65	1.92
Zone	Rx	0.79	1.09
	Ry	0.78	0.98
Plus	Rx	0.89	0.50
	Ry	0.91	0.39

Luo et al. utilised an edge-cloud ensemble network based on deep learning approaches to diagnose ROP (Luo et al., 2023). Although the system achieves a reduced accuracy of 60 % with an AUC of 0.75, it uses a telemedicinal approach to apply in rural areas effectively. The integration of an edge-cloud ensemble network can improve the efficiency of the proposed work and help achieve the diagnosis of ROP in rural areas.

The current study proposes a hybrid quantum network that can accurately predict the stage, zone, and severity of ROP. It does not require any preceding segmentation network and complex preprocessing networks to handle the images. The proposed network is lightweight,

Table 10
Performance of QMViT network in classifying Stage, Zone and severity of ROP using different activation functions.

Classification	Augment	Validation Accuracy	Validation Loss
Stage	Swish	0.65	1.92
	ReLU	0.62	2.05
	GeLU	0.56	2.24
Zone	Swish	0.80	0.98
	ReLU	0.74	1.52
	GeLU	0.65	1.74
Plus	Swish	0.91	0.39
	ReLU	0.81	0.85
	GeLU	0.84	0.93

less complex, and highly efficient, utilising the advantages of CNNs, transformers, and Quantum computing.

5. Conclusion

The proposed work delves into the classification of different stages, zones and severity of ROP in fundus images of preterm infants using ML, transformers and quantum-based transformer networks. The retinal blood vessels were segmented using Swin UNet and MultiResUNet, from which 1000 SIFT, SURF and ORB features are extracted, fused and dimensionally reduced to 25 components using PCA. Three ML classifiers are used to classify the 25 principal components into different stages, zones and severity of ROP. On the other hand, transformer-based networks such as ViT, Swin Transformer, ResViT, MobileViT and Quantum-based QMViT process the raw retinal images and classify them directly into characteristic stages, zones and Plus disease of ROP. The SIFT-SURF-ORB fused features extracted from Swin UNet segmented retinal vessels, with PCA reduced components classified using the SVM classifier, achieve a maximum accuracy of 76.25 %, 77.81 % and 98.33 % accuracy in classification of different stages, zones, plus disease of ROP, respectively. The novel QMViT outperforms all the other state-of-

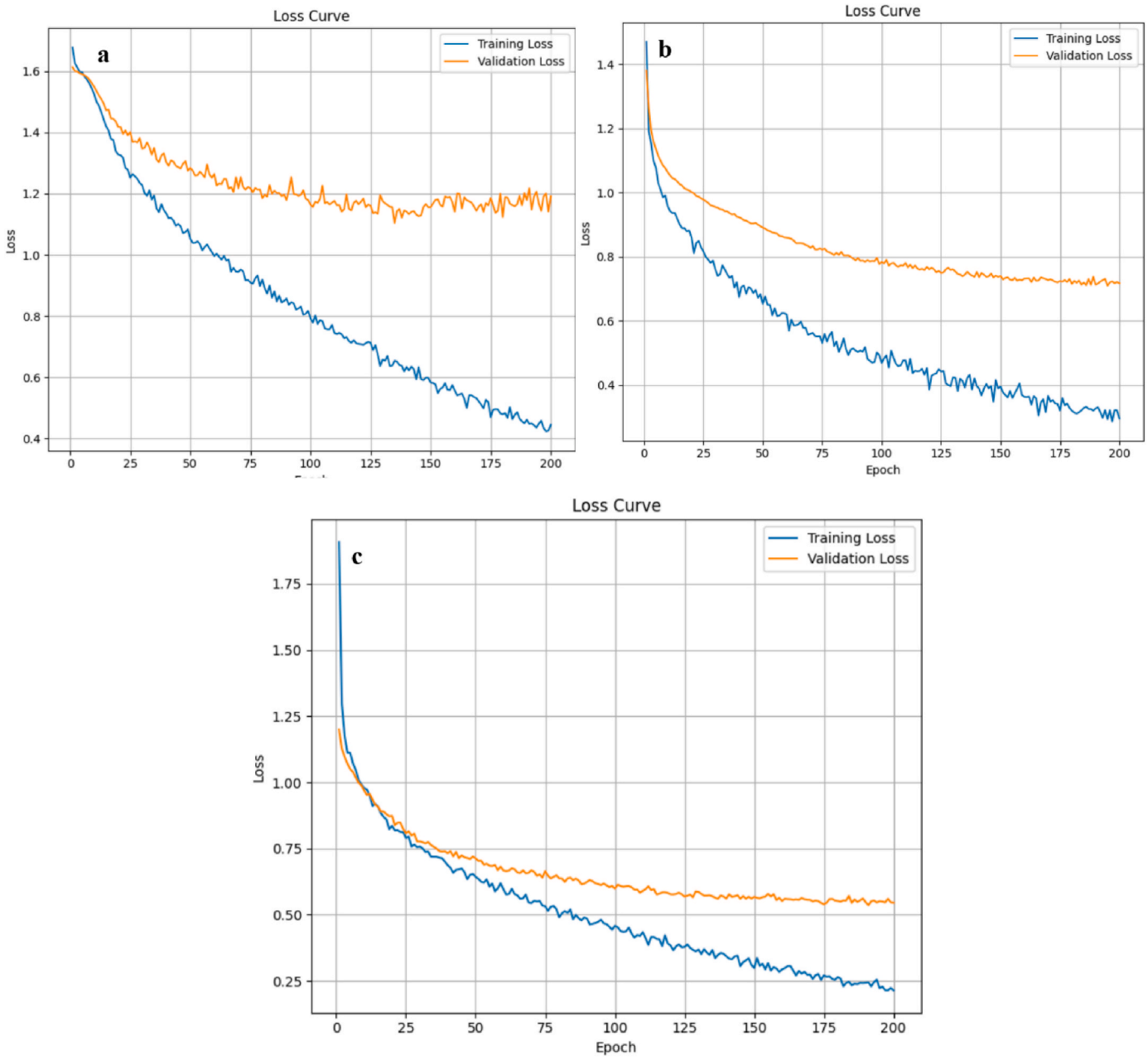

Fig. 10. Loss curves plotted in training and validating the QMViT network in classifying a) Stage, b) Zone, and c) severity of ROP.

Table 11
Comparison of classical and quantum Deep learning networks on the classification of ROP into characteristic stages, zones and severity.

Classification	Network	Sensitivity(%)	Specificity (%)	Precision (%)	NPV (%)	Accuracy (%)	AUC
Stage	Classical CNN	65.32	88.62	68.85	88.64	65.25	0.78
	Quantum CNN	**84.76**	**95.76**	**85.47**	**95.82**	**85.00**	**0.90**
Zone	Classical CNN	74.49	90.69	74.10	90.99	75.63	0.83
	Quantum CNN	**78.22**	**91.53**	**77.55**	**91.75**	**77.50**	**0.85**
Plus	Classical CNN	78.82	88.19	80.31	88.05	77.50	0.86
	Quantum CNN	**82.36**	**90.44**	**85.01**	**90.35**	**82.50**	**0.87**

the-art deep learning networks and ML classifiers, with an accuracy of 95.5 %, 96.88 %, and 96.67 % in classifying stages, zones, and plus disease of ROP, respectively. The proposed work has certain limitations: Firstly, the system detects and classifies ROP that is already present and cannot predict the development of ROP. Hence, predicting the progression of ROP can improve the medical attention necessary for the preterm. Secondly, the network design is trained and tested using limited images collected by the same camera from the same locality. This questions the generalisability of the network. Future work can involve a more generalised dataset of RetCam images collected from various geographic locations. Thirdly, it is important to note that the features extracted by the quanvolutional layer are not clinically

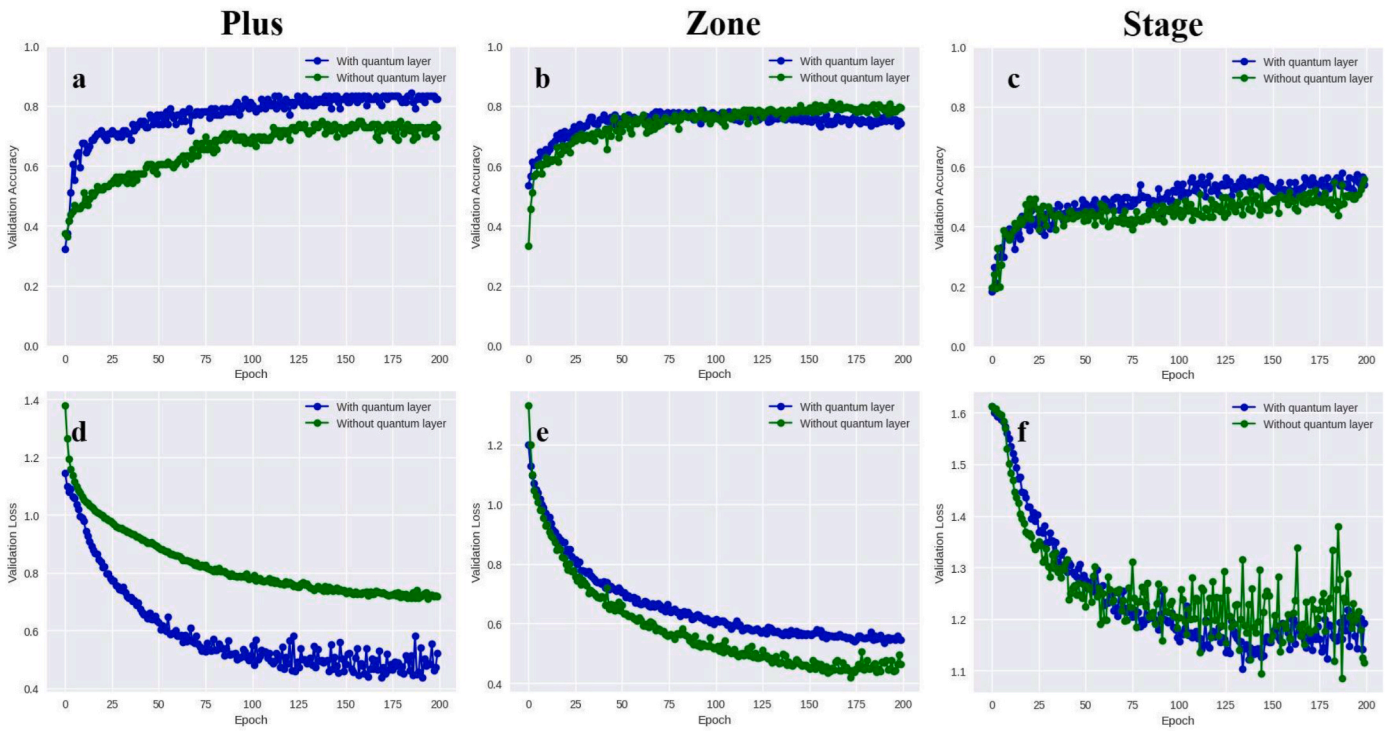

Fig. 11. Validation performances of the custom CNN with and without quantum layer: Validation accuracy curve of a) Plus, b) Zone and c) Stage classification; Validation loss curve of d) Plus, e) Zone and f) Stage classification.

interpretable in a conventional sense. Due to the use of fixed, randomly parameterized quantum circuits, the output features do not correspond to human-recognizable anatomical or pathological structures of retina. While this limits direct interpretability for ophthalmologists, these features can still improve classification performance when used in conjunction with more transparent modules (e.g., attention maps or Grad-CAM applied to later layers in the network). Future work can aim at segmenting demarcation lines, ridges, and blood vessels using RetCam images, which can improve diagnostic accuracy in ML classification. Future work could explore replacing the random circuit with trainable or structured quantum circuits to improve interpretability while retaining quantum advantages. Also, utilising AI to predict the development of ROP in premature babies based on clinical features could enhance the treatment outcomes. Additionally, an edge-cloud ensemble network integrated with the proposed network can increase the efficiency of ROP diagnosis in rural areas. Thus, the developed system can be used as an automated technique in the ancillary diagnosis of the ROP mass population of infants.

CRediT authorship contribution statement

Raja Sankari VM: Writing – original draft, Software, Methodology, Investigation, Data curation, Formal analysis, Conceptualization. **Snekhalatha Umapathy:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Investigation, Data curation, Conceptualization. **Ashok Chandrasekaran:** Writing – review & editing, Resources, Data curation. **Prabhu Baskaran:** Writing – review & editing, Validation, Resources. **Varun Dhanraj:** Writing – review & editing, Visualization, Validation.

Financial Support

None

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors thank the Department of Neonatology of SRM Hospital and Medical College and Research Centre, Kattankulathur, Tamil Nadu, India, and Aravind Hospital, Poonamalle, Tamil Nadu, India, for supporting the research through data collection and annotations.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.engappai.2025.111938.

Data availability

Data will be made available on request.

References

Agrawal, R., Kulkarni, S., Walambe, R., Kotecha, K., 2021. Assistive framework for automatic detection of all the zones in retinopathy of prematurity using deep learning. J. Digit. Imag. 34, 932–947.

Agrawal, R., Kulkarni, S., Walambe, R., Deshpande, M., Kotecha, K., 2022. Deep dive in retinal fundus image segmentation using deep learning for retinopathy of prematurity. Multimed. Tool. Appl. 81, 11441–11460.

Agrawal, R., Walambe, R., Kotecha, K., Gaikwad, A., Deshpande, C.M., Kulkarni, S., 2023. HVDROPDB datasets for research in retinopathy of prematurity. Data Brief 52, 109839.

Ahuja, A.A., Reddy, Y.C.V., Adenuga, O.O., et al., 2018. Risk factors for retinopathy of prematurity in a district in South India: a prospective cohort study. Oman J. Ophthalmol. 11, 33–37.

Bajwa, J., Munir, U., Nori, A., Williams, B., 2021. Artificial intelligence in healthcare: transforming the practice of medicine. Future Healthcare Journal 8.

Bay, H., Tuytelaars, T., Van Gool, L., 2008. Speeded-up robust features (SURF). Comput. Vis. Image Understand. 110, 346–359.
Bowe, T., Nyamai, L., Ademola-Popoola, D., et al., 2019. The current state of retinopathy of prematurity in India, Kenya, Mexico, Nigeria, Philippines, Romania, Thailand, and Venezuela. Digit. J. Ophthalmol. 25, 49–58.
Calonder, M., Lepetit, V., Strecha, C., Fua, P., 2010. BRIEF: binary robust independent elementary features. Computer Vision – ECCV 2010, 778–792.
Chen, J., et al., 2021. Transunet: transformers make strong encoders for medical image segmentation. CoRR, 04306 abs/2102.
Chiang, M.F., Jiang, L., Gelman, R., Du, Y.E., Flynn, J.T., 2007. Interexpert agreement of plus disease diagnosis in retinopathy of prematurity. Arch. Ophthalmol. 125, 875–880.
Chiang, M.F., Jiang, L., Gelman, R., Du, Y.E., Flynn, J.T., 2021. International classification of retinopathy of prematurity. Ophthalmology 128. Third Edition.
Coyner, A.S., et al., 2022. Synthetic medical images for robust, privacy-preserving training of artificial intelligence: application to retinopathy of prematurity diagnosis. Ophthalmology Science. 2, 100126.
Crincoli, E., Sacconi, R., Querques, L., Querques, G., 2024. Artificial intelligence in age-related macular degeneration: state of the art and recent updates. BMC Ophthalmol. 24, 121.
Cunningham, P., Delany, S.J., 2021. k-Nearest neighbour classifiers - a tutorial. ACM Comput. Surv. 54, 1–25.
Cutler, A., Cutler, D.R., Stevens, J.R., 2012. Random Forests. Ensemble Machine Learning, pp. 157–175.
Dalmaz, O., Yurt, M., Cukur, T., 2022. ResViT: residual vision transformers for multimodal medical image synthesis. IEEE Trans. Med. Imag. 41, 2598–2614.
Deng, X., et al., 2023. Vessels characteristics in familial exudative vitreoretinopathy and retinopathy of prematurity based on deep convolutional neural networks. Frontiers in Pediatrics 11, 1252875.
Dosovitskiy, A., et al., 2010. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale, 11929. *arXiv*. 2020.
Fares, A., Abdelmonaim, S., Sayed, D., et al., 2024. Validation of WINROP algorithm as screening tool of retinopathy of prematurity among Egyptian preterm neonates. Eye 38, 1562–1566.
Freitas, A.M., Mörschbächer, R., Thorell, M.R., Rhoden, E.L., 2018. Incidence and risk factors for retinopathy of prematurity: a retrospective cohort study. International Journal of Retina and Vitreous 4, 20.
Gilbert, C., 2008. Retinopathy of prematurity: a global perspective of the epidemics, population of babies at risk and implications for control. Early Hum. Dev. 84, 77–82.
Good, W.V., 2004. Final results of the Early Treatment for Retinopathy of Prematurity (ETROP) randomized trial. Trans. Am. Ophthalmol. Soc. 102, 233–248.
Gordienko, Y., Trochun, Y., Stirenko, S., 2024. Multimodal quanvolutional and convolutional neural networks for multi-class image classification. Big Data and Cognitive Computing 8, 75.
Grzybowski, A., et al., 2020. Artificial intelligence for diabetic retinopathy screening: a review. Eye (Lond.). 34, 451–460.
Hearst, M.A., et al., 1998. Support vector machines. IEEE Intelligent Systems and Applications 13, 18–28.
Hellström, A., Smith, L.E., Dammann, O., 2013. Retinopathy of prematurity. Lancet 382, 1445–1457.
Henderson, M., Shakya, S., Pradhan, S., 2020. Quanvolutional neural networks: powering image recognition with quantum circuits. Quantum Machine Intelligence 2, 2.
Ibtehaz, N., Rahman, M.S., 2020. MultiResUNet: rethinking the U-Net architecture for multimodal biomedical image segmentation. Neural Netw. 121, 74–87.
Jang, J.H., 2024. Characteristics of retinal vascularization in reactivated retinopathy of prematurity requiring treatment and clinical outcome after reinjection of ranibizumab. Sci. Rep. 14, 15647.
Jemshi, K.M., Sreelekha, G., Sathidevi, P., et al., 2024. Plus disease classification in Retinopathy of Prematurity using transform based features. Multimed. Tool. Appl. 83, 861–891.
Kalpathy-Cramer, J., et al., 2016. Plus disease in retinopathy of prematurity: improving diagnosis by ranking disease severity and using quantitative image analysis. Ophthalmology 123, 2345–2351.
Kim, S.J., Port, A.D., Swan, R., Campbell, J.P., Chan, R.V.P., Chiang, M.F., 2018. Retinopathy of prematurity: a review of risk factors and their clinical significance. Surv. Ophthalmol. 63, 618–637.
Kumar, V., Patel, H., Paul, K., Azad, S.V., 2023. Deep learning-assisted Retinopathy of Prematurity (ROP) screening. ACM Transactions on Computing for Healthcare. 4, 1–32.
Le, P.Q., Dong, F., Hirota, K., 2011. A flexible representation of quantum images for polynomial preparation, image compression, and processing. Quant. Inf. Process. 10 (1), 63–84.
Lever, J., Krzywinski, M., Altman, N., 2017. Principal component analysis. Nat. Methods 14, 641–642.
Liu, Z., et al., 2021. Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows. IEEE/CVF International Conference on Computer Vision (ICCV), pp. 9992–10002.
Liu, Y., et al., 2023. An artificial intelligence system for screening and recommending the treatment modalities for retinopathy of prematurity. Asia Pacific Journal of Ophthalmology 12, 468–476.
Lowe, D.G., 2004. Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vis. 60, 91–110.
Luo, Z., Ding, X., Hou, N., Wan, J., 2023. A deep-learning-based collaborative edge–cloud telemedicine System for retinopathy of prematurity. Sensors 23, 276.
Majumder, A., Panigrahi, P.K., Pal, A., 2023. Quantum algorithms for image compression. Computers 12 (8), 185.
Mehta, S., Rastegari, M., 2021. MobileViT: Light-weight, general-purpose, and mobile-friendly vision transformer. arXiv., 02178 abs/2110.
Rahim, S., Sabri, K., Ells, A., et al., 2023. Novel Fundus image preprocessing for retcam images to improve deep learning classification of retinopathy of prematurity. arXiv preprint., 230202524
Rao, D.P., et al., 2023. Development and validation of an artificial intelligence based screening tool for detection of retinopathy of prematurity in a South Indian population. Frontiers in Pediatrics 11, 1197237.
Ronneberger, O., Fischer, P., Brox, T., 2015. U-Net: Convolutional networks for biomedical image segmentation. Lect. Notes Comput. Sci. 9351, 234–241.
Salih, N., et al., 2023. Prediction of ROP zones using deep learning algorithms and voting classifier technique. Int. J. Comput. Intell. Syst. 16, 86.
Sankari, V.M.R., Umapathy, U., Alasmari, S., Aslam, S.M., 2023. Automated detection of retinopathy of prematurity using quantum machine learning and deep learning techniques. IEEE Access 11, 94306–94321.
Sen, P., Jain, S., Bhende, P., 2018. Stage 5 Retinopathy of Prematurity: an update. Taiwan Journal of Ophthalmology 8, 205–215.
Sen, P., Wu, W.C., Chandra, P., et al., 2020. Retinopathy of prematurity treatment: asian perspectives. Eye (Lond.). 34, 632–642.
Sharma, S., 2022. Quantum algorithms for simulation of quantum chemistry problems by quantum computers: an appraisal. Found. Chem. 24, 263–276.
Smith, L.E., 2004. Pathogenesis of Retinopathy of Prematurity, vol. 14. Growth Hormone & IGF Research.
Srivastava, S., Dwivedi, A.D., 2021. Review of quantum image processing. ResearchGate preprint. https://doi.org/10.13140/RG.2.2.28458.95683.
Subramaniam, A., et al., 2023. Image harmonization and deep learning automated classification of plus disease in retinopathy of prematurity. J. Med. Imaging 10, 061107.
Syed, S., Islam, M.M., Sadi, T.A., Hasan, H., Mahdy, M.R.C., 2024. Edge detection quantumized: A novel quantum algorithm for image processing. arXiv (Cornell University) preprint arxiv:2404.06889.
The PennyLane Team, 2023. RandomLayers — PennyLane documentation. Available from: https://docs.pennylane.ai/en/stable/code/api/pennylane.templates.layers.RandomLayers.html.
Vallero, M., Dri, E., Giusto, E., Montrucchio, B., Rech, P., 2024. Understanding logical-shift error propagation in quanvolutional neural networks. IEEE Transactions on Quantum Engineering 5, 3100914.
Vaswani, A., et al., 2017. Attention is all you need. Neural Information Processing Systems (NIPS) 9992–10002.
Xiao, X., Lian, S., Luo, Z., Li, S., 2018. Weighted res-unet for high-quality retina vessel segmentation. 9th International Conference on Information Technology in Medicine and Education, pp. 327–331.
Yenice, E.K., Kara, C., Yenice, M., Erdas, C.B., 2023. Retinopathy of prematurity in late preterm twins with a birth weight discordance: can it be predicted by artificial intelligence? Beyoglu Eye Journal 8, 287.
Young, B.K., et al., 2023. Efficacy of smartphone-based telescreening for retinopathy of prematurity with and without artificial intelligence in India. JAMA Ophthalmology 141, 582–588.
Zhang, Y., Lu, K., Gao, Y., Wang, M., 2013. NEQR: a novel enhanced quantum representation of digital images. Quant. Inf. Process. 12 (8), 2833–2860.
Zheng, C., Johnson, T.V., Garg, A., Boland, M.V., 2019. Artificial intelligence in glaucoma. Curr. Opin. Ophthalmol. 30, 97–103.
Zhou, Z., Rahman Siddiquee, M., Tajbakhsh, N., Liang, J., 2018. Unet++: a nested U-Net architecture for medical image segmentation. Lect. Notes Comput. Sci. 3–11.