IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 8 July 2023, accepted 22 August 2023, date of publication 1 September 2023, date of current version 7 September 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3311346

RESEARCH ARTICLE

Automated Detection of Retinopathy of Prematurity Using Quantum Machine Learning and Deep Learning Techniques

V. M. RAJA SANKARI[1], U. SNEKHALATHA[1], (Member, IEEE), SULTAN ALASMARI[2], AND SHABNAM MOHAMED ASLAM[3], (Member, IEEE)

[1]Department of Biomedical Engineering, College of Engineering and Technology, SRM Institute of Science and Technology, Kattankulathur, Tamil Nadu 603 203, India
[2]Department of Information Systems, College of Computer and Information Sciences, Majmaah University, Al Majmaah 11952, Saudi Arabia
[3]Department of Information Technology, College of Computer and Information Sciences, Majmaah University, Al Majmaah 11952, Saudi Arabia

Corresponding author: U. Snekhalatha (snehalau@srmist.edu.in)

This work was supported by the Deanship of Scientific Research at Majmaah University under Project R-2023-570.

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Institutional Ethics Committee, SRM Medical College Hospital, and Research Centre, under Clearance No. 8241/IEC/2022, and performed in line with the Declaration of Helsinki.

ABSTRACT Retinopathy of prematurity (ROP) is a vasoproliferative retinal disease that affects premature infants and causes permanent blindness if left untreated. Automated retinal diagnosis from the Retinal fundus images aid in the early detection of many pathological conditions. The low-level statistical features used in literatures have not provided the complete ROP-specific profile, and hence it has to be replaced by high-level features. The proposed system involves extracting Scale Invariant Feature Transform (SIFT) - Speeded Up Robust Features (SURF) combined high-level features from the SegNet segmented retinal vessels and classified using the Quantum Support Vector Machine (QSVM) classifier. This study aims (i) to segment retinal vessels from the acquired fundus images using SegNet and extract their features using the SURF and SIFT Feature Extraction method, (ii) to classify the Normal and ROP retinal vessels using four classical machine learning classifiers such as Support Vector Machine (SVM), Reduced Error Pruning (REP) tree, K-Star, and LogitBoost and Quantum SVM classifier, (iii) to develop a novel transformer-based Swin-T ROP model to classify ROP from normal Neonatal fundus images, (iv) to compare the performance characteristics of the proposed QSVM model with the Resnet50, DarkNet19, and classical machine learning classifiers. The study is conducted using 200 fundus images, including 100 normal and 100 ROP-positive neonatal retinal images. The machine learning classifiers such as SVM, REP Tree, K-Star, and Logit Boost Classifiers attained accuracy of 86.7%, 75%, 74%, and 76.5%, respectively, in classifying ROP from normal retinal images. The deep learning networks such as ResNet50 and DarkNet19 classified ROP from normal fundus images with an accuracy of 92.87% and 89%, respectively. The Quantum machine learning classifier outperforms the classical machine learning classifiers, Pre-trained Convolutional Neural Networks (CNN) and SwinT-ROP in terms of classification accuracy (95.5%), sensitivity (93%), and specificity (98%). The proposed system accurately diagnoses ROP from the neonatal fundus images and could be used in point-of-care diagnosis to access diagnostic expertise in underserved regions.

INDEX TERMS QSVM, transformer, transfer learning, quantum classifier, retinal image processing.

I. INTRODUCTION

Imaging the microcirculation of the retinal layer allows the diagnosis of ocular diseases and the wellness of the entire

The associate editor coordinating the review of this manuscript and approving it for publication was Zhen Ren.

circulatory system and brain [1]. The retinal fundus images aid in diagnosing many pathological conditions such as Diabetic Retinopathy (DR), Retinopathy of Prematurity (ROP), Glaucoma, Macular Edema, etc., which could cause partial to complete blindness. The prevalence of visual impairment due to ROP is estimated to be 10% worldwide [2]. About 490000 preterm neonates survive in India annually, of which at least 5000 neonates are diagnosed with severe ROP [3].

Jalali et al. [4] emphasize that early screening of ROP is crucial in restoring the visual acuity of neonates with the least complication. Since neonatal retinal images are acquired non-invasively, it has a better risk-benefit ratio in diagnosing ROP. A fundus camera captures the retinal images where the light enters and leaves through the pupil. ROP is a proliferative retinal disorder that affects premature neonates, which, when left untreated, could lead to permanent blindness. Several studies predicted the changes in retinal vessels as markers for detecting ROP. Vaso proliferation, tortuosity, and changes in the diameter of vessels indicate the progression of ROP [5].

Due to high variability and inter-observer inconsistency in the diagnosis [6], [7], there is a significant need to develop an automated system for the prediction of ROP that could be used in the mass screening of infants. Automated diagnosis of the pathologies related to retinal vessels recently involves machine Learning and deep learning techniques. Machine learning classifiers interpret complicated and hard-to-perceive patterns from massive data. It automatically classifies the images into normal and pathological retinal vessels by training the classifier based on the given extracted features. But it requires several segmentation algorithms [8] and feature extraction as additional steps to extract the distinct attributes of the specific classes. Quantification of retinal vessel features are complex in fundus image and subject to errors. Hence the automated analysis of the retinal vessels is crucial in accurate pathological diagnosis.

Quantum computation utilizes the quantum principle of superposition and entanglement to the computational problems that cannot be solved by a classical computer [9]. Quantum algorithms involve qubits, which could have values 0, 1, or both simultaneously instead of bits, as in classical computers. Quantum machine learning (QML) integrates quantum computing approaches in machine learning techniques to speed up the computation with non-linear data [10]. The proposed study focuses on the segmentation of retinal vessels, feature extraction, and prediction of ROP in neonatal fundus images. The novelty of the study is as follows:

- A dedicated retinal vessel segmentation algorithm is developed to process the raw fundus images without the involvement of human annotations. The effect of the SIFT-SURF combined high-level features on ROP prediction is highlighted in the study.
- In addition, a novel transformer-based Swin-T ROP CNN with Relative Positional Encoding is developed to predict the ROP in neonatal fundus images.
- Finally, the quantum classifier (QSVM) is utilized to classify the ROP from the normal images and compare its performance with the classical machine learning classifiers, which has not been carried out in any previous studies.

The significant contributions of the paper are as follows:

- The retinal vessels are segmented using the SegNet CNN. The Speeded Up Robust Features (SURF) and Scale Invariant Feature Transform (SIFT) features are extracted from the segmented images and fused to obtain a 50-length feature vector.
- The ten best features are selected from the fused feature vector using the Univariate feature selection method and classified using the machine learning classifiers.
- Geometrically augmented fundus images are classified into normal and ROP using the ResNet50 and DarkNet19 networks by the process of transfer learning.
- A novel Swin-T ROP CNN with Relative Positional Encoding is designed to classify the ROP from the normal fundus images.
- The selected 10 features are translated into quantum features and classified using a Quantum Support Vector Machine (QSVM) classifier. A comparison of the performance metrics is carried out between the QSVM, ResNet50, DarkNet19 CNN, Swin-T transformer, and the Classical machine learning classifiers.

  This study aims (i) to segment retinal vessels from the acquired fundus images using SegNet and extract their features using the SURF and SIFT Feature Extraction method, (ii) to classify the Normal and ROP retinal vessels using four classical machine learning classifiers such as SVM (Support Vector Machine), REP (Reduced Error Pruning) tree, K-Star, and LogitBoost and Quantum SVM classifier. (iii) to develop a novel transformer-based Swin-T ROP model to classify ROP from normal Neonatal fundus images. (iv) to compare the performance characteristics of the proposed QSVM model with the Resnet50, DarkNet19, and classical machine learning classifiers.

This paper is organized as follows: Section II demonstrates the procedure of data collection, Methods involved in the segmentation of Retinal vessels, and classification using several machine learning classifiers and Deep Learning networks. Section III presents the performance of different classifiers and methods. The implications of the observations are characterized in Section IV. Finally, section V includes the conclusion of the article.

II. LITERATURE REVIEW

Prediction of ROP using Artificial Intelligence has evolved in recent years. The handcrafted features could be used in clinical studies of ROP classification. Bayraktar and Boyraz [11] compared the performance of different feature descriptors such as Features from Accelerated Segment Test (FAST), Binary Robust Independent Elementary Features (BRIEF), Oriented FAST and Rotated BRIEF (ORB), SIFT, and SURF and their 23 different combinations in image localization. Their results emphasize that the SIFT-SURF combination

TABLE 1. Identified Research gap in previous studies involving Prediction of ROP.

Year	Author	Dataset	Dataset Quantity	Methodology	Performance	Research Gap
2021	Chen et al. [23]	Private Dataset	11,849	ResNet152	94% Sensitivity and 0.99 AUC	· Utilised only one complex pre-trained network for classification. · The dataset was pre-processed using CLAHE (Contrast Limited Adaptive Histogram Equalization) and Weiner filter, and hence it does not represent the real-time images.
2020	Yildiz et al. [24]	Private Dataset 1	5512	U-Net segmentation with Optic Disc Center Detection Vessel tracing, and feature extraction. The classifiers used are Logistic regression, SVM, and Fully connected Multilayer Perceptron classifiers	0.94 AUC	· The statistical features extracted do not represent the discriminative details of the segmented and traced vessels. · No state of art CNN networks were utilized in classification.
2016	Worrall et al. [25]	Cannada Dataset	1459	Modified GoogLeNet and Bayesian CNN	91.8% Accuracy and 82.5% Sensitivity	GoogLeNet is not specific to ROP diagnosis and computationally expensive.
2021	Huang et al. [26]	Private Dataset	11372	Five Convolutional layered CNN	92.23% Accuracy and 96.14% Sensitivity	Five convolutional layers with batch normalization is utilized where the need for batch normalization is not justified.
2018	Brown et al. [27]	Private Dataset	5511	UNet and Inception V1	91% Accuracy and 93% Sensitivity	The system is not specific to ROP diagnosis and is computationally expensive.
2020	Ding et al. [15]	Private Dataset	1657	Mask R-CNN with ResNet 101	67% Accuracy	Performance of the pre-trained network is very less for clinical diagnosis.

yields maximum classification accuracy of 98.41% compared to individual and other combinations of features. Bansal et al. [12] studied the performance of three feature descriptors such as, SIFT, SURF, and ORB, and their combinations in object detection using the Caltech-101 public dataset. It is evident from their study that the SIFT-SURF feature combination surpasses the performance of these individual feature descriptors. The extraction of handcrafted features is time-consuming and increases the complexity of the classifier for a vast dataset [13]. But when there is a need to obtain better accuracy of the classification problems with a large dataset, a Convolution Neural Network (CNN) is employed. CNN is a deep-learning network known for self-training the given images and classifying them with better accuracy without requiring manual feature maps [14]. Many pre-trained models such as Alexnet, GoogleNet, ResNet, and SqueezeNet were able to classify the images with pre-existing knowledge.

Ding et al. [15] developed a hybrid system ensembling object detection and a pre-trained classification network to predict ROP. Mask R-CNN is utilized to detect the demarcation lines and ridges, followed by the Resnet-101 for the classification of various stages of ROP. Although the performance of their developed system is not promising, the authors suggested that object detection in classification could improve the stability of ROP prediction. Staal et al. [16] extracted the pixel-based feature vectors related to the retinal vascular ridges using the retinal images obtained from the DRIVE Database. But the images are classified using a supervised k-NN classifier which requires the annotated ground truth images. Soares et al. [17] used a supervised method to segment the retinal vessels from the DRIVE Database and extracted their features using two dimensional (2D) Gabor wavelet transform. The authors then classified them using a Bayesian classifier. This method failed to segment the thin vessels, which are rarely perceived by human annotations. You et al. [18] designed a semi-supervised algorithm ensembling SVM with a self-training classifier. They extracted the retinal vessels in images from DRIVE and STARE databases based on radial projections and extracted the features using steerable complex wavelet transform. But their method fails to detect several vessels in pathological regions and provides misclassified results.

Jelinek et al. [19] developed an automated vessel segmentation technique with a two-dimensional retinal vessel profile tracking model. They used the Naive Bayes classifier to categorize the vessels using the eight features estimated from the mean and standard deviation of the segmented vessels from the Red, Green, Blue, and Hue regions. But the authors suggested that the performance of the classification

could further be improved by selecting more vessel-specific high-level features. Ooi et al. [20] segmented the retinal blood vessels using a Canny edge detector. But the complete system is a Graphical User Interface, where when the cursor is pressed with a single left click and traced over the vessel, edge detection will be carried out to obtain the segmented vessel. This semi-automated system requires an interactive session from the user and still fails to detect the complete edges. This segmentation method is very time-consuming as well as prone to errors. Li et al. [21] combined the UNet with the DenseNet to segment the retinal vessels. But their method requires the additional step using fusion limited contrast histogram equalization. The utilized Dense-UNet has 89 convolutional layers, which increases the complexity of the system.

Amin et al. [22] employed a 4-qubit-quantum circuit to detect breast malignancy using histopathological images. But their method proved to surpass the classical classifiers with low computational complexity in binary classification applications. The majority of previous studies by Chen et al. [23], Worrall et al. [25], Huang et al. [26], Brown et al. [27], and Ding et al. [15] predicted ROP involving digital fundus images and proved to be successful. Chen et al. [23] achieved a sensitivity of 94% and 0.99 AUC in ROP prediction using a ResNet152 pre-trained network. Worrall et al. [25] utilized Modified GoogLeNet and Bayesian CNN to classify ROP in digital fundus images with an accuracy of 91.8% and a sensitivity of 82.5%. Huang et al. [26] designed a five-layered CNN for detecting ROP in digital fundus images and achieved an accuracy of 92.23% and a sensitivity of 82.5%. Brown et al. [27] classified ROP using the Inception V1 network in neonatal retinal images with an accuracy of 91% and a sensitivity of 93%. Similarly, Ding et al. [15] classified ROP using a ResNet 101 network in digital fundus images and achieved an accuracy of 67%. Table 1 provides a comparison of the recent literature on the prediction of ROP. The prediction of ROP using Artificial Intelligence in literature proved to require additional pre-processing and segmentation techniques for CNN and is computationally expensive. The low-level statistical features used in previous studies [17], [19], [23], [24] may not provide the complete ROP-specific profile, and hence it has to be replaced by high-level features. Studies state that using a quantum classifier to replace the classical machine learning classifier could lead to better prediction performance with lower complexity. Hence, the proposed system involves extracting SIFT-SURF combined high-level features from the SegNet segmented retinal vessels and classifying using the QSVM classifier.

III. METHODOLOGY

A. DATA COLLECTION

This study includes 200 neonatal fundus images collected retrospectively from SRM Medical College and Research Centre, Kattankulathur, Tamil Nadu, India, from February 2019 to February 2022. The retinal fundus images were obtained from the neonates after dilating the pupil using mydriatic drugs. The image resolution of 1024×1024 was acquired by trained professionals using 3Nethra Neo (Forus Health, Bangalore) mydriatic digital wide field fundus camera having a 120° Field of View. A total of 200 fundus images, which consist of 100 Normal and 100 ROP, are included in the proposed study as in Table 2. The certified Ophthalmologists identified and labeled the images of normal and diseased cases separately. The images are captured from preterm infants of average gestational age of 30.3 ± 1.2 weeks with an average birth weight of 1.47 ± 0.12 kg. Neonates having congenital abnormalities such as microphthalmia are excluded from the study. Approval from the Institutional Ethics Committee, SRM Medical College Hospital, and Research Centre, was obtained for the current study with the Clearance number: 8241/IEC/2022, and adhered to the tenets of the Declaration of Helsinki. The ethical approval with its forms and sample images are uploaded in the link: https://drive.google.com/drive/folders/1o1T837-S7FUf9-1P4vioWJZhogVcgAfP?usp=drive_link

TABLE 2. Description of the images utilized in the study.

Characteristics	Normal	ROP
Number of Images	100	100
Average Gestational Age (in weeks)	29.8 ± 1.6	31.84± 1.56
Birth weight (kg)	1.7 ± 0.25	1.38 ± 0.31
Number of Male infants	38	43
Number of Female infants	62	57
Number of Stage 1 ROP Neonates	NA	32
Number of Stage 2 ROP Neonates	NA	45
Number of Stage 3 ROP Neonates	NA	23

*NA- Not Applicable

B. PROPOSED WORK FLOW

The block diagram of the proposed study is depicted in Figure 1. The input retinal fundus images are preprocessed to remove the noise and then segmented to extract the vessels using SegNet architecture. The uncorrelated features of segmented vessels are obtained using the SIFT and SURF Feature extraction methods. The extracted 50 features are fused, and the ten best features are selected using the Univariate feature selection method and given to the different classical and quantum machine learning classifiers to classify the normal and ROP. In the case of Deep learning techniques, the raw retinal images are directly given to the pre-trained networks, Swin-T ROP model, and are classified into normal and ROP vessels.

1) RETINAL VESSEL SEGMENTATION

SegNet is a deep CNN to perform pixel-wise semantic segmentation [28]. The SegNet comprises a 16-layer encoder and

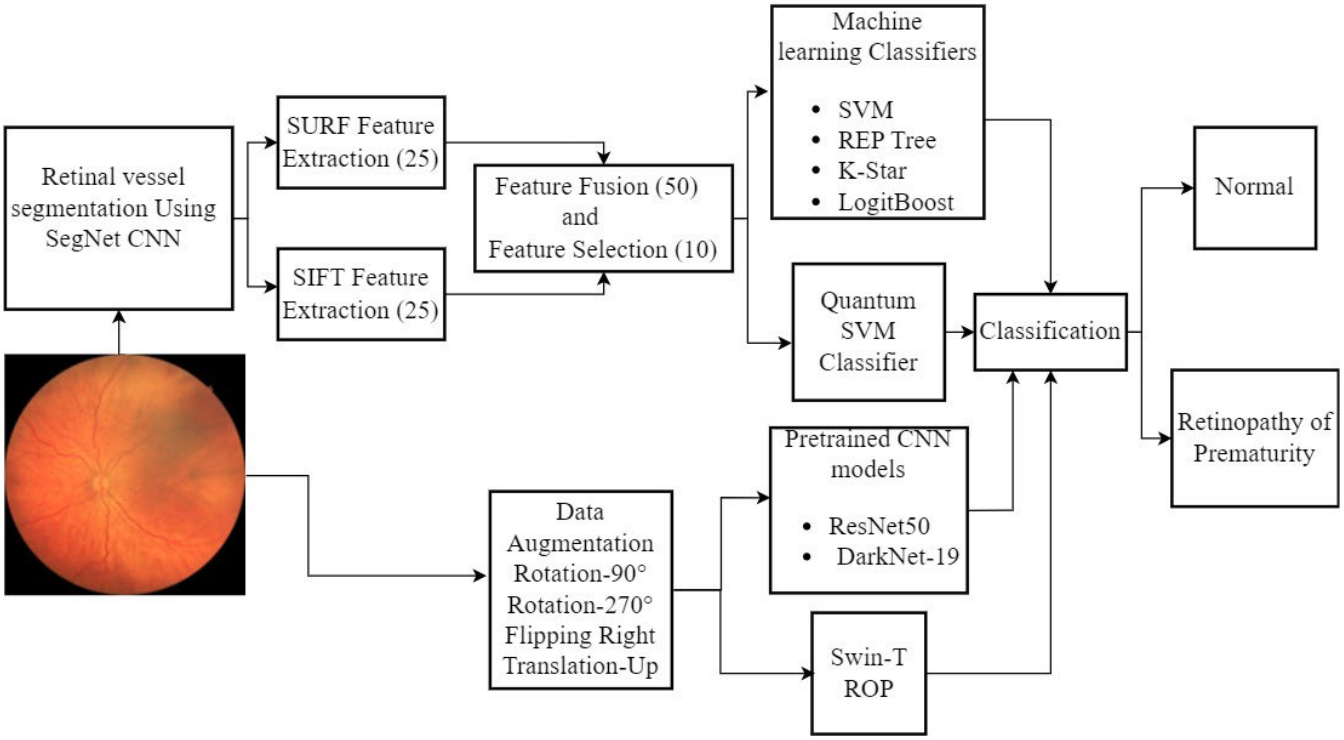

FIGURE 1. Block Diagram of the ROP Prediction Procedure depicting the classification of the input neonatal retinal images into Normal and ROP using classical machine learning classifiers, QSVM, Pre-trained Deep learning models, and Swin-T ROP model.

corresponding decoder block, followed by a pixel-wise classification layer. The encoder block generates the low-resolution feature maps from the input retinal image. The feature maps are then batch normalized, and a Rectified Linear Unit Activation function (ReLU) is applied. Sixteen convolutional layers in the decoder block upsample the feature maps obtained from the encoder block into high-resolution feature maps. The decoder block utilizes pooling indices of the corresponding encoder block to execute non-linear upsampling. The pooling indices correspond to the locations of the retinal vessels stored in the corresponding encoder block. The feature maps from the decoder block are given to the multi-class soft-max classifier, where each neuron corresponds to the class probabilities of every pixel. The illustration of SegNet CNN is given in Figure 2.

The SegNet is trained with the images and their corresponding ground truth annotations from databases such as DRIVE [29], STARE [30], and CHASE [31]. DRIVE, STARE, and CHASE databases contain 20, 20, and 28 images, respectively, along with their corresponding masks. The retinal RGB images are initially resized to dimensions $256\times256\times3$ using the Bicubic Interpolation method. It is used to ensure that all the images utilized have uniform dimensions. Then the resized images are normalized using Min-Max scaling, where all the pixels in the image are scaled to lie between 0 and 1. These 68 images are augmented using geometric transformation algorithms such as rotation at 90 and 270 degrees, flipping right, and translation up to yield 272 images. The SegNet is trained with these

TABLE 3. The Optimal hyperparameters chosen to train the SegNet network.

SI.No	Parameter	Tuned variable
1	Optimizer	ADAM
2	Loss function	Binary Crossentropy
3	Learning rate	$1 \times e^{-1}$
4	Weight Decay	$1 \times e^{-4}$
5	Momentum	0.99
6	batch size	64

augmented images using Adam Optimizer for 200 epochs with a batch size of 64. After successful training, 200 fundus images are provided to the network to extract the segmented retinal vessels. SegNet is trained with the tuned hyperparameters, as mentioned in Table 3. The optimal learning rate, Weight decay, momentum, and batch size were tuned using the Grid search mechanism. The optimizer and loss function are determined by controlling the error by the trial-and-error method. Optimizers such as ADAM, SGD, and RMSprop, and Loss functions such as Hinge, Hinge squared, and Binary crossentropy functions were used in both training and testing processes to choose the optimal parameters.

2) FEATURE EXTRACTION

The SURF and SIFT features were extracted from the segmented retinal vessels' features. The SURF feature descriptors describe an image based on the intensity distribution of the pixels within the neighborhood of the region of interest. Surf detects the point of interest using the determinant of

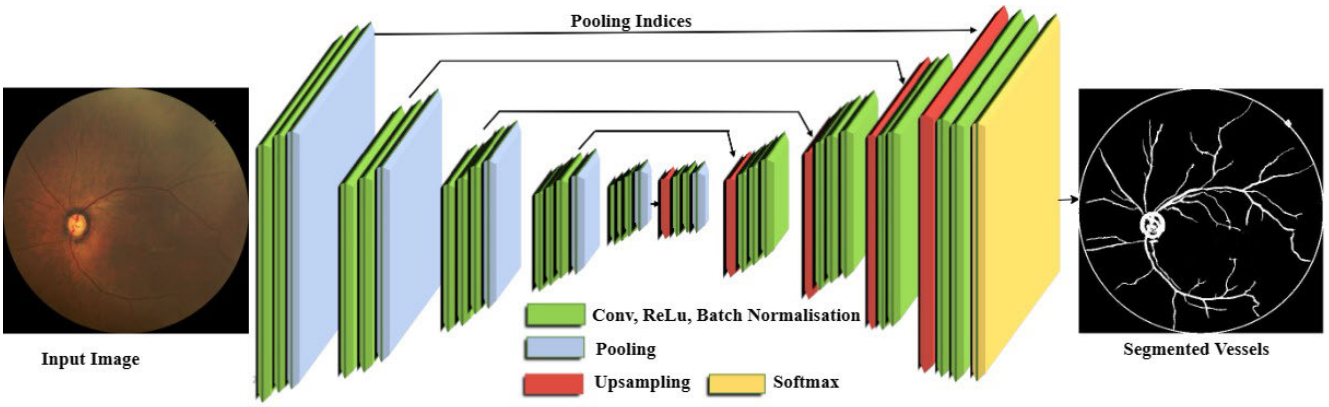

FIGURE 2. The Architecture of SegNet CNN in segmentation of retinal vessels from neonatal images. The encoder blocks (green color) generate the low-resolution features that are converted into sparse feature maps using the decoder blocks (Red color) using the pool indices. The Softmax layer (Yellow) performs pixel-wise classification to provide the vessel segmented output.

the Hessian blob Detector [32]. It calculates the features based on the sum of the Haar Wavelet response. SIFT feature extractor detects the local features from the image based on the gradients of pixels to its surroundings [33]. These features are highly robust and invariant to scale, orientation, and affine distortion [34]. SIFT features are extracted from the high-contrast regions in the image, such as lines and ridges. These are utilized in extracting the features related to the retinal vessels specific to ROP. From the input image, the Haar wavelet responses are obtained using 5×5 equally spaced seed points. Hence, twenty five SURF features and 25 SIFT features are obtained from every image using the SURF and SIFT feature extractors, respectively.

3) FEATURE FUSION AND SELECTION

Twenty-five SURF and Twenty-five SIFT features are fused together to obtain a 50-length feature vector. The uncorrelated optimal features are used to train the classifiers to achieve greater accuracy without overfitting. In order to avoid leakage of data during the training process, feature selection is carried out after a 10-fold cross-validation using machine learning classifiers [35]. The Chi-square statistic-based Univariate feature selection is employed to select 10 best features from the initial feature vector. The most significant ten features are selected by calculating the dependence of the target to the individual non-negative feature at a 95% confidence interval.

C. MACHINE LEARNING CLASSIFIERS

The current study used machine learning classifiers from Scikit-learn Machine learning library version 1.2. Four machine learning classifiers such as SVM, REP Tree, K-star, and LogitBoost, are used to classify the retinal blood vessels.

1) SVM CLASSIFIER

SVM is a supervised machine learning classifier [36] with the maximum margin between the given classes to have minimum classification error [37]. It classifies two classes by identifying the optimal hyperplane that marks the decision boundaries [38]. In the proposed study, the Radial basis function (RBF) is chosen as the kernel to classify the ROP neonates from normal ones. The hyperparameters of the SVM classifier are the cost parameter C, which is chosen to be 1, and a kernel coefficient gamma value is 0.01 to optimize 10-fold validation accuracy. In the proposed work, the SVM classifier is implemented using the sklearn.svm package.

2) REP TREE CLASSIFIER

REP tree is a type of top-down decision tree learner that computes decisions faster. In retinal vessel classification, the REP tree generates binary classes using the discrete feature vectors based on gain and prunes it using reduced error pruning with back fitting. REP Tree selects the most predictive feature and splits the current node of the decision tree based on the selected feature. Thus, each node becomes the indicator of a particular feature [39]. Finally, the nodes that have the maximum gain are estimated to generate the resulting class. The maximum depth of the tree is set as -1 for 100 trees per batch. The minimum proportion of variance is designated as 0.001, with the minimum total weight of instances over a leaf is given as 2. REP Tree classification is executed with the sklearn.tree package using REP as the pruning algorithm.

3) K-STAR CLASSIFIER

K-Star classifier is an instance-based learner that classifies the retinal vessels based on the entropy-distance function. It has the advantage of consistency over other classifiers using real-valued features of retinal blood vessels [40]. Here, the features of the training instances are mapped to the classes initially defined. And the new instance is included in the class, which has the same features. The parameter of global blending B is set as 20 with batch size as 100. K-star is implemented using the sklearn and BaseEstimator packages.

4) LOGITBOOST CLASSIFIER

LogitBoost is a statistical meta-algorithm based boosting classifier with a logistic regression function in a generalized additive booster. It combines several weak base classifiers

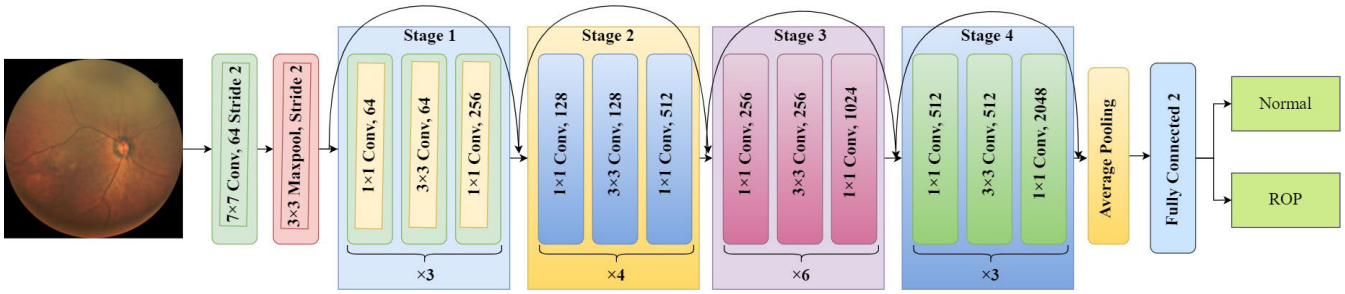

FIGURE 3. Architecture of ResNet50 CNN for the prediction of ROP. The input image is given to stacked blocks of 50 convolutional layers with residual skip connections. The feature map thus produced is fed to Fully connected layer to generate the two-class classification of ROP and normal.

to obtain a robust base classifier with better accuracy and minimal logistic loss. The classification of the retinal vessels is achieved by generating the decision function, which is the linear sum of the decision function of all the weak classifiers [41]. The base classifier used in logit boost is decision Stump with batch size of 100. The total number of iterations to be performed is designated as 10, and the likelihood improvement threshold is -1.79 E 308. LogitBoost classification is carried out using the LogitBoost class cloned on top of Scikit learn package.

D. DEEP LEARNING ALGORITHM

Lecun et al. [42] demonstrated that overfitting of CNN could be avoided by using a vast number of ground truth images. The pre-trained networks utilize previously acquired knowledge of weights by training the images from the ImageNet database. By altering the final classification layer of the model, the pre-trained networks could be utilized to distinguish ROP images from the normal using the transfer learning approach. Data augmentation is employed when the number of available ground truth images is limited to generalize the learning process of the CNN implicitly. The acquired 200 retinal fundus images (100 normal and 100 ROP) are augmented to 800 images using four geometric transformation techniques such as Rotation-90°, Rotation-270°, Flipping-Right, and Translation-Up directions. The acquired 200 images are combined with the augmented 800 images to obtain 1000 fundus images for training and testing the CNN. These 1000 images include 500 normal and 500 ROP images. Geometric transformations remove the image impairments due to positional biases in the retinal vessels.

ResNet50 is a deep learning residual network designed with fifty layers (48 convolutional layers, 1 Max Pooling, and 1 Average pooling layer) and is developed by He et al. [43]. This Residual Network has less time complexity when compared with VGG16 or VGG19 models. Residual networks reduce the complexity of CNN using gated recurrent skip connections. Consider a layer x present between layer n and layer (x+n). If layer x does not contain features that fit the model, it generates a skip connection from layer n to layer (x+n). These skip connections reduce the layers of computation and perform faster than VGG16 and VGG19 models.

The architecture of ResNet50 CNN is shown in Figure 3. The input image is given to a convolution layer which has 64 kernels of size 7×7 with a stride size of 2. A Max Pooling layer follows it with a stride of size 2. Four stages of three-layer convolutions are constructed with kernels 1×1, 3×3, and 1×1. The 1×1 layer reduces and increases (restoring) dimensions which causes the middle 3×3 layer to be left with smaller input and output dimensions [44]. All three convolution layers are repeated thrice in the first stage, four times in stage 2, six times in stage 3, and thrice in stage 4. Finally, downsampling is carried out using a convolutional layer of stride 2, and the features are fed to the global average pooling layer. The retinal images are classified into binary outputs in the fully connected dense layer with a softmax activation function. The model requires 1.4 hours to train, and the average testing time is 0.07 seconds. Darknet-19 is a faster and stronger pre-trained network having 19 convolutional layers and five max pooling layers, developed by Redmon et al. [45]. DarkNet-19 network is built using YOLOv2 as the base model. It predominantly uses filters of size 3 × 3 in the convolutional layers. The number of channels is doubled after each max pooling layer. The architecture of DarkNet-19 CNN is illustrated in Figure 4.

The final convolutional layers result in 7 × 7 features for all the images tested, based on which the classification is made. The Global Average pooling layer is used at the end of the 19th convolutional layers to make the predictions and compress the feature representations. DarkNet-19 employs batch normalization as a crucial step in accelerating convergence and regularizing the model. By replacing the softmax layer with 1000 neurons by a new softmax classification layer with 2 neurons, the DarkNet-19 is employed to classify ROP from normal images. The time taken to train the DarkNet-19 is 1 hour, and the average testing time is 0.09 seconds.

The pre-trained networks are executed using a system with an Intel Core i7-7700 processor (3.6 GHz, 8 MB cache), 16 GB RAM with an NVIDIA TITAN XP GPU(12 GB, 1582 MHz).

1) SWIN-T ROP

Swin-T ROP is a modified hierarchical transformer that utilizes Relative Positional Encoding (RPE) in traditional Vision Transformer (ViT) [46] and Swin Transformer [47]. These Positional encodings in self-attention help the model

FIGURE 4. Architecture of DarkNet-19 Network in ROP prediction. Nineteen convolution layers (green color), along with five Maxpooling layers (blue color) generate high-resolution feature maps, which are fed to the Average pooling layer (yellow) with Softmax Dense classification layer (Red color), where the output classification result as ROP or normal is obtained.

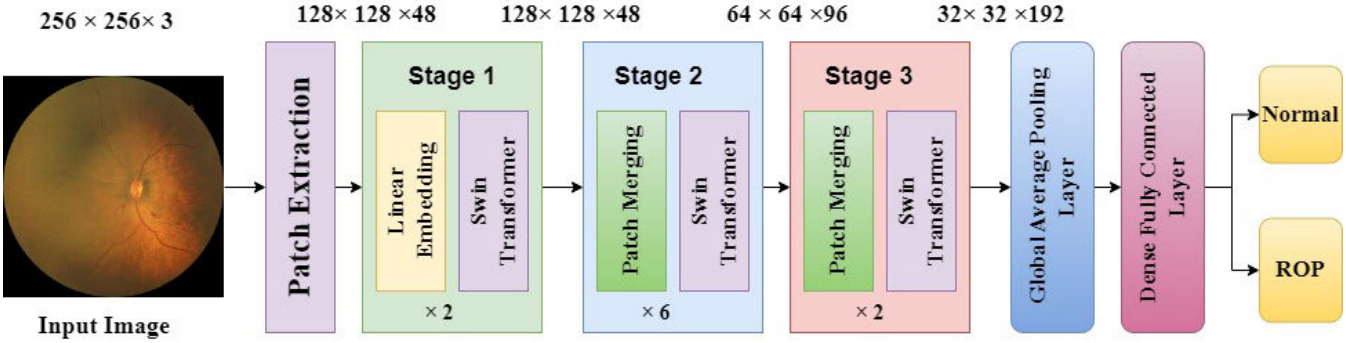

FIGURE 5. Architecture of Swin-T ROP in ROP prediction. Input image of size 256×256×3 is converted into 128×128×48 patches that are given to three stages of the transformer. The local and global dependencies from the three stages are provided to the Average pooling and Dense fully connected layer to give the final classification output as normal or ROP.

to understand the relative positions of the generated tokens. These relative positional encodings provide additional information about the distance or relative positions between tokens. These encodings could enhance the model's ability to capture both global and local dependencies on the retinal vessels. The architecture of the Swin-T ROP model is depicted in Figure 5.

The Swin-T ROP model extracts 128×128 non-overlapping patches, known as tokens, from the input raw image of size 256×256×3. Each token is of size 2×2 in three channels, with the feature dimension of 2×2×3=12. These patches are given to a linear embedding layer and then to a Swin transformer block. The Swin Transformer block has a shifted and a non-shifted window RPE, with a two-layer multi-layer perceptron (MLP) and an Exponential Linear Unit (ELU) Activation function. In order to get more prominent features, the extracted patches are merged by concatenating the features, reducing the number of tokens to 64 × 64 in the second stage and 32 × 32 in the third stage. Finally, the local and global dependencies are obtained from regular and shifted window-based attention blocks from all three stages. These feature maps are fed to the global average pooling layer and a dense classification layer with a sigmoid activation function to classify the images into ROP and normal. The hyperparameters involved were a window size of 7×7 with a shift size of 2. The model is trained with a learning rate of 1×e-3 and a batch size of 64 for 100 epochs.

Pseudocode for Swin-T ROP model designed to predict ROP

Step1: Initialise the input image and resize it to 256×256×3
Step2: Extract 128×128 image patches of size 2×2 from the input image
Step3: Embed the image patches using an embedding layer
Step4: Add relative positional encoding to the patch embeddings
Step5: Initialise the number for stages = 0
Step6: Perform shifted window attention within the stage
Step7: Apply a multi-layer perceptron (MLP) to each patch embedding
Step8: Increment stages by 1.
Step9: Repeat Steps 6 to 8 until stages = 3
Step10: Flatten the embeddings and apply them to the dense layer for classification.

E. QUANTUM MACHINE LEARNING CLASSIFIER

Quantum computing classifies the features in a quantum-enhanced feature space using the kernel matrix constructed by training the data [48], [49]. The quantum classifier interprets the complex features based on the quantum feature maps and kernel functions which is a difficult task for conventional classifiers [50]. SVM classifies the instances based on the

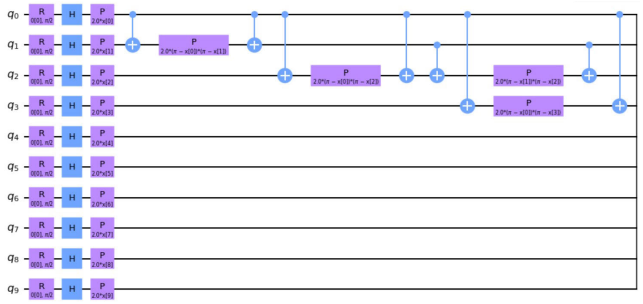

FIGURE 6. Quantum Circuit Implementation with ten Qubits (q0 × q9) of Second-order Pauli-Z features for ten classical features. H represents the Hadamard gate, P represents the Phase shift gate, and R represents the Rotation gate of the qubits .

optimal hyperplane in the kernel feature space [51]. The QSVM classifier is used to classify the normal images and ROP owing to its advantage of parallelism and estimate the maximum kernel vectors to optimize the hyperplane. Let $I \in X$ be the classical data mapped to quantum feature space as in equation 1.

$$\phi(i) = U(i)|0^n\rangle\langle 0^n|U^*(i) \tag{1}$$

where n is the number of qubits used in the unitary circuit U(i). Quantum Kernel estimator (QKE) measures the kernel plane on a quantum computer by evolving the initial state $|0^n\rangle$ with $U^*(i)U(i)$ for the frequency of the entire results 0^n [52] The cost function of QSVM is estimated by maximizing the upper bound to the generalization error greater than $\pm$ and minimizing the upper bound less than ψ as in equation 2.

$$f(\alpha, \psi) = \sum_{x=1}^{h} \alpha_x - \frac{1}{2} \sum_{x,y=1}^{h} (\alpha_x \alpha_y)(\beta_x \beta_y) N \psi(I_x, I_y) \tag{2}$$

where the constraints are $0 \leq \alpha_x \leq C$ and $\sum_x \alpha_x \beta_x = 0$ with C as the regularization parameter in QSVM [53]. QSVM utilizes Grover's search technique to accelerate the process of unstructured search instead of the minimal sequential optimization in a classical SVM classifier [54]. The selected ten features from the 200 images are split to give 80% data for training and 20% for testing. Ten qubits are used (Since the number of features =10) to design the Quantum circuit as in Figure 6.

The ten classical features are converted to a quantum features map using Second-order Pauli-Z evolution. The Quantum Kernel is trained with the feature map using Quantum Kernel trainer with SVC Loss as the loss function. Simultaneous Perturbation Stochastic Approximation (SPSA) optimizer is utilized with 0.1 perturbation, a learning rate of 0.005 for a maximum of 100 iterations. The model is executed using IBM Qiskit with Python version 3.7. The time taken to build the model is 24 minutes.

F. STATISTICAL ANALYSIS

In the proposed work, the t-test primarily assessed the statistical significance between the features extracted from the segmented vessels of ROP and Normal fundus images. Cohen's d effect size associated with the t-test is used to provide information about the magnitude of the difference between the features extracted. Paired comparison is carried out to calculate the significance of obtained results with the networks utilized in the literature such as ResNet (p=0.0047(<0.05), at 95% CI) and SVM classifier (p=0.0092(<0.05), at 95% CI). This is found to be in agreement with the work carried by Ding et al. [15], Chen et al. [23], and Yildiz et al. [24]. Paired comparison is carried out by pairing predicted labels from the proposed method with the predicted labels from the models listed in the literature. The selected ten features are also tested for significance with a 95% confidence interval. All statistical analyses are carried out in IBM SPSS software version 29.

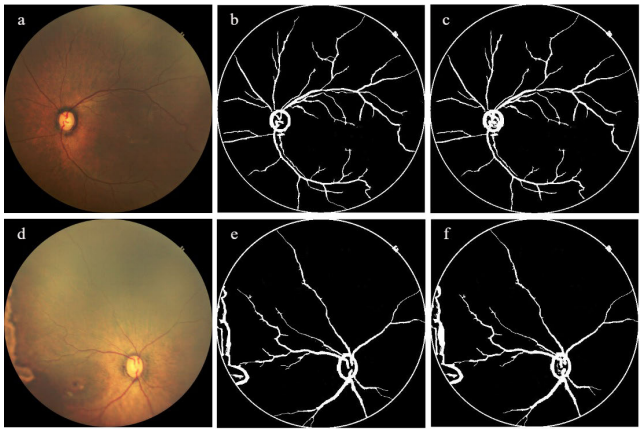

FIGURE 7. a) Normal fundus image of an infant with GA of 34 weeks; b) Ground truth of the normal fundus image; c) Segmented Retinal Vessels using SegNet; d) Fundus image of Stage 3 Zone III ROP of an infant with GA of 32 weeks; e) Ground truth of the ROP image; f) Segmented Retinal Vessels using SegNet .

IV. RESULTS

The retinal vessel segmented output was obtained by SegNet from the neonatal fundus images as shown in figure 7. Figures 7a and 7d are the input retinal images of Normal and ROP neonate respectively. Figures 7b and 7e are the ground truth images of the retinal vessels of the images 7a and 7d. Whereas, Figures 7c and 7f are the segmented retinal vessels obtained from the trained SegNet architecture.

The number of features from the descriptor depends on the image complexity. Since the utilized image is a binary segmented image with retinal vessels, more features add unnecessary complexity by detecting false key points. Figure 8a denotes the SIFT feature extractor using 100 key points. It is seen that many outliers are present in the detected features that cause decreased classification rate in the image. It is essential to eliminate the weak key points that are sensitive to noise or have no edge effect. Figures 8b and 8c demonstrate the 25 features extracted from SIFT and SURF feature extractors, respectively. It is seen that in reduced feature extractions, the outliers are eliminated and are more

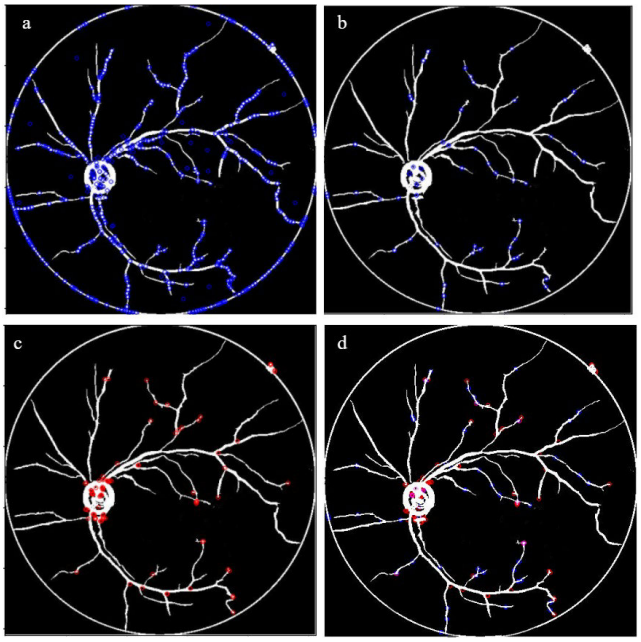

FIGURE 8. a) 100 key point features extracted using the SIFT feature extractor; b) 25 key point features extracted using the SIFT feature extractor; c) 25 key point features extracted using the SURF feature extractor; d) 50 length feature vectors obtained by fusing SIFT and SURF features .

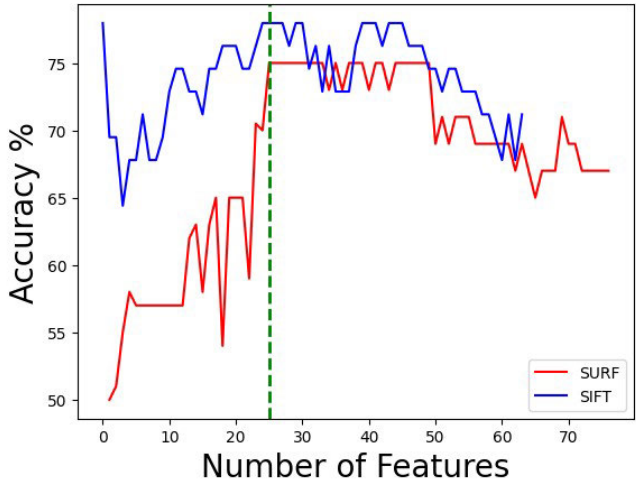

FIGURE 9. The plot describes the change in accuracy with respect to the change in the number of features extracted from the feature descriptors. The red and blue lines indicate the accuracy change from the SURF and SIFT features, respectively. Both feature descriptors yield maximum accuracy initially using 25 features.

specific to retinal vessels. Hence the more prominent 25 features are selected from SIFT and SURF feature descriptors and fused as in Figure 8d.

Twenty five features are selected by the parameter tuning mechanism using the SVM classifier. Figure 9 shows the dependence of the accuracy of the ROP classification on the number of features extracted from SIFT and SURF features. In both SIFT and SURF feature extraction, it is seen that the maximum performance is achieved when the number of features is around 25.

TABLE 4. Effect size obtained from the selected ten features used to train the network.

Feature	Effect Size d
F1	0.863
F2	0.838
F3	0.841
F4	0.644
F5	0.721
F6	0.72
F7	0.785
F8	0.642
F9	0.622
F10	0.61

Table 4 depicts the effect size of each selected feature in the prediction of ROP with a CI of 95%. Since the features extracted are high-level key point descriptors, they couldn't be named and described in the table. It is seen that the effect sizes of the features F1-F3 (d>0.8) are high, and F4-F10 are medium to high, indicating stronger to moderate relationships with the targets and statistically significant.

The performance of the classical machine learning classifiers, CNN models and QSVM is evaluated by constructing the Confusion matrix. The primary parameters present in the confusion matrix are True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). TP and TN denote whether the normal retinal vessels are classified correctly or not. Similarly, FP and FN denote whether the retinal anomaly vessels are classified correctly or not. Based on the parameters mentioned above, specificity, sensitivity, accuracy, Positive Predictive Value (PPV), and Negative Predictive Value (NPV) are derived. Cross-validation is carried out by ensuring that 1/10th part of the entire data set is kept hidden from the network while training is performed. The network tests the hidden data, which does not intersect with the training data, to evaluate its performance. Similarly, in 75-25% hold-out validation, 25% data, which is kept idle from the training data, is used to test the network to prevent data leakage. Table 5 describes the dataset split for 75-25% hold-out validation. In Machine learning classifiers, 25 images from each group are taken to construct a 50-length test set. 150 images are split in a 75-25% ratio to obtain the training and validation set. Hence the classifiers are trained and validated using 112 and 38 images, respectively, and tested using 50 images. In CNN networks, 1000 images are split into 552, 188, and 250 images to obtain training, validation, and testing data sets.

The performance metrics evaluated from the machine learning classifiers such as SVM, REP Tree, K-Star, LogitBoost, Deep learning CNN such as ResNet50, DarkNet19 Networks with and without augmentation, SwinT-ROP and QSVM using 10-fold and hold-out validation are shown in Tables 6 and 7 respectively.

TABLE 5. Description of the Training, testing and validation datasets.

Classifier	Methodology	Training	Testing	Validation
Machine learning classifiers	75-25% Hold Out Validation	112	50	38
CNN	75-25% Hold Out Validation	552	250	188

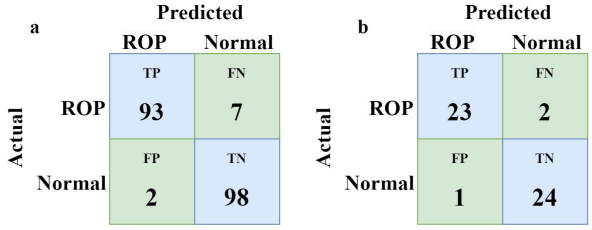

FIGURE 10. a) Confusion matrix of the QSVM classifier using 10-fold cross validation including 200 subjects; b) Confusion matrix of the QSVM classifier using 25% hold out validation including 50 subjects.

The SwinT-ROP model is built using four optimizers such as Yogi, ADAMW (Adam with momentum using decoupled weight decay), SGDW (Stochastic Gradient Descent with momentum using decoupled weight decay), and RMSprop at different learning rates as in table 8. The model achieves the best accuracy using Yogi optimizer at the learning rate of $1 \times e^{-3}$.

The results of the ablation studies are given in Table 9. Swin-T ROP with shifted window partitioning has an 8.4% increase in accuracy compared with the regular single windowing technique. Second, the SwinT-ROP network is investigated with absolute instead of relative positional encoding. The absolute positional encoding reduces the accuracy of the model by 20.5%. Ablations of the proposed model with different windowing sizes and number of stages were performed. The results depict that the Swin-T built on shifted windows (window size=7 $\times$ 7) using relative positional encoding with three stages of Swin-Transformers performs better than the other network variants.

The confusion matrix of the Swin-T ROP and QSVM classifier in Figures 10 and 11 depict that the false negatives (Actual ROP predicted as Normal) in ROP prediction are high. In the case of machine learning classifiers, all 200 subjects were included in the confusion matrix obtained using 10-fold validation, and 50 subjects (25%) were included in the hold-out validation. In case of Swin-T ROP, 1000 subjects were included in the confusion matrix obtained using 10-fold validation, and 250 subjects (25%) were included in the hold-out validation.

The Receiver operating characteristic (ROC) is a graph plotted between the true positive or detection rate and False positive rate. The Area under ROC (AUC) measures the total area in the ROC plot. It quantitatively gives the performance measure of classifiers at all classifying thresholds. Figure 12 shows the ROC curve obtained from the utilized classifiers

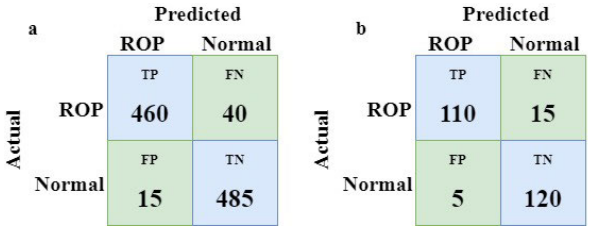

FIGURE 11. a) Confusion matrix of the Swin-T ROP using 10-fold cross-validation including 1000 subjects; b) Confusion matrix of the Swin-T ROP CNN using 25% hold out validation including 250 subjects.

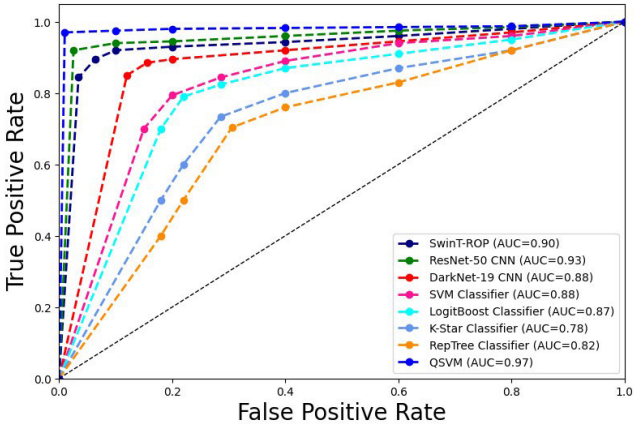

FIGURE 12. ROC curve obtained from utilized classifiers and networks. QSVM classifier covers the maximum area in the plot with AUC=0.97.

and networks. When there are equal instances in groups, the shape of the ROC curve decides the best model for making classifications. The more the ROC curve covers the upper left corner of the plot, the better the model does at classifying the data. Based on the curve provided, it appears that the QSVM classifier is more effective at distinguishing ROP images from normal images. This suggests that the QSVM algorithm could be used as a valuable tool for medical professionals in identifying cases of ROP.

The Machine learning classifiers such as SVM, REP Tree, K-Star, and Logit Boost Classifiers attained accuracy of about 86.7%, 75%, 74%, and 76.5%, respectively. The sensitivity achieved by SVM, REP Tree, K-Star, and Logit Boost Classifiers is 87.88%, 73.15%, 75%, and 75.73%, respectively. The specificity of SVM, REP Tree, K-Star, and Logit Boost Classifiers are 70%, 77.17%, 73.08%, and 77.3%, respectively. But the ResNet50 attained a classification accuracy of 91.5%, with a sensitivity of 90% and specificity of 93%. ResNet50 after data Augmentation improves the accuracy to 92.87%, with a sensitivity of 88.88% and specificity of 79.59%. SVM, REP Tree, K-Star, and Logit Boost Classifiers achieved AUC values of 0.876, 0.82, 0.77, and 0.86, respectively. But, the QSVM classifier outperforms all these classifiers with an accuracy of 95.5%, sensitivity of 93%, and specificity of 98%.

V. DISCUSSION

The automated detection of ROP could be able to provide the appropriate treatment to neonates in rural areas. The

TABLE 6. Comparison of Performance of REP tree, K-Star, LogitBoost Classifiers, and ResNet50 CNN in ROP detection of Retinal image using 10-fold cross validation.

Performance Metrics		Accuracy (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	AUC
REP Tree Classifier		75	73.15	77.17	79	71	0.823
K-Star Classifier		74	75	73.08	72	76	0.775
LogitBoost Classifier		76.5	75.73	77.32	78	75	0.865
SVM Classifier		86.7	87.88	70	86.67	85.18	0.876
ResNet50 CNN	Without Data Augmentation	91.5	90	93	92.8	90.3	0.92
	With Data Augmentation	92.87	88.88	79.59	98	87.75	0.93
DarkNet19 CNN	Without Data Augmentation	84.5	84.69	87.76	83.84	85.15	0.85
	With Data Augmentation	89	89.79	91.84	88	90	0.88
SwinT-ROP		94.5	92	97	96.84	92.38	0.90
QSVM Classifier		95.5	93	98	97.89	93.33	0.97

TABLE 7. Comparison of Performance of REP tree, K-Star, LogitBoost Classifiers and ResNet50 CNN in ROP detection of Retinal image using Hold Out validation.

Performance Metrics		Accuracy (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	AUC
REP Tree classifier		76	81.82	90.91	69.23	83.33	0.81
K-Star Classifier		72	73.91	82.61	68	76	0.72
LogitBoost Classifier		74	72	76	75	73.07	0.78
SVM Classifier		86	87.5	91.67	84	88	0.89
ResNet50 CNN	Without Data Augmentation	90	88	92	91.67	88.46	0.95
	With Data Augmentation	92	92.31	84.62	92.31	91.67	0.92
DarkNet19 CNN	Without Data Augmentation	82	76.92	80.77	86.96	77.78	0.85
	With Data Augmentation	88	91.3	100	84	92	0.87
SwinT-ROP		92	88	96	95.65	88.88	0.89
QSVM Classifier		94	92	96	95.83	92.31	0.92

TABLE 8. Performance of the SwinT-ROP using different optimizers at several learning rates in ROP detection.

Optimizer	Learning rate	Accuracy (%)	Sensitivity %)	Specificity %)	PPV (%)	NPV (%)
YOGI	0.1	89.1	87	91.2	90.8	87.5
	1×e-2	93.3	93.8	92.8	92.8	93.73
	1×e-3	94.5	92	97	96.84	92.38
	1×e-4	93.6	94.4	92.9	94.3	93.6
ADAMW	0.1	90.8	90	91.6	91.46	90.15
	1×e-2	92.6	90	95.2	94.9	90.49
	1×e-3	93.4	91.6	95.2	95	91.89
	1×e-4	93.4	92	94.8	94.65	92.21
SGDW	0.1	74.1	76.2	72	73.12	75.15
	1×e-2	85.8	86.4	85.2	85.3	86.23
	1×e-3	90.29	90	92.4	90.36	90.2
	1×e-4	89	86.4	91.6	91.14	87
RMSprop	0.1	70.2	64.8	75.6	72.6	68.2
	1×e-2	75.9	77	74.8	75.34	76.48
	1×e-3	84.3	83	85.2	84.9	83.69
	1×e-4	81.1	77	85.2	83.87	78.74

earlier detection could slow down the progression of ROP and avoid permanent blindness in the affected neonates. The proposed work emphasizes the clinical implications of ROP in the mass screening programs of preterm infants with the least burden to health care professionals. In the proposed study, ROP is predicted from neonatal fundus images using a QSVM classifier, classical machine learning classifiers, and pre-trained CNN, and their performances were evaluated. Tong et al. [55] classified ROP using ResNet 101 and Faster-RCNN networks and achieved an accuracy of 90.3%. A Morlet wavelet transform-based retinal vascular segmentation method created by Soares et al. [17] exhibited a mean accuracy of 70%. The feature vector is built using the intensity of pixels and Morlet wavelet responses at various scales. The morlet wavelet features extracted are then provided as an input to the Naive Bayes Classifier. But the Morlet

TABLE 9. Results of Ablations on Swin-T ROP in Prediction of ROP.

SI.No	Techniques	Accuracy (%)
1	No ablation (Baseline)	94.5
2	Ablation on shifted Windows (regular windowing)	87.17
3	Ablation on Relative Positional Encoding (Absolute Positional Encoding)	78.58
4	Ablation on Window size (with window size =5)	77.45
5	Ablation on Window size (with window size =3)	74.96
6	Ablation on Stages (Removal of 3rd stage Swin-Transformer)	78.63
7	Ablation on Stages (Removal of 2nd stage Swin-Transformer)	61.25

feature extraction requires optical disc boundary detection to track the retinal vessels.

Fraz et al. [56] used bootstrapped ensemble classifier to classify the retinal vessels and obtained a classification rate of 83%. Han et al. [57] developed an unsupervised automated detection model using a generative adversarial network (Skip-GANomaly) for identifying six different ophthalmic diseases from four real-time datasets. In the classification of anomaly fundus images, they achieved an accuracy of 80.72%, AUC of 0.896 with a sensitivity of 82.69% and a specificity of 82.63%. The current study improved the classification accuracy with 86.7% in Classical SVM and 95.5% in QSVM classifier. Niemeijer et al. [58] developed a supervised approach for classifying the retinal vessels based on the morphological features obtained from the manually segmented vessels. They used SVM and KNN classifiers for the classification of retinal vessels. The authors attained a maximum performance for k-NN classifier with an AUC value of 0.88. The current study of retinal vessel segmentation produced a better AUC value of 0.93 using the ResNet50 network. Desiani et al. [59] created a retinal vessel segmentation model using CLAHE with U-Net and LadderNet. The accuracy obtained by their model is 95.47% which is significant with that of the accuracy obtained from our proposed model. But the need for complex U-Net and LadderNet does not impact the accuracy. Also, it is evident that the SegNet yields good segmentation results on Retinal vessels.

Ahalya et al. [50] predicted rheumatoid arthritis using a QSVM classifier. The authors achieved an improved accuracy of 92.7% compared with classical machine learning classifiers. Akpinar et al. [60] compared the classical SVM with QSVM classifier using multiparametric MRI tissue characteristics of whole lesion in differentiating Medulloblastoma from Ependymoma. The authors concluded that the QSVM classifier achieved the superior performance with an accuracy of 90%. In the proposed study, Quantum machine learning classifiers attained 10.15% increase in the accuracy compared to the classical machine learning classifiers. The deep learning network like Swin T ROP has an 8.99% increase in accuracy compared to the classical machine learning classifier like SVM classifier. Hence, QSVM has better ROP detection than the machine learning classifiers and CNNs. Yet, the data augmentation in CNN improved the accuracy by 1.5%.

An AUC value of 0.5 denotes that the classification accuracy of the classifier is more random, whereas an AUC value of 1 denotes that the classifier recognition is accurate. Higher the AUC, the better is the classification performance. The AUC value of 0.97 indicates that the QSVM classifier has yielded a better classification of ROP.

The majority of the works in retinal vessel segmentation [16], [17], [59] have been carried out using the image in the databases such as DRIVE and STARE. Also, the number of images used from the DRIVE and STARE databases are limited to 40. At the same time, the current study overcomes this limitation by constructing and validating the networks using 200 real-time retinal fundus images. Wang et al. [61] suggested that cardiovascular diseases could be identified several years before their onset by identifying the narrowing of retinal vessels [61]. Thus, the proposed system of classifying retinal images based on the vascular structure finds its broad scope of application as a risk factor in many fatal diseases.

A study by Quinn and Vinekar [62] demonstrated a sensitivity range for diagnosing ROP of 57% to 100% (95% CI) using fundus images. However, the sensitivity revealed by these authors is based on how well the retinal imaging performs in detecting various stages of ROP. The scope of the proposed work is limited to predicting the existence of ROP in retinal images rather than classifying its stages. Hence the accuracy in predicting ROP using the proposed network is high, with a sensitivity of 93% (95% CI 89.3 - 96.7), contradicting the study conducted by Quinn et al.

The limitation of the proposed work is that it could just predict the presence of ROP in the fundus image, rather than classifying it into different stages or identifying the zones. The images utilized were obtained from the same device in a single hospital with similar population characteristics. This might reduce the data diversity that influences the generalizations of the proposed system in ROP diagnosis.

The study conducted by Biten et al. [63] demonstrates the sensitivity of the original diagnosis of the ROP category as 84% [95% CI]. The confusion matrix shows that the false negatives (Actual ROP predicted as Normal) in ROP prediction are high. The network is trained and tested using the neonatal retinal images of the infants with the annotations resulting from the manual ophthalmoscopic examinations conducted by pediatric ophthalmologists. The proposed network is trained based on the annotations from the original diagnosis,

which may have inherent errors. Hence, false negatives from the proposed work may account due to the errors and glare limit in the retinal image captured. A follow-up study of the neonates may ensure whether the false positives (Normal cases predicted as ROP) are low-grade actual ROP cases or may develop ROP in a later stage, which is beyond the scope of the study. This may be carried out in the future to eliminate false positives and enhance the telemedicine approach in ROP diagnosis. CNN-based annotation tools could be used to provide automated annotations [64], and their influence could be studied. Mobile Edge Computing [65] could be carried out in future work to enable real-time processing of retinal images by the end user.

VI. CONCLUSION

The proposed model is aimed at predicting ROP in retinal fundus images having abnormalities in retinal vessel morphology. The study achieved state-of-the-art performance on the stated task with improved performance using real-time retinal fundus images. The classical machine learning classifiers such as SVM, REP Tree, K-Star, and Logit Boost Classifiers attained an accuracy of 86.7%, 75%, 74%, and 76.5%, respectively. ResNet50 and DarkNet19 CNN classified ROP from normal fundus images with an accuracy of 92.87% and 89%, respectively. Among these machine learning classifiers, the SVM classifier achieves the maximum sensitivity of 87.88%, specificity of 70%, and AUC value of 0.876. ResNet50 Network obtained the highest classification performance than DarkNet19 CNN with an accuracy of 92.87%, a sensitivity of 88.88%, and a specificity of 79.59%. The proposed SwinT-ROP model achieved an accuracy of 94.5%, sensitivity of 92%, and specificity of 97%. It is observed that the QSVM classifier surpasses the classical machine learning classifiers and CNN networks with a maximum classification accuracy of 95.5%, a sensitivity of 93%, and a specificity of 98% with an AUC value of 0.97. The utilization of Artificial Intelligence in the classification of different stages of ROP could be carried out in the future. Future research could focus on training the network using images obtained from various localities to enhance the versatility of the network. The proposed approach finds application in point-of-care diagnosis of ROP to gain access to diagnostic expertise in underserved regions.

REFERENCES

[1] J. Flamme, K. Konieczka, R. M. Bruno, A. Virdis, A. J. Flammer, and S. Taddei, "The eye and the heart," *Eur. Heart J.*, vol. 34, no. 17, pp. 1270–1278, May 2013.

[2] H. Blencowe, S. Moxon, and C. Gilbert, "Update on blindness due to retinopathy of prematurity globally and in India," *Indian Pediatrics*, vol. 53, no. 2, pp. S89–S92, Nov. 2016.

[3] S. Chawanpaiboon, J. P. Vogel, A.-B. Moller, P. Lumbiganon, M. Petzold, D. Hogan, S. Landoulsi, N. Jampathong, K. Kongwattanakul, M. Laopaiboon, C. Lewis, S. Rattanakanokchai, D. N. Teng, J. Thinkhamrop, K. Watananirun, J. Zhang, W. Zhou, and A. M. Gülmezoglu, "Global, regional, and national estimates of levels of preterm birth in 2014: A systematic review and modelling analysis," *Lancet Global Health*, vol. 7, no. 1, pp. e37–e46, Jan. 2019.

[4] S. Jalali, R. Azad, H. Trehan, M. Dogra, L. Gopal, and V. Narendran, "Technical aspects of laser treatment for acute retinopathy of prematurity under topical anesthesia," *Indian J. Ophthalmol.*, vol. 58, no. 6, p. 509, 2010.

[5] C. Solarte, A. Awad, C. Wilson, and A. Ells, "Plus disease: Why is it important in retinopathy of prematurity?" *Middle East Afr. J. Ophthalmol.*, vol. 17, no. 2, p. 148, 2010.

[6] M. F. Chiang, "Interexpert agreement of plus disease diagnosis in retinopathy of prematurity," *Arch. Ophthalmol.*, vol. 125, no. 7, p. 875, Jul. 2007.

[7] J. Kalpathy-Cramer, "Plus disease in retinopathy of prematurity: Improving diagnosis by ranking disease severity and using quantitative image analysis," *Ophthalmology*, vol. 123, no. 11, pp. 2345–2351, 2016.

[8] A. Norouzi, M. S. M. Rahim, A. Altameem, T. Saba, A. E. Rad, A. Rehman, and M. Uddin, "Medical image segmentation methods, algorithms, and applications," *IETE Tech. Rev.*, vol. 31, no. 3, pp. 199–213, May 2014, doi: 10.1080/02564602.2014.906861.

[9] S. Sharma, "Quantum algorithms for simulation of quantum chemistry problems by quantum computers: An appraisal," *Found. Chem.*, vol. 24, no. 2, pp. 263–276, May 2022.

[10] G. Sergioli, "Quantum and quantum-like machine learning: A note on differences and similarities," *Soft Comput.*, vol. 24, no. 14, pp. 10247–10255, Oct. 2019.

[11] E. Bayraktar and P. Boyraz, "Analysis of feature detector and descriptor combinations with a localization experiment for various performance metrics," *TURKISH J. Electr. Eng. Comput. Sci.*, vol. 25, no. 3, pp. 2444–2454, 2017, doi: 10.3906/elk-1602-225.

[12] M. Bansal, M. Kumar, and M. Kumar, "2D object recognition: A comparative analysis of SIFT, SURF and ORB feature descriptors," *Multimedia Tools Appl.*, vol. 80, no. 12, pp. 18839–18857, Feb. 2021, doi: 10.1007/s11042-021-10646-0.

[13] L. Alzubaidi, J. Zhang, A. J. Humaidi, A. Al-Dujaili, Y. Duan, O. Al-Shamma, J. Santamaría, M. A. Fadhel, M. Al-Amidie, and L. Farhan, "Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions," *J. Big Data*, vol. 8, no. 1, pp. 1–74, Mar. 2021.

[14] R. Yamashita, M. Nishio, R. K. G. Do, and K. Togashi, "Convolutional neural networks: An overview and application in radiology," *Insights Into Imag.*, vol. 9, no. 4, pp. 611–629, Aug. 2018.

[15] A. Ding, Q. Chen, Y. Cao, and B. Liu, "Retinopathy of prematurity stage diagnosis using object segmentation and convolutional neural networks," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2020, pp. 1–6.

[16] J. Staal, M. Abramoff, M. Niemeijer, M. Viergever, and B. Ginneken, "Ridge-based vessel segmentation in color images of the retina," *IEEE Trans. Med. Imag.*, vol. 23, no. 4, pp. 501–509, Apr. 2004.

[17] J. V. B. Soares, J. J. G. Leandro, R. M. Cesar, H. F. Jelinek, and M. J. Cree, "Retinal vessel segmentation using the 2-D Gabor wavelet and supervised classification," *IEEE Trans. Med. Imag.*, vol. 25, no. 9, pp. 1214–1222, Sep. 2006.

[18] X. You, Q. Peng, Y. Yuan, Y.-M. Cheung, and J. Lei, "Segmentation of retinal blood vessels using the radial projection and semi-supervised approach," *Pattern Recognit.*, vol. 44, nos. 10–11, pp. 2314–2324, Oct. 2011.

[19] H. Jelinek, C. Deaprdieu, C. Lucas, D. J. Cornforth, W. Huang, and M. J. Cree, "Towards vessel characterisation in the vicinity of the optic disc in digital retinal images," in *Proc. Image Vis. Comput. New Zealand (IVCNZ) Int. Conf. (IVCNZ)*, Mar. 2005, p. 6.

[20] A. Z. H. Ooi, Z. Embong, A. I. A. Hamid, R. Zainon, S. L. Wang, T. F. Ng, R. A. Hamzah, S. S. Teoh, and H. Ibrahim, "Interactive blood vessel segmentation from retinal fundus image based on Canny edge detector," *Sensors*, vol. 21, no. 19, p. 6380, Sep. 2021.

[21] Z. Li, M. Jia, X. Yang, and M. Xu, "Blood vessel segmentation of retinal image based on dense-U-Net network," *Micromachines*, vol. 12, no. 12, p. 1478, Nov. 2021.

[22] J. Amin, M. Sharif, S. L. Fernandes, S. Wang, T. Saba, and A. R. Khan, "Breast microscopic cancer segmentation and classification using unique 4-qubit-quantum model," *Microsc. Res. Technique*, vol. 85, no. 5, pp. 1926–1936, Jan. 2022.

[23] J. S. Chen, A. S. Coyner, S. Ostmo, K. Sonmez, S. Bajimaya, E. Pradhan, N. Valikodath, E. D. Cole, T. Al-Khaled, R. V. P. Chan, P. Singh, J. Kalpathy-Cramer, M. F. Chiang, and J. P. Campbell, "Deep learning for the diagnosis of stage in retinopathy of prematurity," *Ophthalmol. Retina*, vol. 5, no. 10, pp. 1027–1035, Oct. 2021, doi: 10.1016/j.oret.2020.12.013.

[24] V. M. Yildiz, P. Tian, I. Yildiz, J. M. Brown, J. Kalpathy-Cramer, J. Dy, S. Ioannidis, D. Erdogmus, S. Ostmo, S. J. Kim, R. V. P. Chan, J. P. Campbell, and M. F. Chiang, "Plus disease in retinopathy of prematurity: Convolutional neural network performance using a combined neural network and feature extraction approach," *Translational Vis. Sci. Technol.*, vol. 9, no. 2, p. 10, Feb. 2020, doi: 10.1167/tvst.9.2.10.

[25] D. E. Worrall, C. M. Wilson, and G. Brostow. (2016). *Automated Retinopathy of Prematurity Case Detection with Convolutional Neural Networks*. Accessed: Jul. 7, 2023. [Online]. Available: https://www.semanticscholar.org/paper/Automated-Retinopathy-of-Prematurity-Case-Detection-Worrall-Wilson/005dad88402e020e503f192435d1ebd9f70039cf

[26] Y.-P. Huang, H. Basanta, E. Y.-C. Kang, K.-J. Chen, Y.-S. Hwang, C.-C. Lai, J. P. Campbell, M. F. Chiang, R. V. P. Chan, S. Kusaka, Y. Fukushima, and W.-C. Wu, "Automated detection of early-stage ROP using a deep convolutional neural network," *Brit. J. Ophthalmol.*, vol. 105, no. 8, pp. 1099–1103, Aug. 2021, doi: 10.1136/bjophthalmol-2020-316526.

[27] J. M. Brown, J. P. Campbell, A. Beers, K. Chang, S. Ostmo, R. V. P. Chan, J. Dy, D. Erdogmus, S. Ioannidis, J. Kalpathy-Cramer, and M. F. Chiang, "Automated diagnosis of plus disease in retinopathy of prematurity using deep convolutional neural networks," *JAMA Ophthalmol.*, vol. 136, no. 7, p. 803, Jul. 2018, doi: 10.1001/jamaophthalmol.2018.1934.

[28] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder–decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.

[29] 21. M. Niemeijer and B. Ginneken. (2002). *Drive Database*. [Online]. Available: http://www.isi.uu.nl/Research/%20Databases/DRIVE/

[30] A. D. Hoover, V. Kouznetsova, and M. Goldbaum, "Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response," *IEEE Trans. Med. Imag.*, vol. 19, no. 3, pp. 203–210, Mar. 2000.

[31] C. G. Owen, "Measuring retinal vessel tortuosity in 10-year-old children: Validation of the computer-assisted image analysis of the retina (CAIAR) program," *Investigative Ophthalmol. Vis. Sci.*, vol. 50, no. 5, pp. 2004–2010, May 2009.

[32] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (SURF)," *Comput. Vis. Image Understand.*, vol. 110, no. 3, pp. 346–359, Jun. 2008.

[33] Y. Liu, J. Tian, R. Hu, B. Yang, S. Liu, L. Yin, and W. Zheng, "Improved feature point pair purification algorithm based on SIFT during endoscope image stitching," *Frontiers Neurorobotics*, vol. 16, Feb. 2022, Art. no. 840594.

[34] J. Philbin, O. Chum, M. Isard, J. Sivic, and A. Zisserman, "Object retrieval with large vocabularies and fast spatial matching," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2007, pp. 1–8.

[35] A. Demircioğlu, "Measuring the bias of incorrect application of feature selection when using cross-validation in radiomics," *Insights Into Imag.*, vol. 12, no. 1, pp. 1–10, Nov. 2021, doi: 10.1186/s13244-021-01115-1.

[36] F. Charih, "Machine learning in audiology: Applications and implications," M.S. thesis, ECE, Carleton Univ., Ottawa, ON, Canada, 2019.

[37] K. V. Park, K. H. Oh, Y. J. Jeong, J. Rhee, M. S. Han, S. W. Han, and J. Choi, "Machine learning models for predicting hearing prognosis in unilateral idiopathic sudden sensorineural hearing loss," *Clin. Experim. Otorhinolaryngol.*, vol. 13, no. 2, pp. 148–156, May 2020.

[38] G. Wu, R. Zheng, Y. Tian, and D. Liu, "Joint ranking SVM and binary relevance with robust low-rank learning for multi-label classification," *Neural Netw.*, vol. 122, pp. 24–39, Feb. 2020.

[39] W. N. H. W. Mohamed, M. N. M. Salleh, and A. H. Omar, "A comparative study of reduced error pruning method in decision tree algorithms," in *Proc. IEEE Int. Conf. Control Syst., Comput. Eng.*, Nov. 2012, pp. 392–397.

[40] J. G. Cleary and L. E. Trigg, "K*: An instance-based learner using an entropic distance measure," in *Machine Learning Proceedings*. Tahoe City, CA, USA: Morgan Kaufmann 1995, pp. 108–114.

[41] B. T. Pham and I. Prakash, "Evaluation and comparison of LogitBoost ensemble, Fisher's linear discriminant analysis, logistic regression and support vector machines methods for landslide susceptibility mapping," *Geocarto Int.*, vol. 34, no. 3, pp. 316–333, Feb. 2019.

[42] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.

[43] K. He, X. Zhang, S. Ren, and J. Sun. (2016). *Deep Residual Learning for Image Recognition*. [Online]. Available: www.cv-foundation.org

[44] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017.

[45] J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 6517–6525, doi: 10.1109/cvpr.2017.690.

[46] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16×16 words: Transformers for image recognition at scale," 2020, *arXiv:2010.11929*.

[47] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin transformer: Hierarchical vision transformer using shifted windows," 2021, *arXiv:2103.14030*.

[48] V. Havlíček, A. D. Córcoles, K. Temme, A. W. Harrow, A. Kandala, J. M. Chow, and J. M. Gambetta, "Supervised learning with quantum-enhanced feature spaces," *Nature*, vol. 567, no. 7747, pp. 209–212, Mar. 2019.

[49] M. Schuld and N. Killoran, "Quantum machine learning in feature Hilbert spaces," *Phys. Rev. Lett.*, vol. 122, no. 4, Feb. 2019, Art. no. 040504.

[50] R. K. Ahalya, S. Umapathy, P. T. Krishnan, and A. N. Joseph Raj, "Automated evaluation of rheumatoid arthritis from hand radiographs using machine learning and deep learning techniques," *Proc. Inst. Mech. Eng., H, J. Eng. Med.*, vol. 236, no. 8, pp. 1238–1249, Aug. 2022.

[51] A. K. Bishwas, A. Mani, and V. Palade, "An all-pair quantum SVM approach for big data multiclass classification," *Quantum Inf. Process.*, vol. 17, no. 10, pp. 1–16, Sep. 2018.

[52] J. R. Glick, T. P. Gujarati, A. D. Corcoles, Y. Kim, A. Kandala, J. M. Gambetta, and K. Temme, "Covariant quantum kernels for data with group structure," 2021, *arXiv:2105.03406*.

[53] B. Bullins, C. Zhang, and Y. Zhang, "Not-so-random features," 2017, *arXiv:1710.10230*.

[54] F. de Lima Marquezino, R. Portugal, and C. Lavor, *Grover's Algorithm for Unstructured Search*. Cham, Switzerland: Springer, 2019, pp. 35–55.

[55] Y. Tong, W. Lu, Q.-Q. Deng, C. Chen, and Y. Shen, "Automated identification of retinopathy of prematurity by image-based deep learning," *Eye Vis.*, vol. 7, no. 1, pp. 1–12, Aug. 2020.

[56] M. M. Fraz, A. R. Rudnicka, C. G. Owen, D. P. Strachan, and S. A. Barman, "Automated arteriole and venule recognition in retinal images using Ensemble classification," in *Proc. Int. Conf. Comput. Vis. Theory Appl. (VISAPP)*, Jan. 2014, pp. 194–202.

[57] Y. Han, W. Li, M. Liu, Z. Wu, F. Zhang, X. Liu, L. Tao, X. Li, and X. Guo, "Application of an anomaly detection model to screen for ocular diseases using color retinal fundus images: Design and evaluation study," *J. Med. Internet Res.*, vol. 23, no. 7, Jul. 2021, Art. no. e27822.

[58] M. Niemeijer, B. van Ginneken, and M. D. Abramoff, "Automatic classification of retinal vessels into arteries and veins," *Proc. SPIE*, vol. 7260, Feb. 2009, Art. no. 72601F.

[59] A. Desiani, S. Suprihatin, and S. B. Agustina, "A robust techniques of enhancement and segmentation blood vessels in retinal image using deep learning," *Biomed. Eng., Appl., Basis Commun.*, vol. 34, no. 4, Feb. 2022, Art. no. 2250019.

[60] E. Akpinar, N. M. Duc, and B. KesercI, "The role of quantum-enhanced support vector machine using multiparametric MRI parameters in differentiating medulloblastoma from ependymoma," in *Proc. IEEE Int. Conf. Quantum Comput. Eng. (QCE)*, Sep. 2022, pp. 882–885.

[61] J. J. Wang, "Generalized retinal arteriolar narrowing predicts 5-year cardio-vascular and cerebro-vascular mortality: Findings from the blue mountains eye study," *Investigative Ophthalmol. Visual Sci.*, vol. 43, no. 13, p. 4396, 2002.

[62] G. E. Quinn and A. Vinekar, "The role of retinal photography and telemedicine in ROP screening," *Seminars Perinatol.*, vol. 43, no. 6, pp. 367–374, Oct. 2019, doi: 10.1053/j.semperi.2019.05.010.

[63] H. Biten, T. K. Redd, C. Moleta, J. P. Campbell, S. Ostmo, K. Jonas, R. V. P. Chan, and M. F. Chiang, "Diagnostic accuracy of ophthalmoscopy vs telemedicine in examinations for retinopathy of prematurity," *JAMA Ophthalmol.*, vol. 136, no. 5, p. 498, May 2018, doi: 10.1001/jamaophthalmol.2018.0649.

[64] M. M. Adnan, M. S. M. Rahim, A. R. Khan, T. Saba, S. M. Fati, and S. A. Bahaj, "An improved automatic image annotation approach using convolutional neural network-slantlet transform," *IEEE Access*, vol. 10, pp. 7520–7532, 2022.

[65] A. Rehman, T. Saba, K. Haseeb, T. Alam, and J. Lloret, "Sustainability model for the Internet of Health Things (IoHT) using reinforcement learning with mobile edge secured services," *Sustainability*, vol. 14, no. 19, Sep. 2022, Art. no. 12185.

V. M. RAJA SANKARI received the B.E. degree in biomedical engineering from the Sri Ramakrishna Engineering College, Coimbatore, India, in 2018, and the M.Tech. degree in biomedical engineering from the SRM Institute of Science and Technology, Kattankulathur, India, in 2020, where she is currently pursuing the Ph.D. degree in biomedical engineering. She holds three patents, of which one patent is granted and two are in the published stage. Her research interests include medical image processing and instrumentation.

U. SNEKHALATHA (Member, IEEE) received the bachelor's degree in electronics and communication engineering from Madras University, in 2002, the master's degree in medical electronics from Anna University, in 2005, and the Ph.D. degree in biomedical engineering from SRMIST, Kattankulathur, in 2015. She is currently a Professor with the Department of Biomedical Engineering, SRMIST. She has 20 years of teaching experience. She has published 111 research articles in reputed peer-reviewed international journals and international conferences. She has authored a book titled *Artificial Intelligence-Based Infrared Thermal Image Processing and its Applications* (CRC Press Taylor and Francis Group). She has filed seven Indian patents, of which two patents are granted and five are in the published stage. Her research interests include biomedical signal processing, medical image processing, biomedical instrumentation, machine learning, and deep learning techniques. She obtained the Best Researcher Award for publications in *Nature* indexed journal during the Research Day function held on 1 March 2021. She has obtained the best paper award and gold medals for some research paper publications.

SULTAN ALASMARI received the master's degree from Depaul University, Chicago, in 2011, and the Ph.D. degree in software and information systems from the University of North Carolina at Charlotte, in 2022. He was the Information Security Manager in the financial sector. He is currently an Assistant Professor with the College of Computer and Information Sciences, Almajmaah University. His research interests include cybersecurity enforcement and smart health. He specializes in health analytics.

SHABNAM MOHAMED ASLAM (Member, IEEE) is an Associate Professor with the College of Computer and Information Sciences, Majmaah University, Saudi Arabia. Her specializations include soft computing, HCI, and cyber security. Her current research interests are analyzing machine learning and deep learning algorithms with applicability to different domains. Research interest in the fields of e-learning, cyber security, and digital forensics. She has overall 17 years of experience in academic teaching at different universities in India.