International Journal of Advanced Engineering Application (IJAEA)
Volume 3 Issue 11 Nov 2026
ISSN :3048-6807

Adaptive Quantum Feature Encoding for Interpretable and Efficient Retinal Diagnostics

Bonu Harish[1], Chebrolu Hanish[2], Karnam Likith Babu[3], Mrs. P. Padmavathi[4]

[123] Department of Information Technology, GMRIT Deemed to be University, Rajam, India

[4] Assistant Professor, Department of Information Technology, GMRIT Deemed to be University, Rajam, India

Email: harishbonu3@gmail.com

Abstract—

Diabetic retinopathy (DR) is a leading cause of avoidable vision loss, so timely and accurate grading matters. Manual examination of retinal images is slow, needs trained specialists and is hard to arrange where resources are limited, while many automated graders overlook subtle visual patterns, work with redundant data and explain their outputs poorly. This study presents RetinaFormer, a framework for five-class DR grading that aims to be accurate, compact and easy to interpret. A Vision Transformer (ViT) extracts features from fundus images, adaptive feature selection and Principal Component Analysis (PCA) remove redundant information, and a four-qubit Variational Quantum Circuit (VQC) learns a richer representation before classification. The framework also produces visual explanations that mark the retinal regions influencing each prediction. Trained and evaluated on APTOS 2019 and FGADR images, it reached 81.95% accuracy, an 80.09% macro F1-score and a quadratic weighted kappa of 90.94% on a held-out test set of 277 images. The system is presented as a research framework and not as a clinical diagnostic tool.

Keywords-Diabetic Retinopathy, Vision Transformer, Variational Quantum Circuit, Medical Image Classification, Explainable AI.

I. INTRODUCTION

Diabetic retinopathy is a diabetes-associated disorder characterized by damage to the blood vessels of the retina. As the disease progresses, retinal changes such as microaneurysms, hemorrhages and abnormal blood vessel patterns may appear. Early detection and proper clinical evaluation are important for reducing the risk of vision loss.

Traditionally, diabetic retinopathy is diagnosed by trained eye-care professionals who examine retinal fundus images. This takes considerable time and expertise, particularly when large numbers of images must be processed. Automated analysis of retinal images has therefore become an important research area in ophthalmology.

Deep learning approaches, especially Convolutional Neural Networks (CNNs), have demonstrated encouraging performance in retinal image classification by automatically learning visual features. However, CNNs mainly focus on local image patterns and may require additional mechanisms to model dependencies between spatially distant regions within an image.

Vision Transformers (ViTs) address this limitation by partitioning images into smaller patches and applying self-attention mechanisms to learn relationships between different regions. These models can capture global contextual information and have shown potential in diabetic retinopathy grading [1].

Despite these advances, challenges such as imbalanced datasets, variations in image quality, and subtle differences between disease stages remain. High-dimensional features may also increase computational requirements, while understanding model predictions is important for improving interpretability.

To address these limitations, we developed RetinaFormer, which combines Vision Transformer-based feature extraction, adaptive feature selection, Principal Component Analysis (PCA) and a four-qubit Variational Quantum Circuit (VQC), and adds visual explanations of the predictions. The main contributions of this work are: (i) a compact ViT–PCA–VQC pipeline for five-class DR grading; (ii) an explanation stage that highlights the retinal regions influencing each prediction; and (iii) an evaluation on APTOS 2019 and FGADR that reports the data split and measured class-wise and overall results.

II. LITERATURE REVIEW

Gulshan et al. [1] showed that a deep learning algorithm can detect diabetic retinopathy in fundus photographs, and Abràmoff et al. [3] studied automated analysis of retinal images for referable disease. Ting et al. [2] evaluated a deep learning system on images from several ethnic populations, which underlined the importance of generalization. Zhou et al. [4] introduced the FGADR

benchmark for lesion segmentation, grading and transferability, and IDRiD [14] is another public dataset for DR research. Convolutional backbones such as ResNet [6] and EfficientNet [7] became standard tools for this task, but Vision Transformers [8], which relate image patches through self-attention, have recently drawn attention for DR grading. Nazih et al. [5] applied a ViT to severity prediction, and later studies explored ViTs combined with residual attention [16], visual prompt learning [17] and softmax and pooling operators [19], as well as comparisons between ViTs and CNNs [18]. Explainable and fusion-based designs [15], [21], [22] and the explanatory framework of Quellec et al. [20] aim to make grading decisions easier to inspect, and a systematic review of explainable AI for DR is given in [25]. For feature compression we rely on PCA [9], which keeps the main directions of variation in the data. Quantum machine learning has been surveyed by Biamonte et al. [10], Schuld and Killoran [11] studied how classical data can be encoded in quantum feature spaces, and hybrid quantum–classical models have been applied to image classification [23], [24]. To explain predictions, attention rollout and attention flow [12] and Grad-CAM [13] trace which parts of an image drive a model's output. Building on these ideas, RetinaFormer integrates a Vision Transformer, PCA, a Variational Quantum Circuit and explainable AI to classify diabetic retinopathy into five severity levels, and it is evaluated here on held-out test data.

COMPARISON TABLE

REF.	AUTHOR AND YEAR	METHODOLOGY	MAJOR CONTRIBUTION	LIMITATION / RESEARCH GAP
[1]	GULSHAN ET AL. (2016)	DEEP CNN	DESIGNED AN AUTOMATED FRAMEWORK TO IDENTIFY DIABETIC RETINOPATHY IN RETINAL FUNDUS IMAGES.	PRIMARILY FOCUSES ON REFERABLE DR DETECTION RATHER THAN COMPREHENSIVE EXPLAINABLE FIVE-CLASS GRADING.
[2]	TING ET AL. (2017)	DEEP LEARNING	DEVELOPED AND EVALUATED A SYSTEM USING RETINAL IMAGES COLLECTED FROM DIVERSE ETHNIC POPULATIONS.	GENERALIZATION TO DIFFERENT DATASETS AND IMAGING CONDITIONS REQUIRES FURTHER EVALUATION.
[4]	ZHOU ET AL. (2021)	SEGMENTATION, GRADING AND TRANSFER LEARNING	INTRODUCED FGADR FOR RETINAL LESION SEGMENTATION, DR GRADING AND TRANSFERABILITY RESEARCH.	CROSS-DATASET DIFFERENCES CAN AFFECT MODEL PERFORMANCE AND TRANSFERABILITY.
[5]	NAZIH ET AL. (2023)	VISION TRANSFORMER	INVESTIGATED TRANSFORMER-BASED PREDICTION OF DIABETIC RETINOPATHY SEVERITY.	FURTHER INVESTIGATION OF EXPLAINABILITY AND HYBRID QUANTUM CLASSIFICATION IS POSSIBLE.
[8]	DOSOVITSKIY ET AL. (2021)	VISION TRANSFORMER (VIT)	INTRODUCED PATCH-BASED IMAGE REPRESENTATION AND SELF-ATTENTION FOR IMAGE RECOGNITION.	REQUIRES SUITABLE TRAINING DATA AND COMPUTATIONAL RESOURCES FOR EFFECTIVE APPLICATION.
[9]	JOLLIFFE AND CADIMA (2016)	PRINCIPAL COMPONENT ANALYSIS	REVIEWED DIMENSIONALITY REDUCTION FOR RETAINING IMPORTANT DATA VARIATION.	VARIANCE PRESERVATION DOES NOT NECESSARILY RETAIN ALL DISEASE-RELEVANT FEATURES.
[10]	BIAMONTE ET AL. (2017)	QUANTUM MACHINE LEARNING	REVIEWED QUANTUM COMPUTING APPROACHES AND THEIR POTENTIAL APPLICATIONS IN MACHINE LEARNING.	PRACTICAL IMPLEMENTATION FACES HARDWARE AND COMPUTATIONAL LIMITATIONS.
[12]	ABNAR AND ZUIDEMA (2020)	ATTENTION ROLLOUT AND ATTENTION FLOW	PROPOSED METHODS FOR ESTIMATING INFORMATION FLOW THROUGH TRANSFORMER ATTENTION LAYERS.	ATTENTION VISUALIZATIONS ALONE DO NOT GUARANTEE CLINICALLY CORRECT EXPLANATIONS.

III. METHODOLOGY

A. Overview of the Proposed Framework

The RetinaFormer pipeline is made up of several steps:

- Input retinal fundus image.
- Image preprocessing.
- Transformer-based feature extraction.
- Adaptive feature selection.
- PCA-based dimensionality reduction.
- Four-qubit Variational Quantum Circuit.
- Classical classification.
- Prediction visualization and explainability.

The full process can be written as

$I \longrightarrow P(I) \longrightarrow f_\theta\big(P(I)\big) \longrightarrow S(\cdot) \longrightarrow \mathrm{PCA}(\cdot) \longrightarrow Q_\phi(\cdot) \longrightarrow g_\psi(\cdot)$In this, I is the input image, P(I) is the preprocessed image, f_{θ} is the feature extraction network, S(·) is the feature selection function, PCA is used for dimensionality reduction, Q_{φ}(·) is the quantum circuit, and g_{ψ}(·) is the classification network.

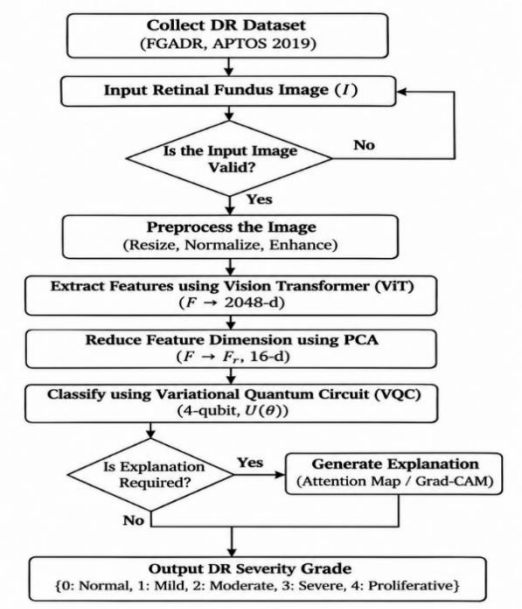

Fig. 1. Proposed RetinaFormer architecture, showing the stages from retinal image preprocessing to classification and explanation.

B. Image Preprocessing

Let I be the input retinal fundus image. Preprocessing standardizes the image before feature extraction, giving the preprocessed image

$$\tilde{I} = P(I).$$

Here P(·) covers resizing the image to the input size expected by the backbone, normalizing the pixel values, and applying image-quality operations. Any data augmentation was restricted to the training images, so the validation and test images were left unchanged.

Image size, normalization and quality filtering were kept identical for the training, validation and test images.

C. Transformer-Based Feature Extraction

The preprocessed image is processed through a feature extraction network. The feature extraction can be written as

$$z = f_\theta(\tilde{I})$$

where $\tilde{I}$ represents the preprocessed image, $f\theta$ denotes the feature extraction network, and θ represents its parameters.

The resulting feature vector is defined as

$$z \in \mathbb{R}^{2048}$$

.

In other words, every image is described by 2,048 numerical values produced by the Transformer-based feature extractor.
This representation is the input to the feature selection step.

D. Adaptive Feature Selection

The extracted representation may contain redundant or irrelevant components. Adaptive feature selection keeps only the features that contribute most to the classification.
The selected features are given by

$$z_s = S(z),$$

where S(·) is the method used for feature selection and z_s is the resulting feature representation.
Adaptive feature selection scored the 2,048 extracted features and kept the most informative ones, which were then passed to PCA. The selector was fitted on the training data only, and the validation set was used to choose the final configuration.

E. Principal Component Analysis

PCA is applied to transform the selected features into a lower-dimensional, more compact representation.
The dimensionality reduction process is expressed

$$u = \mathrm{PCA}(z_s)$$

where u is the reduced feature vector.
In the proposed setup,

$$u \in \mathbb{R}^{16}$$

Thus, the pipeline transforms the feature representation into 16 numerical values before quantum encoding.
PCA was fitted on the training set, and the same transformation was then applied unchanged to the validation and test sets. The 16 retained components form the input to the quantum circuit.

F. Quantum Feature Encoding and Variational Circuit

The compressed feature representation is then mapped into a quantum state. In a system comprising four qubits, this encoding is expressed as follows:

$$|\psi(u)\rangle = U_{\text{enc}}(u)|0\rangle^{\otimes 4},$$

where U_{enc} is the feature encoding operation and |0⟩^{⊗4} is the initial state of the four qubits.
This encoded state is then processed by a parameterized quantum circuit.

$$|\psi_\phi(u)\rangle = U_{\text{var}}(\phi)|\psi(u)\rangle$$

where U_{var}(φ) is the variational quantum circuit and φ are its trainable parameters.
When we measure the quantum state, we get a classical feature vector. For instance, using Pauli-Z measurements on the four qubits,

$$q = [\langle Z_1\rangle, \langle Z_2\rangle, \langle Z_3\rangle, \langle Z_4\rangle],$$

where each expectation value is defined as

$$\langle Z_i\rangle = \langle \psi_\phi(u)|Z_i|\psi_\phi(u)\rangle$$

Here, Z_i denotes the Pauli-Z operator applied to qubit *i*, while *q* represents the four measurement outcomes, one from each qubit.
Implementation details: the 16 PCA features were encoded into the four-qubit circuit by the encoding operation U_enc, followed by the trainable variational block U_var(φ). Pauli-Z expectation values were then measured on all four qubits, and the resulting four-value vector q was passed to the classical classifier.

G. Classical Classification

The quantum measurement vector is then passed to a classical classification network.
The classification step is written as

$$\hat{y} = g_\psi(q)$$

where g_ψ represents the classical classification network, and ψ are its trainable parameters.
For a five-class classification task, the output probability vector is

$$\hat{y} = [p_0, p_1, p_2, p_3, p_4],$$

with the constraint that

$$\sum_{k=0}^{4} p_k = 1$$

The predicted category is selected by identifying the class with the highest probability:

$$\hat{c} = \arg \max_{k \in \{0,1,2,3,4\}} p_k$$

The five output classes are:

- Class 0: Absence of diabetic retinopathy (No DR).
- Class 1: Mild diabetic retinopathy.
- Class 2: Moderate diabetic retinopathy.
- Class 3: Severe diabetic retinopathy.
- Class 4: Proliferative diabetic retinopathy (PDR).

The model's predicted class is an automated classification result and not a confirmed clinical diagnosis.

H. Model Training

The model was trained on annotated retinal images as a multiclass problem. The standard cross-entropy loss is defined as follows:

$$L = -\frac{1}{N} \sum_{i=1}^{N} \sum_{k=0}^{4} y_{ik} \log(p_{ik})$$

Here, N denotes the total number of training samples, y_{ik} represents the actual class label for class k in sample i, and p_{ik} indicates the model's predicted probability for that class.

The classical classification layers and the quantum circuit parameters were trained with the cross-entropy loss defined above. The training set was used to update the parameters, and the validation set was used to select the final model.

FGADR and APTOS 2019 were both used, with their grading labels mapped to the same five-grade scale. The splits were made to avoid data leakage, and the test set was not used for any tuning decision.

I. Explainability

The explainability stage is designed to help users examine the retinal areas that have the greatest influence on the model's predicted outcome.

Because the backbone is a Vision Transformer, explanations were generated with attention-based attribution such as attention rollout [12]. Gradient-based methods such as Grad-CAM [13] are the natural alternative for CNN backbones.

The resulting heatmap is overlaid on the retinal image to highlight regions with higher attribution under the chosen method.

The heatmap describes the model's behavior. It is not evidence of a specific lesion, and it does not confirm that the predicted stage is clinically correct.

J. Web-Based Prototype

The framework was integrated into the RetinaFormer web application to demonstrate the workflow. The interface takes a retinal fundus image as input, displays the predicted class with its probabilities, shows the explanation overlay and reports any processing errors.

IV. Experimental Setup

A. Datasets

Two public datasets were used in this study.

1) FGADR [4]: provides image-level DR grades together with lesion-level annotations. It was used in this study together with APTOS 2019, in line with the dataset's terms of use.

2) APTOS 2019 [26]: a collection of retinal fundus images graded into five severity levels. It served as the main five-class evaluation set.

Image counts and labels were checked against the image and label files themselves and not taken from the archive descriptions.

B. Data Partitioning

The data were divided into training, validation and test subsets. The training set was used to fit the model parameters, the validation set for hyperparameter tuning and model selection, and the test set only for the final evaluation.

The test set contained 277 images, distributed across the five grades as follows: No DR 15, Mild DR 32, Moderate DR 90, Severe DR 97 and Proliferative DR 43. The same preprocessing was applied to all three subsets.
The test set was kept untouched until the final evaluation.

C. Evaluation Metrics

Several metrics were used, because a single score can be misleading and can hide poor performance on under-represented classes.
Accuracy is calculated as the ratio of correctly classified samples to the total number of samples:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

for binary classification. For the five-class task, overall accuracy is the percentage of correctly classified samples out of all samples classified.
Precision and recall were computed for each class as follows:

$$\text{Precision} = \frac{TP}{TP + FP}, \quad \text{Recall} = \frac{TP}{TP + FN}$$

The F1-score is

$$F_1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

Macro-F1 is the simple average of the F1-scores of the five classes, so every class carries equal weight. Because DR grades are ordinal, the quadratic weighted kappa (QWK) was also reported, together with weighted F1-score, balanced accuracy and the Matthews correlation coefficient (MCC). Per-class recall corresponds to sensitivity.

V. RESULTS & DISCUSSIONS

RetinaFormer was evaluated on a held-out test set of 277 images that was kept separate from training and validation. The tables below report the measured results.

Performance metric	Calculated value
Accuracy	81.95%
Macro Precision	79.97%
Macro Recall	80.23%
Macro F1-score	80.09%
Weighted F1-score	81.98%
Balanced Accuracy	80.23%
QWK	90.94%
MCC	75.35%

Overall performance metrics

DR class		Precision	Recall	F1-score	Support
No DR		0.800	0.800	0.800	15
Mild DR		0.697	0.719	0.708	32
Moderate DR		0.833	0.833	0.833	90
Severe DR		0.854	0.845	0.850	97
Proliferative DR		0.814	0.814	0.814	43

Class-wise performance metrics

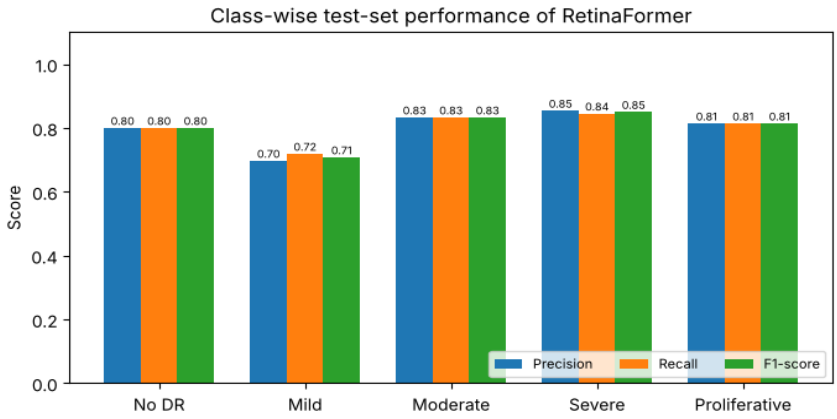

Fig. 2. Class-wise precision, recall and F1-score on the test set.

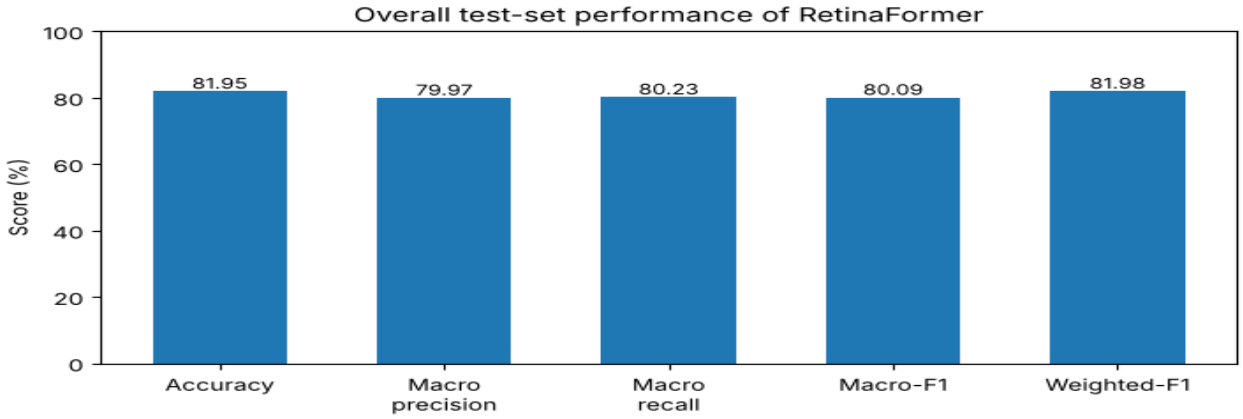

Fig. 3. Overall test-set metrics of RetinaFormer.

RetinaFormer reached an accuracy of 81.95%, a macro F1-score of 80.09% and a quadratic weighted kappa of 90.94% on the test set. The high QWK is consistent with errors falling mostly between neighboring grades rather than distant ones. Severe DR (F1 0.850) and Moderate DR (F1 0.833) were handled best, while Mild DR was the hardest class (F1 0.708), which is plausible given that its lesions are small and easily confused with the adjacent grades. The No DR class has only 15 test images, so its figures are less stable than those of the larger classes.

Ablation Study

An ablation study is used to understand the contribution of individual components of the RetinaFormer architecture.

Different configurations are compared by removing or replacing selected components while keeping the experimental conditions the same.

The following configurations are considered:

Configuration	Description	Purpose
A0	Complete RetinaFormer	Evaluates the complete proposed framework.
A1	Without adaptive feature selection	Examines the contribution of feature selection.
A2	Without PCA	Examines the effect of dimensionality reduction.
A3	VQC replaced with MLP	Compares quantum and classical processing.
A4	Classical ViT-based baseline	Provides a classical reference model.

Each configuration is assessed with accuracy, macro F1-score and QWK on the same test set.

The comparison helps show how individual components influence classification performance and computational requirements.

Conclusions from the ablation are drawn only from measured results and not from assumed improvements.

Discussions

RetinaFormer pairs the global context captured by a Vision Transformer with a compact feature pipeline (feature selection and PCA) and a hybrid quantum–classical classifier.

Reducing the representation to 16 features and using four qubits was a design choice. It does not by itself guarantee that useful information is preserved, or that the quantum part is cheaper than a classical one. This is why the VQC is best compared directly with a classical MLP and a classical ViT baseline.

The explanation maps help show which regions the model relied on for a prediction. They have not been validated against clinician-marked lesions, so they should be read as a view into model behavior and not as a clinical finding.

The main limitations are the small test set (277 images, with only 15 No DR cases), which makes per-class estimates, especially for No DR and Mild DR, somewhat uncertain; results from a single data split; and the absence of cross-dataset and clinical validation. The results support RetinaFormer as a research framework, not as a diagnostic tool.

VI. CONCLUSIONS AND FUTURE DIRECTIONS

In this paper we presented RetinaFormer, a hybrid classical–quantum framework that grades diabetic retinopathy severity from retinal fundus images. It combines a Vision Transformer feature extractor, adaptive feature selection, PCA, a four-qubit Variational Quantum Circuit and a classical classifier, together with an explanation stage that helps users inspect predictions.

On a held-out test set of 277 images, the framework achieved 81.95% accuracy, an 80.09% macro F1-score and a QWK of 90.94%, performing best on Severe and Moderate DR and least well on Mild DR.

Future work includes handling class imbalance more directly, testing across datasets, measuring computational cost with larger circuits or real quantum hardware, studying how reliable the explanations are, and carrying out independent clinical validation.

These results come from a single dataset configuration and do not demonstrate quantum advantage or clinical effectiveness. Both would need repeatable evidence from further experiments.

VII. REFERENCES

[1] V. Gulshan *et al*., “Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs,” *JAMA*, vol. 316, no. 22, pp. 2402–2410, 2016, doi: 10.1001/jama.2016.17216.

[2] D. S. W. Ting *et al*., “Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes,” *JAMA*, vol. 318, no. 22, pp. 2211–2223, 2017, doi: 10.1001/jama.2017.18152.

[3] M. D. Abràmoff *et al*., “Automated analysis of retinal images for detection of referable diabetic retinopathy,” *JAMA Ophthalmology*, vol. 131, no. 3, pp. 351–357, 2013, doi: 10.1001/jamaophthalmol.2013.1743.

[4] Y. Zhou, B. Wang, L. Huang, S. Cui, and L. Shao, “A benchmark for studying diabetic retinopathy: Segmentation, grading, and transferability,” *IEEE Transactions on Medical Imaging*, vol. 40, no. 3, pp. 818–828, 2021, doi: 10.1109/TMI.2020.3037771.

[5] W. Nazih, A. O. Aseeri, O. Y. Atallah, and S. El-Sappagh, “Vision Transformer model for predicting the severity of diabetic retinopathy in fundus photography-based retina images,” *IEEE Access*, vol. 11, pp. 117546–117561, 2023, doi: 10.1109/ACCESS.2023.3326528.

[6] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.

[7] M. Tan and Q. V. Le, “EfficientNet: Rethinking model scaling for convolutional neural networks,” in *Proc. 36th Int. Conf. Machine Learning (ICML)*, 2019, pp. 6105–6114.

[8] A. Dosovitskiy *et al*., “An image is worth 16×16 words: Transformers for image recognition at scale,” in *Proc. Int. Conf. Learning Representations (ICLR)*, 2021.

[9] I. T. Jolliffe and J. Cadima, “Principal component analysis: A review and recent developments,” *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 374, no. 2065, Art. no. 20150202, 2016, doi: 10.1098/rsta.2015.0202.

[10] J. Biamonte *et al*., “Quantum machine learning,” *Nature*, vol. 549, pp. 195–202, 2017, doi: 10.1038/nature23474.

[11] M. Schuld and N. Killoran, "Quantum machine learning in feature Hilbert spaces," *Physical Review Letters*, vol. 122, no. 4,Art.no.040504,2019,doi:10.1103/PhysRevLett.122.040504.
[12] S. Abnar and W. Zuidema, "Quantifying attention flow in transformers," in *Proc. 58th Annual Meeting of the Association for Computational Linguistics*, 2020, pp. 4190–4197, doi: 10.18653/v1/2020.acl-main.385.
[13] R. R. Selvaraju *et al*., "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *Proc. IEEE Int. Conf. Computer Vision (ICCV)*, 2017, pp. 618–626, doi: 10.1109/ICCV.2017.74.
[14] A. Porwal *et al*., "Indian diabetic retinopathy image dataset (IDRiD): A database for diabetic retinopathy screening research," *Data*, vol. 3, no. 3, Art. no. 25, 2018, doi: 10.3390/data3030025.
[15] M. Rahaman, M. A. Muni, S. Tasnim, S. S. Juthi, R. Islam, and K. R. Ahmed, "Explainable AI for diabetic retinopathy detection using vision transformers," *Scientific Reports*, vol. 16, Art. no. 27354, 2026, doi: 10.1038/s41598-026-56569-4.
[16] Z. Gu et al., "Classification of diabetic retinopathy severity in fundus images using the vision transformer and residual attention," Computational Intelligence and Neuroscience, vol. 2023, Art. no. 1305583, 2023, doi: 10.1155/2023/1305583.
[17] Y. Zhang et al., "Semantic-oriented visual prompt learning for diabetic retinopathy grading on fundus images," IEEE Transactions on Medical Imaging, vol. 43, no. 8, pp. 2960–2969, 2024, doi: 10.1109/TMI.2024.3383827.
[18] "Comparative analysis of vision transformers and conventional convolutional neural networks in detecting referable diabetic retinopathy," Ophthalmology Science, vol. 4, no. 6, Art. no. 100552, 2024, doi: 10.1016/j.xops.2024.100552.
[19] C. Liu et al., "Lesion classification and diabetic retinopathy grading by integrating softmax and pooling operators into vision transformer," Frontiers in Public Health, vol. 12, Art. no. 1442114, 2025, doi: 10.3389/fpubh.2024.1442114.
[20] G. Quellec et al., "ExplAIn: Explanatory artificial intelligence for diabetic retinopathy diagnosis," Medical Image Analysis, vol. 72, Art. no. 102118, 2021, doi: 10.1016/j.media.2021.102118.
[21] K. A. Alavee et al., "Enhancing early detection of diabetic retinopathy through the integration of deep learning models and explainable artificial intelligence," IEEE Access, vol. 12, pp. 73950–73969, 2024, doi: 10.1109/ACCESS.2024.3405570.
[22] A. Ikram and A. Imran, "ResViT FusionNet model: An explainable AI-driven approach for automated grading of diabetic retinopathy in retinal images," Computers in Biology and Medicine, vol. 186, Art. no. 109656, 2025, doi: 10.1016/j.compbiomed.2025.109656.
[23] J. Landman et al., "Quantum methods for neural networks and application to medical image classification," Quantum, vol. 6, Art. no. 881, 2022, doi: 10.22331/q-2022-12-22-881.
[24] T.-H. Qiu and S.-M. Wang, "An image classification algorithm based on hybrid quantum classical convolutional neural network," Quantum Engineering, 2022, doi: 10.1155/2022/5701479.
[25] T. Ngwazi, B. Ndlovu, and K. Maguraushe, "Early detection of diabetic retinopathy through explainable AI models: A systematic review," International Journal on Informatics for Development, vol. 14, no. 2, pp. 616–628, 2025, doi: 10.14421/ijid.2025.5200.
[26] Asia Pacific Tele-Ophthalmology Society (APTOS), "APTOS 2019 Blindness Detection," Kaggle, 2019. [Online]. Available: https://www.kaggle.com/c/aptos2019-blindness-detection