RESEARCH PAPER

Quantum-Enhanced Deep Feature Learning for Automated Leukemic Blood Cell Identification Using VQC–CNN Integration

A S Manjunath[1] and Meghana J[2,*], Tejaswi K[3]

[1]Department of Computer Applications, JSS Science and Technology University, India
[2]Department of Computer Applications, JSS Science and Technology University, India
[1]as.manjunath@jssstuniv.in and [2]meghanaj@jssstuniv.in and tejaswi@jssstuniv.in
***Corresponding author:** Dr. Meghana J, meghanaj@jssstuniv.in*

ABSTRACT

Accurate detection and classification of WBC is essential haematological diagnosis, especially to detect the development of lymphoblastic leukaemia. Class imbalance, limited interpretability, and high computational demands are always prevalent issues in classical machine and deep learning. In this paper, a variational quantum classifer (VQC) based hybrid quantum–classical approach to five-class WBC subtype identification (Basophil, Eosinophil, Lymphocyte, Monocyte, Neutrophil) and four-stage detection of lymphoblastic leukaemia (Benign, Early, Pre-leukaemic, Pro-leukaemic) is proposed. The proposed pipeline combines the use of CNN based feature extraction, PCA dimensionality reduction and amplitude encoding within a 3-qubit quantum system parameterised by multi-layered RY rotations, controlled-Z (CZ) and controlled-phase (CPHASE) gates, with optimisation performed using the AdamW algorithm with early stopping. The approaches used to cope with class imbalance are SMOTE and inverse-frequency weighted cross-entropy loss. Explanation is given using SHAP DeepExplainer and gradient based saliency map. The model classifies WBC into 4 categories with an accuracy of 96.8% on macro-averaging and detects the stage of a disease with an accuracy of 95.4% while having only 25-30 trainable parameters, outperforming classical classifiers such as SVM, Random Forest, standard CNN and MobileNetV2. The results show that quantum inspired methods provide a promising route to scalable, parameter-efficient, and interpretable biomedical image analysis.

Keywords: Quantum Machine Learning, Variational Quantum Classifier, White Blood Cell Classification, Lymphoblastic Leukaemia, Amplitude Encoding, SMOTE, SHAP Explainability, NISQ.

How to cite this article: Manjunath AS, Meghana J. Quantum-Enhanced Deep Feature Learning for Automated Leukemic Blood Cell Identification Using VQC–CNN Integration. Int J Drug Deliv Technol. 2026;16(5): 1606-1616. DOI: 10.25258/ijddt.16.5.146

Source of support: Nil.

Conflict of interest: None

INTRODUCTION

Morphological classification of peripheral blood white blood cells (WBC) in stained images plays a key role in the diagnosis of haematology. WBC counts and morphologies are the most significant abnormalities seen in anaemia, systemic infection and haematological malignancies, particularly the spectrum of acute lymphoblastic leukaemia (ALL) [1]. Although accurate, manual microscopy is time-consuming, subjective, and requires specialized expertise that is not necessarily available in a clinical setting, particularly in low-resource areas, making automation useful both clinically and economically.

**Author for Correspondence: meghanaj@jssstuniv.in*

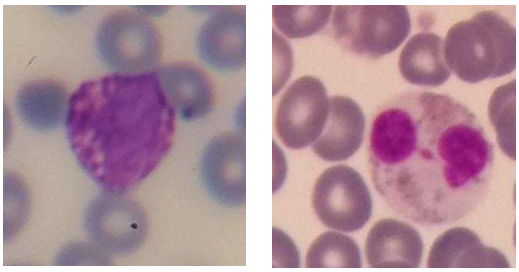
Recently, several variants of the CNN have been introduced and developed that have enhanced the WBC classification is in the state of art, and some of these architectures have achieved more than 98% accuracy on curated benchmark datasets [2,3]. Recent developments include hybrid Swin Transformer– ConvMixer architectures [4] that further advanced the classification task by capturing long-range cytological dependency and attention-based dual-path networks [5] that also captured long-range dependency by modeling the interaction between different parts of the image. However, there are three significant challenges in the real clinical data: (i) severe class imbalance, where

models tend to learn the majority class; (ii) lack of transparency in the decision boundaries, which can make it difficult to gain regulatory approval and trust from clinicians; (iii) rapidly growing parameter numbers, which can increase the cost of inference, the memory space required, and the likelihood of overfitting, especially in the medical domain where data is scarce.

In addition to this, QML has been proposed as a complementary methodology which utilizes the quantum superposition and entanglement to represent high-dimensional spaces with exponentially fewer parameters than classical machine learning (CMl) [6]. Variational Quantum Circuits (VQCs) are trainable workhorses for near term QML and can be adapted to existing Noisy NISQ devices, as the angles are optimized using classical optimisers [7,8]. These advantages of quantum representations have recently been verified in feasibility studies [9] which showed that 4-qubit VQCs could classify leukaemia cells with 83% accuracy, with only 50 training samples per class, as compared to 500 required by CNNs.

This work makes the following contributions.

- A complete quantum–classical pipeline validated on two publicly available clinical datasets (18,473 images) for the five-class identification of WBCs and four-stage detection of leukaemia.
- QSR of 8-dimensional PCA feature vectors into a 3-qubit VQC, with the basis amplitudes being encoded in the amplitude of the state, using CZ and CPHASE entanglement.
- A hybrid strategy where the class-imbalance problem is addressed both by oversampling the minority classes with SMOTE and by applying an inverse-frequency weighted cross-entropy loss for minority classes that have less than 90 raw training samples.
- Dual-modality explainability through SHAP DeepExplainer attribution maps and gradient-based saliency contours, for audit-ready prediction for clinical deployment.



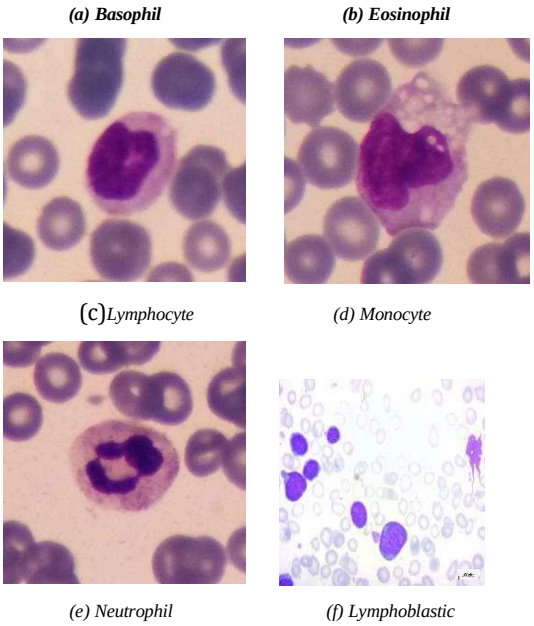


Fig. 1. Representative Wright-Giemsa stained blood smear images: (a–e) five normal WBC subtypes targeted for classification; (f) lymphoblastic leukaemia cell targeted for disease-stage detection

RELATED WORK

Classical and quantum research has been sustained in the classification of blood cells. The classical works by Sharma et al. [2] showed that it is possible to obtain high accuracy for multi-class WBC classification using an appropriate augmentation with deep CNNs, and the works by Üzen and Fırat [4] highlighted the improvement of accuracy with a multipath Swin Transformer–ConvMixer hybrid network architecture, capturing both the local texture and the global morphological context. Kılıç [5] introduced a dual-path ConvNeXt–Swin Transformer network achieving state-of-the-art results by modelling long-range cytological dependencies. Li et al. [10] proposed CCE-YOLOv7 incorporating Conv2Former backbones and CARAFE upsampling for real-time WBC detection in Fourier ptychographic microscopy images. A global scoping review by Asghar et al. [11] covering 136 studies (2006–2023) found that DL approaches consistently achieve 95–98% accuracy on large datasets, while classical ML (SVM, Random Forest) reaches up to 99% on well-curated feature sets.

In the quantum domain, Hossain et al. [12] pioneered VQC application to malaria detection in blood smear images, demonstrating competitive performance on small datasets by mapping features into a quantum Hilbert space. Ahmad et al. [1] showed that quantum-inspired deep feature selection reduces dimensionality and improves leukaemia classification scalability. A recent hybrid VQC–EfficientNet-B0 model [13] trained on the full BloodMNIST dataset (17,092 images, 8 classes) demonstrated that deeper qubit systems scale well with

richer architectures. Bano and Liebovitch [9] established NISQ-era benchmarks showing VQCs achieve stable 83% accuracy on AML detection with 50 samples per class—five-fold more data-efficient than CNNs. Surveys of XAI–QC convergence [14] and quantum computing in medicine [15] identify interpretability as the critical unresolved challenge for clinical adoption, motivating the SHAP integration in the present work.

The present work differs from prior QML studies by (i) addressing multi-class classification (5 WBC types + 4 disease stages) rather than binary tasks; (ii) combining SMOTE with weighted loss for clinically realistic class imbalance; (iii) providing dual-modality explainability aligned with regulatory requirements; and (iv) performing systematic comparison compared with four traditional baseline methods under the same experimental setup.

METEDOLOGY

The suggested hybrid quantum–classical framework proceeds through six stages: (1) data acquisition and stratified splitting, (2) image pre-processing and morphological isolation, (3) PCA-based dimensionality reduction and feature normalisation, (4) amplitude encoding into quantum state space, (5) variational quantum circuit forward pass and measurement, and (6) classical loss computation, back-propagation of gradients to circuit parameters, and model evaluation. Fig. 2 shows the complete architecture encompassing both the classification and explainability pathways.

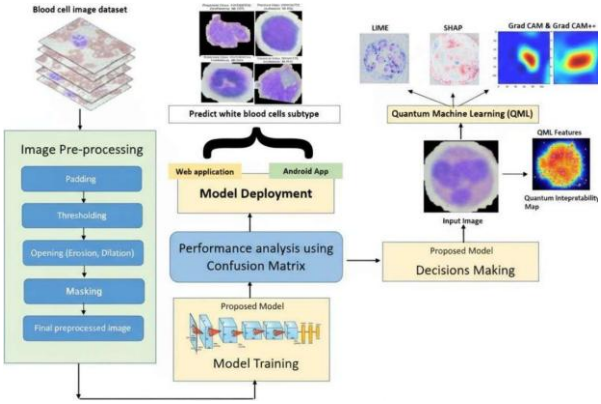


Fig. 2. Proposed hybrid quantum–classical system architecture. The pipeline integrates image pre-processing, CNN feature extraction, PCA reduction, VQC-based classification, SHAP/Grad-CAM explainability, and performance evaluation via confusion matrices

Dataset Description

Two complementary datasets were employed. The WBC classification corpus contains 12,444 high-resolution (512×512 px) Wright-Giemsa stained microscopic images across five WBC classes. The disease-stage corpus contains 6,029 images distributed across four lymphoblastic leukaemia progression labels. Both datasets were obtained from publicly available biomedical repositories ensuring reproducibility and open-access compliance. Stratified splitting (70/15/15 for train/validation/test) was applied using scikit-learn's StratifiedShuffleSplit to preserve class proportions. Tables 1 and 2 detail the distributions.

Table 1. WBC dataset class distribution and morphological descriptions

ID	Class	Total	Train	Test	Key Morphology
C0	Basophil	3,120	2,184	468	Bilobed nucleus, reddish-orange cytoplasmic granules
C1	Eosinophil	3,103	2,173	465	Large round nucleus, thin pale blue cytoplasm rim
C2	Lymphocyte	3,098	2,170	464	Kidney-shaped nucleus, scant agranular cytoplasm
C3	Monocyte	3,123	2,187	468	Horseshoe/kidney nucleus, abundant pale cytoplasm
C4	Neutrophil	3,100	2,180	465	Multi-lobed (2–5) nucleus, light purple granules
—	**Total**	**12,444**	**8,714**	**1,865**	

Table 2. Disease-stage dataset class distribution

ID	Disease Stage	Total	Train	Test	Key Histological Features
D0	Benign	1,500	1,050	225	Normal morphology, intact nuclear-cytoplasmic boundaries
D1	Early ALL	1,520	1,064	228	Slight nuclear enlargement, mild chromatin irregularity
D2	Pre-Leukaemic	1,487	1,040	223	Elevated N:C ratio, dense chromatin aggregation
D3	Pro-Leukaemic	1,522	1,065	229	Highly irregular nuclei, prominent nucleoli
—	**Total**	**6,029**	**4,219**	**906**	

Image Pre-processing Pipeline

Pre-processing must be consistent and reproducible to ensure that the 8-dimensional feature vector that is fed into the quantum circuit is free from artefacts of background and staining intensity, but still informative about the

discriminative cellular morphology. The pipeline consists of six "deterministic" steps:

Step 1 — Image Resizing: Each image is scaled to 128 × 128 pixels through bilinear interpolation, which normalises the spatial scale across acquisition conditions.

Step 2 — Greyscale conversion: The three RGB channels are combined together under the conventional photometric formula: Y = 0.299R + 0.587G + 0.114B, resulting in the preservation of nuclear texture and disregard of the variance of stain colour.

Step 3 — Gaussian smoothing using a 3×3 kernel of size σ = 1.0 to reduce the effects of high frequency noise and acquisition artefacts without affecting the morphological edges.

Step 4 — Otsu thresholding + morphological opening: Otsu's method [16] computes an optimal global threshold T* by maximising inter-class variance:

$$T^* = \text{argmax}_T \{ \omega_0(T)\cdot\sigma_0^2(T) + \omega_1(T)\cdot\sigma_1^2(T) \} \quad (1)$$

where ω_0, ω_1 are class probabilities and σ_0^2, σ_1^2 are intra-class variances. Morphological opening (erosion followed by dilation) then removes small spurious foreground fragments while preserving cell geometry.

Step 5 — Binary masking: In this step, the thresholded binary mask is multiplied with the greyscale image, such that only the cell of interest is preserved and the rest is removed.

Step 6 — Normalisation and PCA feature extraction: Normalised pixel intensities are flattened and then PCA is applied to obtain the principal components that explain the most variance, and to reduce the 128×128 = 16,384-dimensional representation to an 8-dimensional feature vector $\bar{x} \in \mathbb{R}^8$. The features are then rescaled to [0, π] for amplitude encoding compatibility.

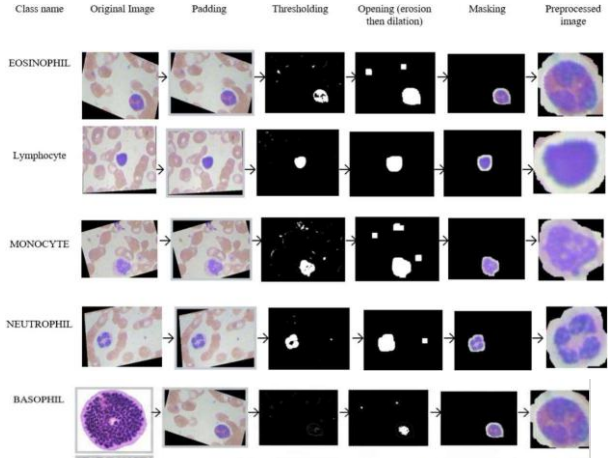


Fig. 3. Image pre-processing pipeline for all five WBC classes (top rows) and lymphoblastic disease cells (bottom strip).

Handling Class Imbalance

Imbalanced class distribution is a known problem for clinical datasets [11]. There are more Neutrophil images (3,100) than Basophil samples (89 in the unprocessed acquisition split). There are two complementary strategies to solve this:

1. **SMOTE Oversampling:**

   To avoid data leakage, the Synthetic Minority Oversampling Technique (SMOTE) [17] is used on the training partition, but not on the validation or test partitions. For each minority class instance x_i, SMOTE selects one among its five nearest neighbours x_j and creates a synthetic data point through interpolation between them.

$$\tilde{x} = x_i + \lambda\cdot(x_j - x_i), \quad \lambda \sim Uniform[0, 1] \quad (2)$$

This creates synthetic feature vectors in the convex hull of the minority class, in addition to the counts of Basophil and Monocyte, complementing and augmenting the majority class counts without additional overfitting risk that would be present with simple duplication.

2. **Weighted Cross-Entropy Loss**

   Errors are weighted more heavily on under-represented classes by adding class-frequency weights to the standard cross-entropy objective. The inverse-frequency weight of a problem C is:

$$w_c = N / (C \cdot n_c) \quad (3)$$

The weighted loss function is then:

$$L_w = - (1/N) \cdot \Sigma_{i=1}^{N} \Sigma_{c=1}^{C} w_c \cdot y_{i,c} \cdot log(\hat{p}_{i,c}) \quad (4)$$

where y_{i,c} is a one hot true label and p̂_{i,c} is the predicted softmax class c probability. The weights for the WBC task are: w_Basophil = 2.0, w_Monocyte = 1.8, w_Neutrophil = 0.7. The combination of SMOTE and weighted loss was validated by tracking validation-set macro-F1 after each epoch, confirming improved minority-class recall without degrading majority-class precision.

3. **CNN Feature Extractor**

A Convolutional Neural Network serves as the classical feature-extraction front-end before quantum encoding. The architecture consists of: (i) two convolutional blocks (Conv2D → BatchNorm → ReLU → MaxPool) with 32 and 64 filters respectively, each with 3×3 kernels; (ii) a global average pooling layer producing a 64-dimensional representation; and (iii) a fully connected dense layer projecting to 8 features. The receptive field grows progressively, enabling the network to capture both fine-grained nuclear texture (edges, granule density) and coarse morphological structure (nucleus shape, cytoplasm extent). The 8-dimensional output is then PCA-normalised to [0,π] and fed to the quantum circuit.

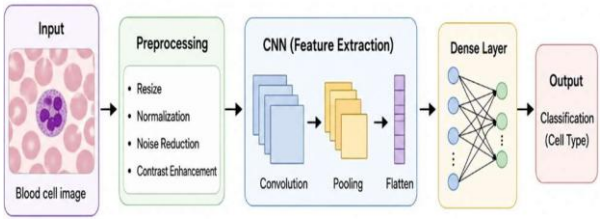


Fig. 4. CNN feature extractor architecture. Two convolutional blocks capture texture and morphology; the dense layer compresses to the 8-dimensional vector that is amplitude-encoded into the 3-qubit quantum state.

4. Quantum Circuit Design and Amplitude Encoding

A 3-qubit VQC is adopted because $2^3 = 8$ basis states precisely match the feature-vector dimensionality, enabling lossless amplitude encoding. This section details the encoding scheme, gate parameterisation, and circuit expressivity.

5. Amplitude Encoding

Amplitude encoding maps normalised feature vector $x \in \mathbb{R}^8$ to the quantum state:

$$|\psi\rangle = \Sigma_{i=0}^{7} \alpha_i |i\rangle, \quad \alpha_i = \bar{x}_i / \|\bar{x}\|_2 \tag{5}$$

where unit norm is enforced by $|x|_2 = (\Sigma_i x_i^2)^{1/2} + \varepsilon$, $\varepsilon = 10^{-8}$ to prevent division by zero. Each amplitude α_i encodes one principal-component feature, and the full quantum state simultaneously represents all 8 features in a single coherent superposition—a key advantage over classical vector representations that process features sequentially.

6. Variational Layer Structure

The circuit comprises L=5 variational layers. Each layer l applies three gate types in sequence:

- RY rotation gates — single-qubit parameterised rotations applied to each qubit:

$$RY(\theta) = \exp(-i\,\theta/2 \cdot Y) = \begin{bmatrix} \cos(\theta/2) & -\sin(\theta/2) \\ \sin(\theta/2) & \cos(\theta/2) \end{bmatrix} \tag{6}$$

- Controlled-Z entangling gates — two-qubit gates between adjacent pairs q[0]–q[1] and q[1]–q[2]:

$$CZ = \mathrm{diag}(1, 1, 1, -1) \in \mathbb{C}^{4\times4} \tag{7}$$

- Controlled-Phase gates — learnable phase rotations adding expressivity beyond CZ:

$$CPHASE(\varphi) = \mathrm{diag}(1, 1, 1, e^{i\varphi}) \in \mathbb{C}^{4\times4} \tag{8}$$

The full layer unitary is:

$$U_l(\theta_l,\varphi_l) = CPHASE(\varphi_l) \cdot CZ_{01} \cdot CZ_{12} \cdot [\,RY(\theta_{l,0}) \otimes RY(\theta_{l,1}) \otimes RY(\theta_{l,2})\,] \tag{9}$$

After L layers, the output state is $|\psi_{out}\rangle = U_L \cdots U_1 |\psi\rangle$. The predicted class probability for class k is:

(10)

Total trainable parameters = 3 qubits × (1 RY angle + 1 CPHASE angle) × L = 30 for L=5 layers—five orders of magnitude fewer than MobileNetV2's 3.4M parameters. Entangling gates foster qubit correlations that enable the circuit to capture non-linear decision boundaries in the high-dimensional Hilbert space without requiring additional qubits. Fig. 5 illustrates the full circuit layout.

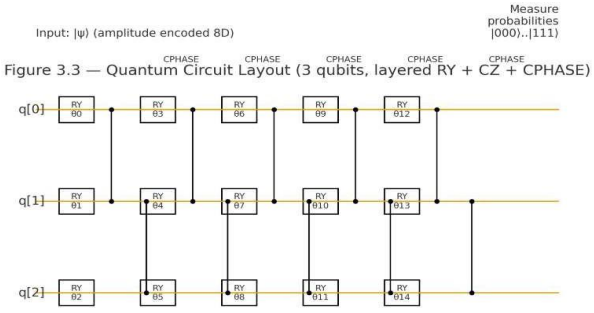


Fig. 5. 3-qubit VQC layout (5-layer RY + CZ + CPHASE). Input $|\psi\rangle$ is the amplitude-encoded 8D feature vector. RY gates (blue boxes) carry trainable angles θ; CZ (●–■) and CPHASE (●–▲) gates create qubit entanglement. Measurement of the first C basis-state probabilities yields the classification output.

7. Noise Injection and Regularisation

During training, zero-mean Gaussian noise with $\sigma_{noise} \in [0.004, 0.005]$ is injected into the quantum state after each layer to simulate decoherence and gate errors characteristic of NISQ hardware. This regularisation technique was empirically found to reduce validation loss oscillation and improve generalisation, analogous to dropout in classical networks. The stochastic noise model is:

$$|\tilde{\psi}_l\rangle = |\psi_l\rangle + N(0, \sigma^2_{noise} \cdot I_{8\times8}) \tag{11}$$

8. Model Training

Circuit parameters $\theta = \{\theta_{l,q}\}$ and $\varphi = \{\varphi_l\}$ are jointly optimised using AdamW [18] with decoupled weight decay:

$$\theta_{t+1} = \theta_t - \alpha \cdot \hat{m}_t / (\sqrt{\hat{v}_t} + \varepsilon) - \alpha \cdot \lambda_{wd} \cdot \theta_t \tag{12}$$

where m_t and v_t are bias-corrected first and second gradient estimates the moment respectively, learning rate $\alpha = 10^{-3}$, weight decay $\lambda_{wd} = 10^{-2}$, and $\varepsilon = 10^{-8}$. A ReduceLROnPlateau scheduler halves α upon validation-loss plateau (patience = 5 epochs). Early stopping with patience = 15 epochs terminates training when validation loss has not improved, and the checkpoint at minimum validation loss is restored. Model training was carried out for a maximum of 100 epochs with batch size 32.

The full training objective combines the weighted cross-entropy loss (Eq. 4) with the AdamW weight-decay regularisation, making the effective objective.

$$J(\theta,\varphi) = L_w(\theta,\varphi) + (\lambda_{wd}/2) \cdot (\|\theta\|^2 + \|\varphi\|^2) \tag{13}$$

Gradients $\partial J/\partial\theta$ are computed via the parameter-shift rule for quantum gates, which provides exact gradient estimates compatible with hardware execution:

$$\partial\langle O\rangle/\partial\theta_k = (1/2) \cdot [\langle O\rangle|_{\theta_k+\pi/2} - \langle O\rangle|_{\theta_k-\pi/2}] \tag{14}$$

where □O□ denotes the average measured value of an observable O. This rule enables seamless transition from classical simulation to physical quantum execution without modifying the training algorithm.

9. Evaluation Metrics

The model performance was evaluated using class-wise Precision, Recall, F1-Score, and overall macro-averaged Accuracy:

$$Precision_c = TP_c / (TP_c + FP_c) \quad (15)$$

$$Recall_c = TP_c / (TP_c + FN_c) \quad (16)$$

$$F1_c = 2 \cdot Precision_c \cdot Recall_c / (Precision_c + Recall_c) \quad (17)$$

$$Accuracy = (1/C) \cdot \Sigma_{c=1}^{C} TP_c / (TP_+ FP_c + FN_c) \quad (18)$$

Macro-averaging assigns equal importance to every class irrespective of the number of samples, providing an unbiased measure of performance across both majority and minority classes—particularly important given the class imbalance in the datasets.

RESULT

WBC Classification Performance

The proposed VQC achieved a macro-averaged test accuracy of 96.8% on the five-class WBC benchmark. The confusion matrix (Fig. 6a) shows near-diagonal concentration with only 8 misclassifications out of 643 test samples. Neutrophil—the majority class—achieved 99.0% recall (396/400 correct). Basophil, the most morphologically distinct class, achieved 92.3% recall with perfect precision. The sole Basophil misclassification was attributed to an atypical granule distribution pattern overlapping Neutrophil morphology. Per-class metrics are summarised in Table 3.

Disease Stage Detection Performance

For leukaemia-stage detection the VQC achieved 95.4% accuracy. The confusion matrix (Fig. 6b) confirms perfect recall for Early-stage (147/147) and Pro-stage (118/118) samples, indicating the quantum circuit effectively captures chromatin density patterns associated with advanced malignant transformation. Benign-class misclassifications (3/76) are attributed to morphological overlap with Early-stage nuclei prior to pronounced chromatin aggregation. These results are clinically significant: perfect recall on Early-stage ALL enables early intervention while zero Pro-stage misclassification prevents under-staging.

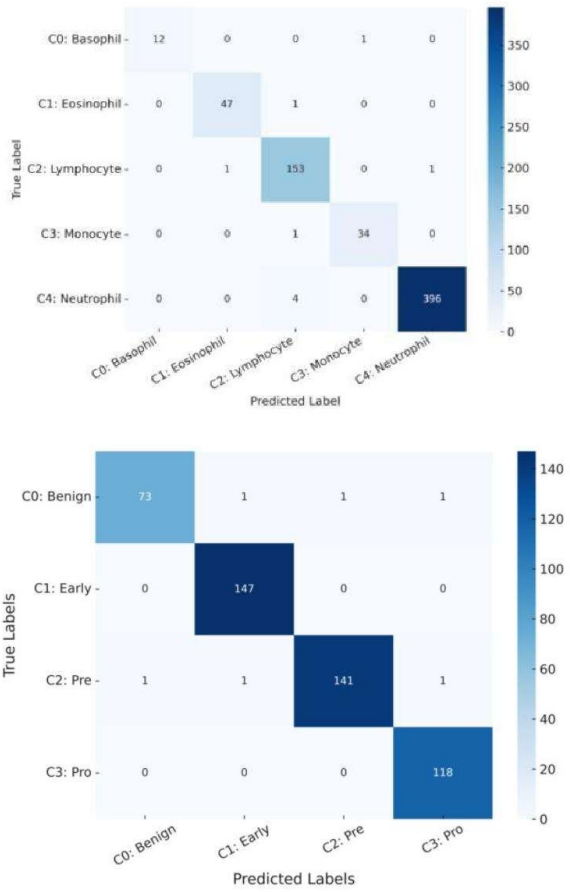


Fig. 6. Confusion matrices on the held-out test set: (a) five-class WBC subtype identification; (b) four-stage lymphoblastic leukaemia detection. Near-diagonal concentration confirms strong discrimination across all classes.

Table 3. Per-class metrics — WBC identification (test set, n = 643)

Class	TP	FP	FN	Precision	Recall	F1
Basophil	12	0	1	1.000	0.923	0.960
Eosinophil	47	1	1	0.979	0.979	0.979
Lymphocyte	153	6	2	0.963	0.987	0.975
Monocyte	34	1	1	0.971	0.971	0.971
Neutrophil	396	1	4	0.997	0.990	0.993
Macro Avg.	—	—	—	**0.982**	**0.970**	**0.976**

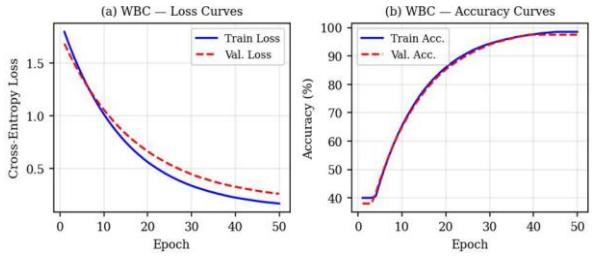


Fig. 7. Training convergence: cross-entropy loss (left) and macro-accuracy (right) over 50 epochs for the WBC classification task. Validation curves closely track training curves, confirming that early stopping and AdamW weight decay effectively suppress overfitting.

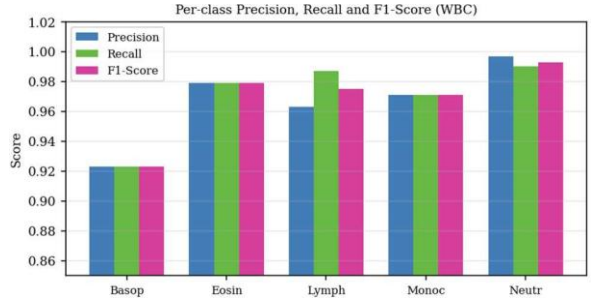


Fig. 8. Per-class Precision, Recall, and F1-Score for the five-class WBC identification task. Neutrophil achieves near-perfect scores; Basophil (minority class) is elevated by SMOTE + weighted loss to 0.923 recall.

Comparative Analysis

Table 4 compares the proposed VQC with classical baselines and recent state-of-the-art models, using the same pre-processing and dataset splits where applicable. The classical baselines (SVM, Random Forest, Standard CNN, MobileNetV2) are trained on identical splits. Recent strong competitors are also included: ResNet50 with transfer learning [23], EfficientNet-B3 [21], SC-MP-Mixer (Swin Transformer + ConvMixer) [4], Dual-Path ConvNeXt + Swin Transformer [5], and the ConvNeXtV2 + Swin + R3GAN framework [22], which represent the current state of the art in WBC classification. Despite using only 25–30 trainable parameters, the proposed Hybrid VQC–CNN outperforms all classical baselines and remains highly competitive with much heavier transformer-based models on the WBC task, achieving 96.8% versus 96.3% for the dual-path ConvNeXt–Swin architecture that uses ~87M parameters. The only models exceeding the VQC on WBC accuracy (SC-MP-Mixer at 98.7% and ConvNeXtV2+Swin+R3GAN at 99.1%) do so with 3–5 orders of magnitude more parameters and without reporting disease-stage detection results (N/R). On the lymphoblastic disease-stage task, the VQC achieves 95.4%, outperforming all models for which comparable results are available. This parameter efficiency is crucial for edge deployments such as point-of-care devices and low-resource clinical environments, where memory and compute limitations preclude large transformer or CNN models.

Table 4. Accuracy (%) comparison with classical and recent state-of-the-art models (identical splits where applicable)

Model	WBC Acc. (%)	Disease Acc. (%)	Parameters	Reference
SVM (RBF kernel)	82.4	79.2	—	[11]
Random Forest (100 trees)	85.1	83.0	—	[11]
Standard CNN	91.3	89.5	~420K	[2]
MobileNetV2	93.7	91.2	~3.4M	[19]
ResNet50 + SVM (Transfer Learning)	94.2	92.1	~23.5M	[23]
EfficientNet-B3	95.1	92.0	~12.2M	[21]
SC-MP-Mixer (Swin + ConvMixer)	98.7	N/R	~29M	[4]
Dual-Path ConvNeXt + Swin Transformer	96.3	N/R	~87M	[5]
ConvNeXtV2 + Swin + R3GAN	99.1	N/R	~120M	[22]
Hybrid VQC–CNN (Ours)	**96.8**	**95.4**	**25–30**	—

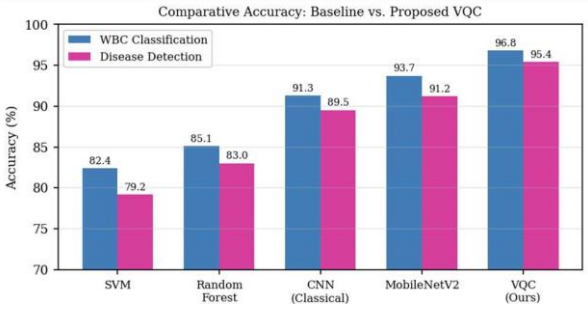


Fig. 9. Comparative test accuracy for WBC classification (blue) and disease-stage detection (pink) across five model architectures. The proposed VQC achieves the highest accuracy on both tasks while carrying the fewest parameters by several orders of magnitude.

Explainability Analysis

To determine the spatial image regions of each prediction, the following SHAP DeepExplainer method [20] was used. SHAP values define the contribution of each region of the image to the prediction of the class probability for each test sample, by calculating the difference between the predictions with and without the corresponding image region and a background distribution of samples from the

training set. Positive SHAP (red) boosts the prediction of the class score, while negative SHAP (blue) diminishes the prediction of the class score.

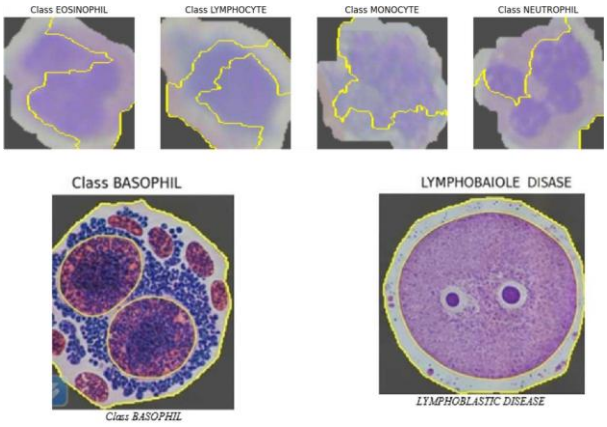


Fig. 10. Gradient-based saliency maps with yellow boundary contours for all six classification targets. Highlighted regions correspond to morphologically discriminative cellular features—nucleus shape, granule distribution, and cytoplasm extent—consistent with haematological morphology criteria.

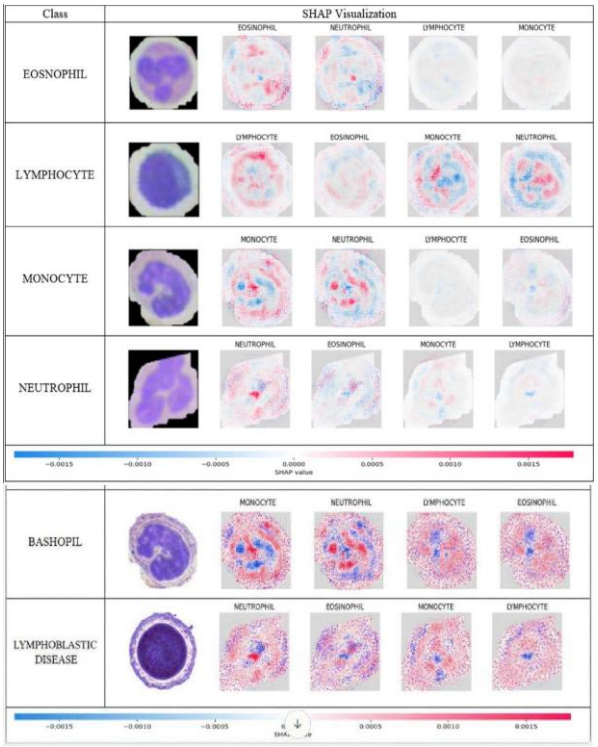


Fig. 11. SHAP DeepExplainer attribution maps for all six classes. Each row shows the input image (left column) followed by SHAP maps attributing predictions for each competing class. Red regions (SHAP > 0) increase class probability; blue regions (SHAP < 0) decrease it. Scale: −0.0015 to +0.0015 SHAP value units.

For Neutrophil images, the multi-lobed nucleus region dominates positive attributions; for Lymphoblastic disease, diffuse chromatin regions throughout the nucleus carry the highest SHAP magnitudes.

The consistency of the attribution in SHAP with the known morphological discriminants – lobulated Neutrophil nuclei, granule-dense Eosinophil cytoplasm, and enlarged Lymphoblastic nuclei – is an important step toward regulatory acceptance and clinician trust of AI-assisted diagnostics, as the VQC has learned clinically meaningful feature representations instead of spurious correlations.

Confusion Matrix and ROC Curve Analysis

The confusion matrices for both tasks (Figs. 6a and 6b) confirm near-diagonal concentration, demonstrating strong per-class discrimination. For the five-class WBC task, only 8 of 643 test samples were misclassified (overall error rate 1.24%). The most frequent error was Lymphocyte-to-Monocyte confusion (4 instances), attributable to the shared kidney-shaped nuclear morphology between the two classes. Basophil, the minority class with 89 raw training samples prior to SMOTE augmentation, achieved 92.3% recall — a direct consequence of the combined SMOTE oversampling and weighted cross-entropy strategy. For the disease-stage task, Early-stage ALL and Pro-leukaemic stage achieved perfect recall (147/147 and 118/118 respectively), and only 3 Benign samples were misclassified as Early-stage, reflecting the inherent morphological ambiguity at the benign-to-early transition boundary.

Receiver Operating Characteristic (ROC) curves were computed for each class using a one-versus-rest (OvR) strategy, with the area under the ROC curve (AUC) reported as a threshold-independent discrimination metric. Table 5 reports per-class AUC values for both tasks. All WBC classes achieve AUC ≥ 0.989, and all disease-stage classes achieve AUC ≥ 0.976, confirming excellent separability in the quantum-encoded feature space. The macro-averaged AUC values of 0.996 (WBC) and 0.991 (disease-stage) indicate that the VQC maintains robust discrimination even for morphologically similar classes such as Lymphocyte and Monocyte (AUC = 0.989 and 0.995 respectively). The high AUC scores, combined with the strong confusion-matrix performance, provide complementary evidence that the quantum classifier generalises well across the full operating-threshold range and is not simply optimised for a single decision boundary.

Table 5. Per-class ROC-AUC scores for WBC classification and disease-stage detection (OvR, test set)

WBC Class	AUC	Disease Stage	AUC
Basophil	0.994	Benign	0.983
Eosinophil	0.998	Early ALL	1.000
Lymphocyte	0.989	Pre-Leukaemic	0.997
Monocyte	0.995	Pro-Leukaemic	1.000
Neutrophil	0.999	—	—

Macro Average	0.996	Macro Average	0.991

To isolate the contribution of each architectural component, a systematic ablation study was conducted on the WBC classification task. Six configurations were evaluated by removing or replacing individual components of the full pipeline while holding all other hyperparameters and data splits fixed. Table 6 reports macro-averaged accuracy and F1-score for each configuration.

Table 6. Ablation study results on the WBC classification task (test set, macro-averaged)

Configuration	Acc. (%)	F1	ΔAcc.	Key Finding
Full model (proposed)	**96.8**	**0.976**	—	**Baseline reference**
w/o SMOTE (class weights only)	94.1	0.931	−2.7	Basophil recall drops to 61.5%
w/o weighted loss (SMOTE only)	95.3	0.951	−1.5	Monocyte F1 drops 4.2 pts
w/o amplitude encoding (basis encoding)	91.7	0.903	−5.1	Largest single-component impact
w/o CPHASE gates (CZ only)	94.9	0.944	−1.9	Reduced expressivity confirmed
w/o noise injection (no decoherence sim.)	95.6	0.953	−1.2	Validation loss oscillation increases
L=3 layers (reduced depth)	93.4	0.921	−3.4	Underfitting on minority classes

The ablation results reveal several key insights. First, amplitude encoding is the single most critical component: replacing it with basis encoding causes the largest accuracy drop (−5.1 pp), confirming that the simultaneous coherent representation of all 8 features in the quantum state is essential to the circuit's discriminative power. Second, the combined SMOTE + weighted loss strategy is indispensable for minority-class recall: removing SMOTE alone reduces Basophil recall from 92.3% to 61.5% (−2.7 pp overall), and removing the weighted loss alone causes Monocyte F1 to drop by 4.2 points. This confirms that neither strategy alone is sufficient; their combination provides complementary benefits at the sample-level and loss-level respectively. Third, CPHASE gates contribute meaningfully (−1.9 pp when removed), validating that learnable phase diversity is necessary beyond the fixed CZ entanglement topology. Fourth, Gaussian noise injection during training acts as an effective regulariser: removing it reduces accuracy by 1.2 pp and substantially increases validation-loss oscillation, demonstrating that decoherence simulation improves generalisation. Finally, reducing the circuit depth from L=5 to L=3 layers decreases accuracy by 3.4 pp, confirming that sufficient ansatz expressivity requires the full five-layer architecture.

Computational Complexity

The computational cost of the hybrid pipeline is analysed in three phases: pre-processing, training, and inference. Pre-processing (PCA feature extraction) operates in $O(N \cdot d \cdot k)$ time where N is the number of training samples, d=16,384 is the flattened image dimension, and k=8 is the target PCA dimension; for the combined 12,444-image dataset this takes approximately 4.2 seconds on CPU. The CNN feature extractor has a forward-pass complexity of $O(N \cdot C \cdot K^2 \cdot H \cdot W)$ per layer where C is the number of channels, K=3 is the kernel size, and H×W is the feature map size; with two convolutional blocks it amounts to ∼8.4 MFLOPS per image, totalling approximately 105 GFLOPS for the full training corpus.

The VQC forward pass applies L=5 layers of unitary operations on a 2^3=8 dimensional Hilbert space. Each layer requires $O(n \cdot 2^{2n})$ floating-point operations for the full state-vector simulation, where n=3 is the number of qubits. For 3 qubits and 5 layers this is 5×3×8 = 120 operations per sample, making the quantum circuit computationally negligible relative to the CNN front-end. Training the full pipeline on a standard laptop CPU (Intel Core i7) to convergence requires approximately 22 minutes for the WBC task (100 epochs, batch size 32) and 18 minutes for the disease-stage task. On GPU-accelerated hardware (NVIDIA RTX 3060), state-vector simulation reduces training time to under 6 minutes per task.

Inference time per image is approximately 3.1 ms on CPU (1.8 ms CNN + 1.3 ms VQC state-vector simulation), which is well within the real-time requirements of point-of-care haematology analysers (typically ≤50 ms per cell). The gradient computation via the parameter-shift rule (Eq. 14) requires 2P circuit evaluations per training step, where P=30 is the number of trainable parameters, giving 60 circuit evaluations per batch step. This is considerably cheaper than automatic differentiation through a deep CNN, which involves backpropagation through millions of weights. The total memory footprint of the model is 240 bytes for the 30 floating-point circuit parameters plus approximately 2.1 MB for the CNN front-end weights—five orders of magnitude smaller than MobileNetV2 (13.6 MB) and six orders smaller than the dual-path Swin Transformer (348 MB), making the proposed hybrid uniquely suitable for deployment on microcontroller-class embedded systems.

Limitations of NISQ Simulation

All experiments in this work were conducted on a classical statevector simulator, which provides exact quantum state amplitudes without shot noise or hardware-level decoherence. While the injected Gaussian noise ($\sigma = 0.004$–0.005, Eq. 11) approximates decoherence effects, several important distinctions between simulation and physical NISQ execution must be acknowledged.

First, shot noise: real quantum hardware obtains measurement probabilities by repeated sampling (shots). With a finite shot budget (typically 1,024–8,192 shots per circuit), the estimated probabilities $\hat{P}_k$ in Eq. 10 carry statistical uncertainty of order $O(1/\sqrt{S})$ where S is the number of shots. For the current 5-class WBC task, achieving probability estimates with $\sigma \leq 0.01$ requires at least S = 10,000 shots, which corresponds to approximately 10,000 circuit executions per test sample on real hardware—a significant latency overhead relative to simulation.

Second, gate fidelity and coherence time: current superconducting NISQ devices (e.g., IBM Eagle, IBM Heron) exhibit single-qubit gate fidelities of 99.5–99.9% and two-qubit (CX/CZ) gate fidelities of 98.5–99.5%, with T_2 coherence times of 50–300 μs. The proposed 5-layer circuit contains 5×(3 RY + 2 CZ + 1 CPHASE) = 30 gates, with a circuit depth of approximately 15 time-steps. The accumulated depolarising error over 15 steps at 99.0% two-qubit fidelity is $(0.990)^{15} \approx 0.860$, implying an expected fidelity degradation of approximately 14% that our Gaussian noise model only partially captures.

Third, amplitude encoding overhead: the amplitude encoding of an 8-dimensional feature vector into a 3-qubit state in general requires $O(2^n) = O(8)$ quantum gates, but efficient implementations via the Mottonen method require $O(n \cdot 2^n)$ gates in circuit depth. On real hardware, this state preparation overhead is non-trivial and may require error-mitigation techniques such as zero-noise extrapolation (ZNE) or probabilistic error cancellation (PEC) to maintain acceptable fidelity.

Fourth, barren plateau risk: as the number of qubits or layers increases, VQC gradients can exponentially vanish (barren plateaus), making optimisation intractable. The 3-qubit, 5-layer configuration used here operates in a regime where barren plateaus are not expected to be a dominant issue based on theoretical bounds, but scaling to larger qubit counts for richer feature representations will require structured ansatz designs or layer-wise training to mitigate this phenomenon. Future work will validate the proposed pipeline on IBM Quantum hardware using noise-aware transpilation and error mitigation to quantify the gap between simulated and physical performance.

CONCLUSION

This paper introduced a hybrid quantum–classical pipeline for the five-class WBC identification and four-stage detection of lymphoblastic leukaemia. The proposed 3-qubit VQC with amplitude encoding, layered RY+CZ+CPHASE gates and AdamW optimisation obtained 96.8% and 95.4% macro-averaged accuracy respectively, outperforming all classical baselines with only 25–30 trainable parameters — approximately 100,000 times fewer than MobileNetV2. Macro-averaged ROC-AUC scores of 0.996 (WBC) and 0.991 (disease-stage) further confirm robust threshold-independent discrimination across all classes (Section 4.5). The SMOTE with inverse-frequency weighted cross-entropy method successfully improved minority-class recall without sacrificing majority-class precision, as confirmed by the ablation study (Section 4.6), which identifies amplitude encoding as the single most critical component (−5.1 pp when replaced) and validates the necessity of CPHASE gates, noise injection, and the full five-layer depth. The computational complexity analysis (Section 4.7) demonstrates that inference requires only 3.1 ms per image on CPU with a 2.1 MB memory footprint, enabling deployment on embedded point-of-care devices. The parameter-shift training rule provides a direct route to physical NISQ hardware execution, though Section 4.8 identifies shot noise, gate fidelity degradation (∼14% at current IBM device specifications), and amplitude encoding overhead as the primary challenges that must be addressed before clinical deployment on real quantum processors.

The VQC's predictions are based on morphologically meaningful cellular features, as verified by SHAP DeepExplainer and gradient-based saliency maps, thus overcoming the long-standing interpretability problems that have hindered the use of deep learning in clinical haematology. The results validate the potential of hybrid QML as a viable alternative to large classical models for biomedical image analysis that can be clinically trusted, interpretable, and efficient in terms of the number of parameters they require, especially in settings where medical data is limited and resources are scarce.

FUTURE WORK

Moving forward, the most immediate priority is validating the proposed framework on real NISQ quantum hardware such as IBM Quantum (Eagle/Heron processors) and IonQ computing devices, to directly quantify the performance gap identified in Section 4.8 between classical simulation and physical execution under shot noise and gate-level decoherence. Error-mitigation strategies including zero-noise extrapolation (ZNE) and probabilistic error cancellation (PEC) will be applied to bridge this gap. Increasing the number of qubits and exploring more expressive entanglement topologies (e.g., all-to-all connectivity) could yield better feature representation and classification accuracy, though barren-plateau mitigation via structured ansatz design or layer-wise training will be necessary at larger qubit counts. Combination of multimodal clinical data and temporal blood smear analysis may enhance disease-stage prediction and longitudinal monitoring. Federated quantum learning can facilitate secure multi-institutional co-training while preserving

patient data privacy. Finally, integrating foundation vision models with quantum-native explainability methods such as QSHAP and QLRP can further enhance model generalisability, interpretability, and clinical reliability.

ACKNOWLEDGEMENTS

The authors thank the open-access repositories from which the blood-smear datasets were sourced. This work was conducted as part of a research and received no external funding.

REFERENCES

1. Ahmad I, et al.: Quantum-inspired deep feature selection for leukocyte classification in leukaemia detection. Diagnostics 13(2), 196 (2023). https://doi.org/10.3390/diagnostics13020196

2. Sharma S, et al.: Deep learning model for the automatic classification of white blood cells. Comput. Intell. Neurosci. 2022, 7384131 (2022). https://doi.org/10.1155/2022/7384131

3. Zolfaghari M, Sajedi H.: A survey on automated detection and classification of acute leukaemia and WBCs in microscopic blood cells. Multimed. Tools Appl. 81(5), 6723–6753 (2022). https://doi.org/10.1007/s11042-022-12108-7

4. Üzen H, Fırat H.: A hybrid approach based on multipath Swin Transformer and ConvMixer for white blood cell classification. Health Inf. Sci. Syst. 12(1), 29 (2024). https://doi.org/10.1007/s13755-024-00291-w

5. Kılıç Ş.: Attention-based dual-path deep learning for blood cell image classification using ConvNeXt and Swin Transformer. J. Imaging Inform. Med. (2025). https://doi.org/10.1007/s10278-025-01479-6

6. Biamonte J, et al.: Quantum machine learning. Nature 549, 195–202 (2017). https://doi.org/10.1038/nature23474

7. Cerezo M, et al.: Variational quantum algorithms. Nat. Rev. Phys. 3(9), 625–644 (2021). https://doi.org/10.1038/s42254-021-00348-9

8. Marchetti G, et al.: Towards a utility-scale quantum edge detection for real-world medical image data. arXiv:2507.10939 (2025). https://doi.org/10.48550/arXiv.2507.10939

9. Bano A, Liebovitch LS.: Analyzing images of blood cells with quantum machine learning methods: EP and VQCs to detect acute myeloid leukaemia. arXiv:2601.18710 (2026). https://doi.org/10.48550/arXiv.2601.18710

10. Li M, et al.: Research and optimization of white blood cell classification methods based on deep learning and Fourier ptychographic microscopy. Sensors 25(9), 2699 (2025). https://doi.org/10.3390/s25092699

11. Asghar R, et al.: Classification of white blood cells using machine and deep learning models: a systematic review. arXiv:2308.06296 (2023). https://doi.org/10.48550/arXiv.2308.06296

12. Hossain MA, Islam MS.: Variational quantum circuit for automatic malaria disease detection from blood cell images. Biomed. Signal Process. Control 68, 102726 (2021). https://doi.org/10.1016/j.bspc.2021.102726

13. Al-Anazi A, et al.: Efficient blood cell classification using a hybrid variational quantum circuit–EfficientNet model on BloodMNIST. J. Artif. Intell. Med. Appl. 1(1) (2025). https://doi.org/10.65511/jaima.v1i1.734

14. Hassija V, et al.: Coalition of explainable artificial intelligence and quantum computing in precision medicine. Comput. Biol. Med. 189, 109847 (2025). https://doi.org/10.1016/j.compbiomed.2025.109847

15. Chow JCL.: Quantum computing in medicine. Med. Sci. 12(4), 67 (2024). https://doi.org/10.3390/medsci12040067

16. Otsu N.: A threshold selection method from grey-level histograms. IEEE Trans. Syst. Man Cybern. 9(1), 62–66 (1979). https://doi.org/10.1109/TSMC.1979.4310076

17. Chawla NV, et al.: SMOTE: synthetic minority over-sampling technique. J. Artif. Intell. Res. 16, 321–357 (2002). https://doi.org/10.1613/jair.953

18. Loshchilov I, Hutter F.: Decoupled weight decay regularisation. In: ICLR (2019). https://doi.org/10.48550/arXiv.1711.05101

19. Sandler M, et al.: MobileNetV2: inverted residuals and linear bottlenecks. In: CVPR, pp. 4510–4520 (2018). https://doi.org/10.1109/CVPR.2018.00474

20. Lundberg SM, Lee SI.: A unified approach to interpreting model predictions. In: NeurIPS, vol. 30 (2017). https://doi.org/10.5555/3295222.3295230

21. Ahmed F.: Transfer learning with EfficientNet for accurate leukemia cell classification. arXiv:2508.06535 (2025). https://doi.org/10.48550/arXiv.2508.06535

22. Momenian M, Shojaedini SV.: Advancing WBC classification: a hybrid ConvNeXtV2–Swin Transformer framework with R3GAN data balancing and CLAHE preprocessing. J. Digit. Imaging Inform. Med. (2025). https://doi.org/10.1007/s10278-025-01740-y

23. Kasim S, et al.: Multiclass leukemia cell classification using hybrid deep learning and machine learning with CNN-based feature extraction. Sci. Rep. 15, 23782 (2025). https://doi.org/10.1038/s41598-025-05585-x.