TELKOMNIKA Telecommunication Computing Electronics and Control
Vol. 24, No. 1, February 2026, pp. 359~370
ISSN: 1693-6930, DOI: 10.12928/TELKOMNIKA.v24i1.27269

Optimizing blood cell classification: evaluating feature dimensionality and validation strategies

Ruaa H. Ali Al-Mallah
Technical Engineering College for Computer and Artificial Intelligence, Northern Technical University, Mosul, Iraq

Article Info

Article history:

Received Jun 6, 2025
Revised Nov 19, 2025
Accepted Dec 8, 2025

Keywords:

Artificial intelligence
Blood cell classification
Blood cell image
Cross validation
InceptionV3
Optimizing

ABSTRACT

Manual blood cell classification is time consuming and may lead to inconsistent results. This study aims to assist pathologists in diagnosing hematological disorders using machine learning (ML) techniques for automated classification of blood cells in multi-color test images, distinguishing red blood cells (RBCs) and white blood cells (WBCs). Features were extracted using the InceptionV3 network, and several ML models were evaluated for classifying blood cells into eight categories. Two validation strategies: a 66%–34% train–test split and 20-fold cross-validation were applied. The effect of dimensionality reduction through principal component analysis (PCA) was also examined, reducing the feature space from 2,048 to 100 components. Among all models, support vector machine (SVM) achieved highest performance, with 93.4% accuracy and an area under the curve (AUC) of 0.996 without PCA, and 90.1% accuracy with an AUC of 0.991 after PCA. Although PCA slightly reduced accuracy, it improved computational efficiency. Overall, SVM provided the most accurate, stable, and generalizable classification results for automated blood cell analysis.

Corresponding Author:

Ruaa H. Ali Al-Mallah
Technical Engineering College for Computer and Artificial Intelligence, Northern Technical University
Mosul, Iraq
Email: ruaa_almallah@ntu.edu.iq

1. INTRODUCTION

Medical image analysis is considered as a critical component in the quickly diagnosis of blood cell abnormalities. The abnormality of the blood cells or potential disorders or diseases can be diagnosed from variations in contain shape, color, or size [1]. However, morphological differences detection in red blood cells (RBCs) still challenging because they are very similar in size and appearance [2]. Raw images of RBCs, white blood cells (WBCs), and platelets often contain widespread morphological features which require advanced computational techniques to ensure machine learning effective analysis [3].

Recently, different machine learning (ML) algorithms, especially those based on image feature extraction, have been used in automated blood cell classification. Accurate classification of blood cells is essential for identifying various hematological diseases efficiently and reliably [4]. Modern developments in artificial intelligence (AI) and deep learning (DL) assisted in enable automated, objective, and high-precision assessments of medical images [5]. Among different approaches convolutional neural networks (CNNs), especially architectures like InceptionV3 pre-trained on ImageNet have been proven effective in extracting features for blood cell classification tasks [6], [7]. Principal component analysis (PCA) has been used in this study to reduce redundant blood cell features, enhancing efficiency and model accuracy [8]. ML algorithms have been used in order to achieve accurate blood cell identification and classification [9].

Researchers are increasingly applying artificial AI in medical imaging to analyze tumors, blood vessels, and cells [10]. Techniques such as computed tomography (CT), magnetic resonance imaging (MRI), and positron emission tomography (PET) provide detailed visual information but generate large, complex datasets that require advanced AI models for accurate analysis and interpretation [11]-[14].

In 2020, Alzubaidi *et al.* [15] introduced an approach to classify three types of RBCs using three DL models consists of parallel and traditional convolutional layers. The best results of their method have been achieved using transfer learning within the same domain, reaching 99.54% accuracy. Similar performance was achieved when a support vector machine (SVM) was used. In the same year, Singh *et al.* [16] developed a custom CNN model to classify WBC types. The network architecture consists of two convolutional layers, two pooling layers, and two fully connected layers and the accuracy was 86%. In 2021, Navya *et al.* [17] proposed a model to segment and classify blood cells into WBCs and RBCs. Gray level co-occurrence matrix (GLCM) was used as a feature extraction. Several classifiers including Naive Bayes (NB), k-nearest neighbors (KNN), decision tree (DT), random forest (RF), logistic regression (LR), artificial neural networks (ANNs), and SVM have been applied. The highest accuracy of 97% achieved when used LR.

Elhassan *et al.* [18] introduced WBC classification approach consisting of two steps. The first one, "geometric transformation-deep convolutional autoencoder (GT-DCAE) WBC augmentation model," extracted an atypical WBC feature using deep convolutional autoencoder and geometric transformations, while the second step, "two-stage atypical WBC classification model," used a DCAE-CNN hybrid as a final classification. The accuracy obtained was 97%. In 2023, Ali *et al.* [19] uses vision transformers (ViTs) and CNNs to study the classification of WBCs. Using two datasets: peripheral blood cell (PBC) (high-quality photos) and blood cell count and detection (BCCD) (noisy, unbalanced images) the authors compare Google ViT and pre-trained ImageNet CNNs for the classification of four WBC kinds (neutrophil, eosinophil, lymphocyte, and monocyte). The limitation of this research ViT is computationally expensive, it beat CNNs, reaching 100% accuracy on PBC and 88.36% accuracy on BCCD. In 2023, Heni *et al.* [20] suggested combining fuzzy C-means and K-means to improves picture segmentation by introducing EK-means segmentation. Visual geometry group 19 (VGG19) model used to optimize for WBC classification, reaches an accuracy of 96.24%. Validated using 12,444 pictures of blood cells (monocytes, lymphocytes, eosinophils, and neutrophils), the model performs better than Inception, ResNet, and recurrent neural network (RNN), especially when data augmentation techniques are used.

In 2025, a new DL technique called Swin–spatial pyramid down(SPD)–Wasserstein distance loss (WDLoss) you only look once (SSW-YOLO) was created to improve blood cell recognition by tackling issues with low-resolution images and small target detection [21]. It incorporates a streamlined c2f module, Swin Transformers for multi-scale attention, spatial pyramid down sampling convolution (SPD-Conv) layers for enhanced feature extraction, and WDLoss for enhanced localization accuracy. These enhancements speed up diagnosis, lower human error, and improve detection accuracy. SSW-YOLO outperforms current techniques with a mean average precision (mAP) of 94.0% when tested on the BCCD dataset.

This work suggests minimizing feature extraction at low cost and reducing feature dimensionality from DL model by PCA technique and used it to evaluate and compare different ML algorithms in order to get a suitable method for applications. Additionally, the effects of different blood cell types on the performance of classification models were analyzed and optimized.

2. METHOD AND MATERIALS

The study presents a blood cell classification system that integrates automatic feature extraction based on DL using InceptionV3 with PCA for dimensionality reduction. Two distinct classification pipelines are used one with PCA and one without it. Both approaches use logistic LR, KNN, neural network (NN), and SVM in the classification task. The dataset is split into 66% for training and 34% for testing, followed by 20-fold cross-validation to assess model robustness. Figure 1 illustrates the flowchart of the proposed system.

2.1. Image acquisition

The dataset used in this study was performed by collecting microscopic PBC images from the CellaVision DM96 analyzer at Barcelona Hospital Clinic. The data set consist of 11,092 images across eight blood cell classes: neutrophils, eosinophils, basophils, lymphocytes, monocytes, immature granulocytes (IGs), erythroblasts, and platelets, each image is (360×363 pixels, JPG) [22]. Figure 2 presents a violin plot of cell size distributions, showing that platelets have the most uniform sizes, while monocytes and IGs exhibit greater variability. The dataset is moderately imbalanced, with fewer basophils (200 images) and IGs (150 images) compared to neutrophils (4,000 images) and lymphocytes (3,200 images). No class weighting or sampling techniques were applied; the models were trained on the original dataset to reflect the natural distribution of cell types in clinical settings.

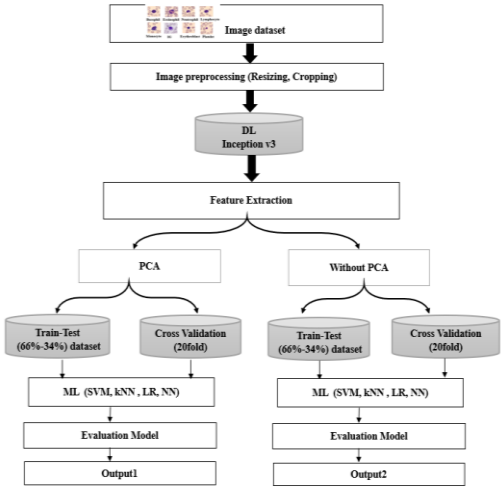


Figure 1. Flowchart of the proposed blood cell classification system

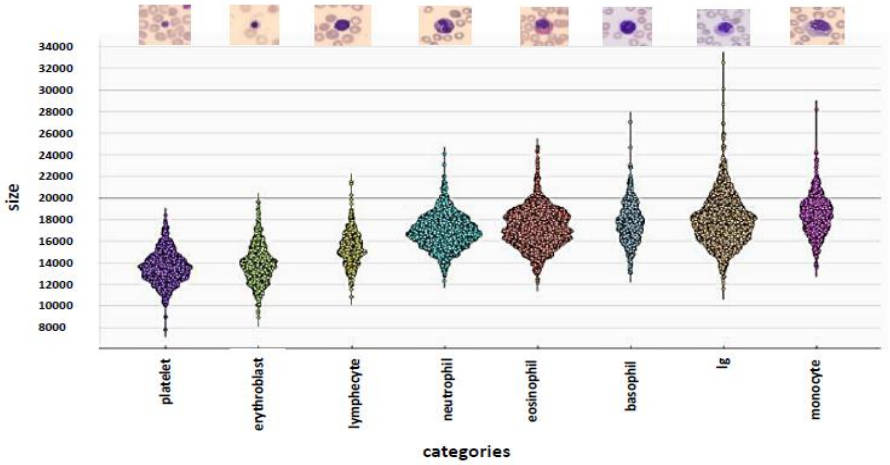


Figure 2. Violin plot showing the distribution of blood cell sizes across eight classes

2.2. InceptionV3 classification

InceptionV3 is a deep CNN developed by Google Research to enhance image classification, detection, and segmentation performance [23]-[25]. It consists of 42 layers and is pretrained on the ImageNet dataset containing 1,000 classes [23]. The architecture includes three main inception modules (A, B, and C) composed of parallel convolutions with different kernel sizes (1×1, 3×3, and 5×5) [24]. An auxiliary classifier is incorporated to improve regularization, while factorized and asymmetric convolutions reduce computational cost and parameters [25], [26]. Figure 3 illustrates the InceptionV3 architecture.

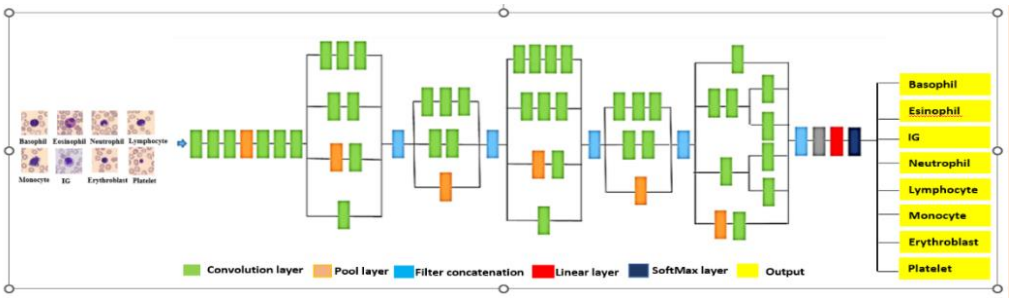


Figure 3. Architecture of the InceptionV3 CNN

2.3. Classification algorithm in ML

ML involves techniques used in different tasks such as, data analysis, pattern recognition, and classification the choice of algorithms and features plays critical roles in the classification effective. This study focuses on four widely used classifiers: SVM, KNN, LR, and NN.

- LR: LR is used when the response variable is categorical, the probability of an event is estimated by fitting data to a logistic (sigmoid) function [27]. The predicted probability is given in (1):

$$(x) = \frac{e^{(b0+b_1^* X)}}{1+e^{(b0+b_1^* X)}} \quad (1)$$

where, $p(x)$: predicted output, b_0: intercept term, b_1^* coefficient for the single input value (x).

- NN: it uses a multilayer perceptron (MLP) which is a feedforward neural network that maps inputs through fully connected hidden layers to outputs [28].
- SVM: is a supervised ML technique used for classification in high-dimensional feature spaces [29]. It separates data by constructing a hyperplane that maximizes the margin between classes [30]. For a linear SVM, the decision function is defined using (2):

$$f(x) = w.x + b \quad (2)$$

Where, x: input feature, w: weight vector orthogonal to the hyperplane, and b is the bias.

- KNN: is a supervised learning algorithm that classifies objects based on the closest data points in a multi-dimensional feature space [31]. During training, it stores the feature vectors and labels. For classification, it computes the distance between a test point and all training points, identifies the KNNs, and assigns the class most common among them. The Euclidean distance is commonly used as in (3):

$$d_{Euclidean}(x, y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \quad (3)$$

2.3.1 Parameters of ML models

The hyperparameters are important as an external configuration for all ML models. The hyperparameters were optimized for classification before training. In this study, the hyperparameters is used as illustrated in Table 1.

Table 1. Optimized hyperparameters for ML models used in blood cell classification

ML algorithm	Parameters
SVM	Polynomial kernel (degree 3), C = 1.0, ε = 0.1, tolerance = 0.0011, iterations = 2000
KNN	Neighbors = 5, Euclidean distance, distance-based weighting
NN	Hidden layers = 100, rectified linear unit (ReLU), stochastic gradient descent (SGD) solver, α = 0.04, iterations = 2000
LR	L2 (ridge) regularization, C = 6

2.4. Principal component analysis

PCA is a linear dimensionality-reduction method which converts high-dimensional data into a set of uncorrelated principal components (PCs), each of them represents a direction of maximum variance in the data [8]. This technique is particularly effective for image-based feature datasets, where different features may be highly correlated due to shared structural or morphological characteristics.

In this study, the Inception-V3 model was used to extract 2,048 deep features from blood cell images. these extracted features exhibited substantial redundancy, due tothe fact that blood cells share common morphological properties such as size, shape, area, and internal texture. To reduce this redundancy and enhance computational efficiency, PCA was applied to transform the original feature space into a lower-dimensional representation. The resulting dataset contains 100 principal components. Before applying PCA and in order to ensure equal contribution of each feature, all features' values were normalized using (4).

$$Z = \frac{X-\mu}{\sigma} \quad (4)$$

where, X refers to the original feature vector, μ represents the mean, and σ is the standard deviation of each feature.

Then, the relationships between features were quantified using the covariance matrix, as in (5).

$$cov(x1, x2) = \frac{\sum_{i=1}^{n}(x1_i - \grave{x1})(x2_i - \grave{x2})}{n} \quad (5)$$

where, n represents the total number of samples, $(x1)\grave{}$ and $(x2)\grave{}$ are the mean values.

The eigenvalues (λ) and eigenvectors (X) of the covariance matrix (A) were computed by solving (6).

$$AX = \lambda X \quad (6)$$

where, A: elements of features as square matrix

The eigenvectors corresponding to the largest eigenvalues form the principal components. These PCs are orthogonal and ranked according to the amount of variance they capture. After eigen-decomposition, the first 100 principal components were selected. These components retained approximately 76% of the total variance. The retained components exhibited relatively uniform mean values as a result, PCA improves both the computational efficiency and classification performance of the downstream machine-learning models used for blood cell recognition [32].

2.5. K-fold cross validation method

In this technique the dataset is divided into n folds, used once for validation while the others used in model training. This process diminishes overfitting, improves generalization, and provides a stable performance estimate by averaging results across folds [33]. In this study 20-fold was used for data validation.

2.6. Performance evaluation metrics

The performance of the proposed architecture for blood cell classification was assessed using the following metrics:

- Accuracy: measures the overall proportion of correct predictions. While useful, it may be misleading for imbalanced datasets.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (7)$$

- Recall: indicates the proportion of actual positives correctly identified. High recall is critical when missing positive cases is costly.

$$Recall = \frac{TP}{TP+FN} \quad (8)$$

- Precision: measures the proportion of positive predictions that are correct, important when false positives are costly.

$$Precision = \frac{TP}{TP+FN} \quad (9)$$

- F1-score: harmonic mean of precision and recall, useful for imbalanced datasets. Higher values indicate a better balance between precision and recall.

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (10)$$

- AUC-ROC: the area under the curve (AUC) of the receiver operating characteristic (ROC) evaluates the ability of the model to distinguish between positive and negative classes. AUC values range from 0.5 which represents random performance to 1 indicating to perfect classification, while the ROC curve plots true positive rate against false positive rate across thresholds [31], [33].

3. RESULTS AND DISCUSSION

In this study, ML techniques with PCA were applied to classify BCs [16]. The dataset used consists of 11,092 images across eight classes. Experiments were conducted using Python on an open-source platform with an NVIDIA GeForce RTX 4050 GPU and an Intel Core i7-13620H processor (2.40 GHz).

3.1. Effect of dimensionality reduction

Two validation strategies have been used in evaluating the classification: 66%/34% train–test split and 20-fold cross-validation with two feature settings to assess their impact on model performance, full 2048 features and 100 features reduced by PCA.

3.1.1. 66-34% train–test split

Using the 66%–34% train–test split, SVM achieved the highest performance without PCA, with 92.6% accuracy and 0.995 AUC, as shown in Figure 4 and Table 2, demonstrating its effectiveness for high-dimensional, multi-class data. NN and LR performed similarly, with accuracy of 92.2% and 92.4%, and AUC of 0.994. KNN showed the lowest performance (73.4% accuracy, 0.935 AUC), indicating its limitations with high-dimensional features [17]. After applying PCA, SVM remained the best (89.3% accuracy, 0.990 AUC), followed by NN (88.0% accuracy, 0.987 AUC) and LR (87.8% accuracy, 0.984 AUC), while KNN showed a slight improvement (72.9% accuracy, 0.920 AUC).

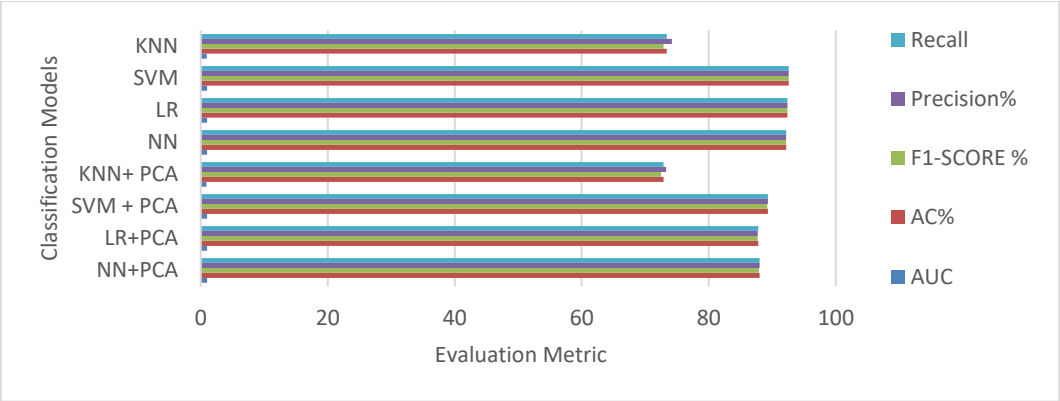


Figure 4. Comparison of four ML models with and without PCA using (66/34% train – test split)

Table 2. Performance of ML models with and without PCA using 66-34% train – test split

Models	AUC	AC%	F1-score%	Precision%	Recall%
KNN	0.935	73.4	72.9	74.2	73.4
SVM	0.995	92.6	92.6	92.6	92.6
LR	0.994	92.4	92.4	92.4	92.4
NN	0.994	92.2	92.2	92.2	92.2
KNN+ PCA	0.92	72.9	72.5	73.3	72.9
LR+PCA	0.984	87.8	87.7	87.7	87.8
SVM+PCA	0.99	89.3	89.2	89.3	89.3
NN+PCA	0.987	88	87.9	88	88

3.1.2. 20-fold cross-validation

In 20-fold cross-validation without PCA, SVM achieved the highest performance (93.4% accuracy, 0.996 AUC), followed by NN (93.1% accuracy, 0.995 AUC) and LR (92.9% accuracy, 0.995 AUC), while KNN performed worst (73.4% accuracy, 0.925 AUC), confirming its limitations with high-dimensional data. With PCA applied, SVM remained the best-performing model, achieving 90.1% accuracy and an AUC of 0.991. Meanwhile, KNN showed a slight improvement in accuracy (74.2%) due to reduced feature space noise, as illustrated in Figure 5 and Table 3). Overall, PCA slightly reduced the accuracy of the top models but preserved key features while lowering computational complexity. The 20-fold cross-validation provided robust performance estimates across all classifiers.

The mean values of the first 100 PCA components are shown in Figure 6, which arrange in a narrow range of approximately 0.292 to 0.308. This uniform distribution indicates consistent scaling and balanced contribution across components after normalization. The selected 100 components maintain 76% of the total variance, confirming that PCA preserves the essential structure of the original 2,048-dimensional feature space while reducing redundancy. This balance across components indicates that the PCA is suitable for dimensionality reduction and ensures that the most informative features remain available for downstream classification tasks.

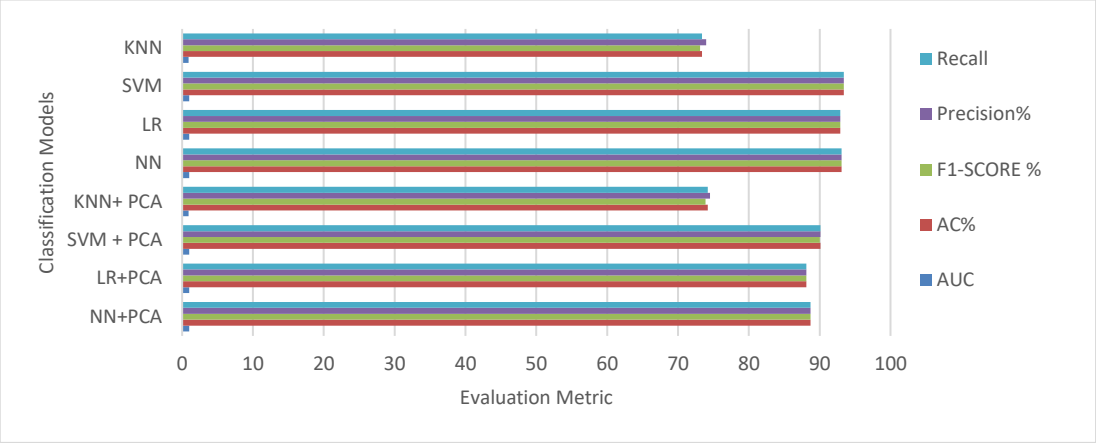


Figure 5. Comparison of four ML models with and without PCA using 20-fold cross validation

Table 3. Performance of ML models with and without PCA using 20-fold cross-validation

Models	AUC	AC%	F1-score%	Precision%	Recall%
NN+PCA	0.989	88.7	88.7	88.7	88.7
LR+PCA	0.985	88.1	88.1	88.1	88.1
SVM+PCA	0.991	90.1	90	90.1	90.1
KNN+PCA	0.928	74.2	73.9	74.5	74.2
NN	0.995	93.1	93.1	93.1	93.1
LR	0.995	92.9	92.9	92.9	92.9
SVM	0.996	93.4	93.4	93.4	93.4
KNN	0.925	73.4	73.1	74	73.4

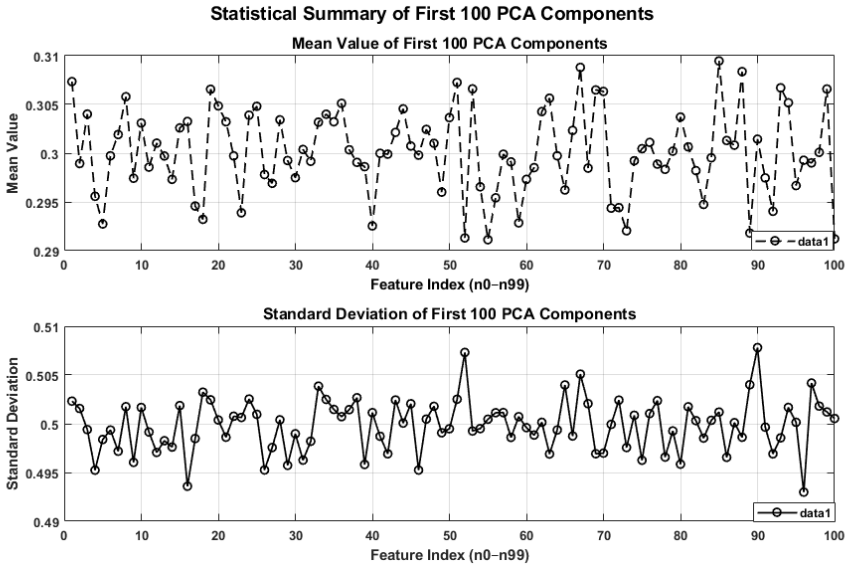
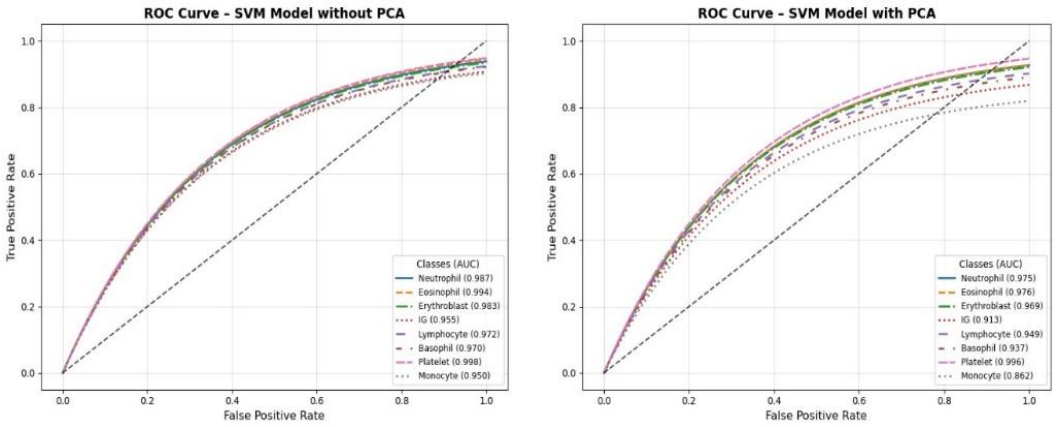


Figure 6. Statistical Summary of the first 100 PCA components

3.2. Class-wise performance

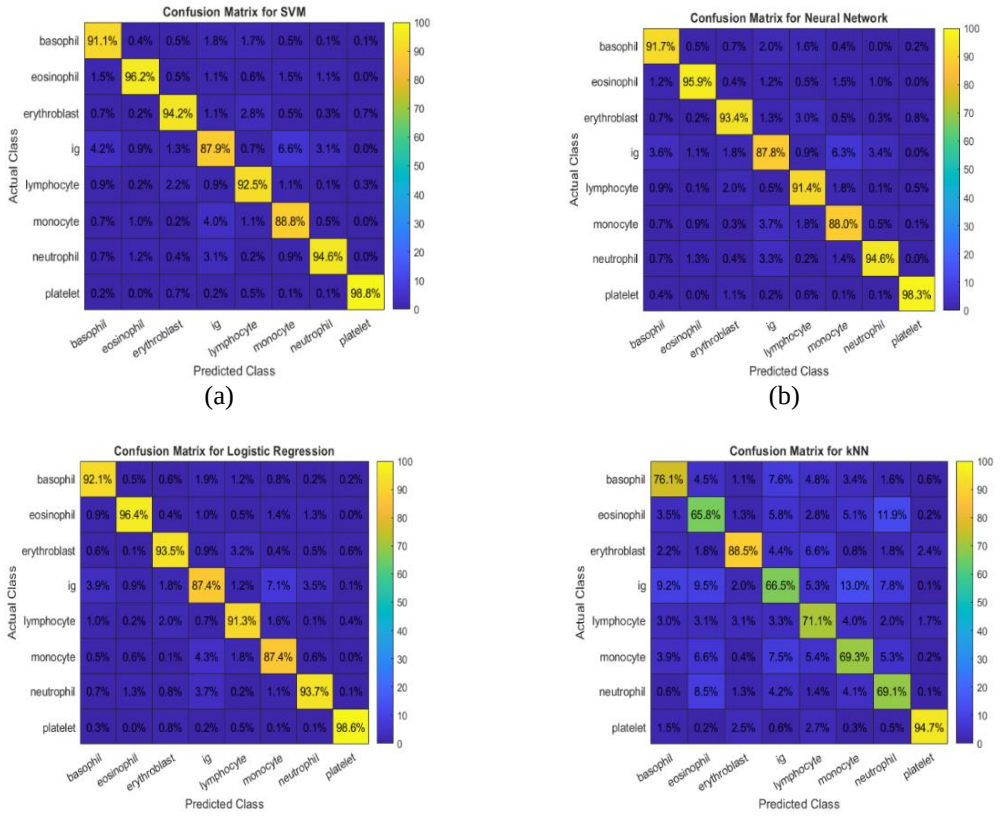
Figures 7(a) and (b) illustrates the Class-wise ROC analysis, confirming that superior and consistent performance across all blood cell types was achieved by SVM model. Without PCA, the model obtained the highest AUC values with (0.998) for platelets, (0.994) for eosinophils, and (0.987) for neutrophils. After applying PCA, the performance declined slightly, most notably for monocytes (AUC decline from 0.950 to 0.862) and IGs (from 0.955 to 0.913), This may be caused because of the loss of fine morphological details which preserved only in the original high-dimensional features.



(a) (b)

Figure 7. Class-wise ROC curves for all blood cell types using the SVM model with 20-fold cross-validation: (a) without PCA and (b) with PCA

The confusion matrices obtained from 20-fold cross-validation ML (SVM, NN, LR, and KNN) without PCA are illustrated in Figures 8(a)-(d) respectively. These results showing that neutrophils, eosinophils, and platelets were classified consistently with high precision and recall across all models. As a result, this can reflect clear morphological distinctions. While monocytes and IGs showed higher misclassification rates, indicating that these cell types require the full high-dimensional feature set to preserve subtle structural differences.



(a) (b)

(c) (d)

Figure 8. Confusion matrices of KNN, SVM, LR, and NN without PCA using 20-fold cross-validation: (a) SVM, (b) NN, (c) LR, and (d) KNN

Figure 9 presents the confusion matrices after applying PCA with ML models. The dimensionality reduction improved overall model stability and reduced noise, Figure 9(a) particularly for SVM and Figure 9(b) for NN, both of which maintained high accuracy and balanced performance across all classes. LR is illustrated in Figure 9(c), also demonstrated more consistent predictions after PCA. However, KNN was more negatively affected, as illustrated in Figure 9(d), showing a noticeable decline in precision due to its sensitivity to feature compression and distance distortions introduced by PCA [15].

Overall, combining PCA with 20-fold cross-validation enhanced computational efficiency and model robustness. Among all methods, SVM and NN achieved the most reliable and stable blood cell classification under both original and reduced feature spaces.

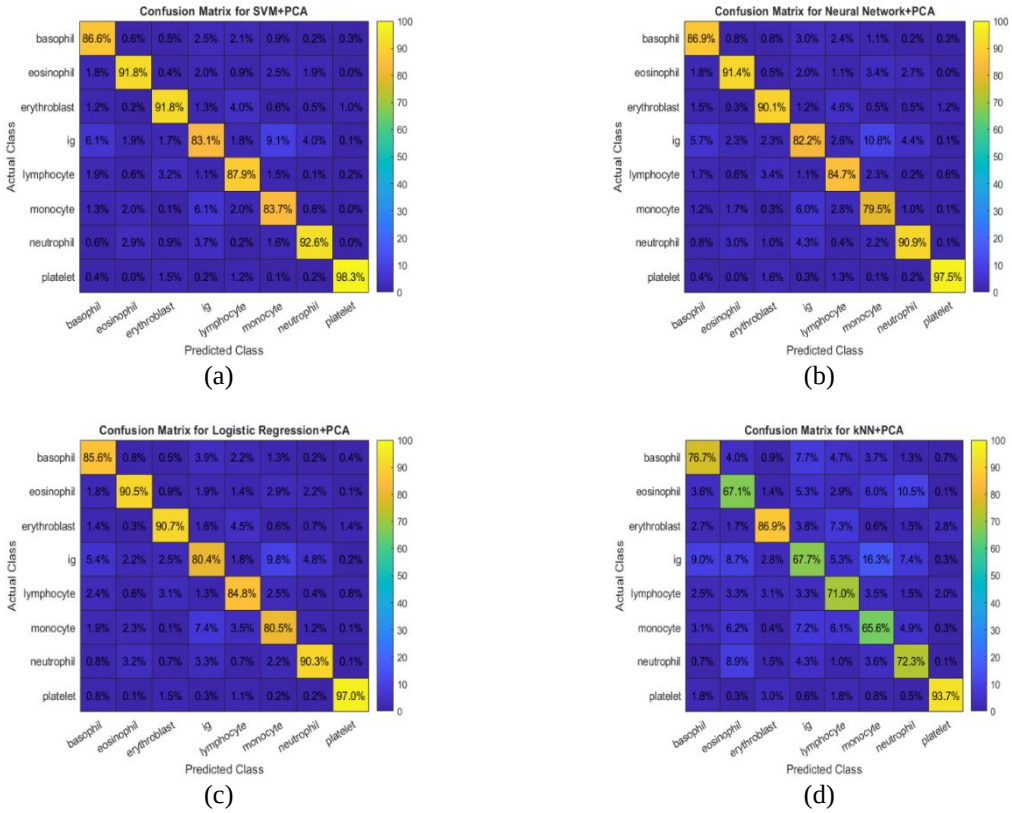


Figure 9. Confusion matrices of SVM, NN, LR, and KNN with PCA using 20-fold cross-validation: (a) SVM (b) NN, (c) LR, and (d) KNN

3.3. Comparative model analysis

The performance of SVM and KNN with and without PCA under different validation strategies is summarized in Table 4. Overall, SVM achieved the highest accuracy of (93.4%) and AUC of (0.996) in the 20-fold cross-validation without PCA. This confirms the robustness of the models when operating on the full 2,048-dimensional feature set. After applying PCA, the accuracy of SVM decreased slightly to (90.1%), reflecting the trade-off between computational efficiency and the loss of subtle morphological cues. The reduction in discriminative performance particularly affected monocytes and IGs, which rely on fine-grained features that may be attenuated during dimensionality reduction.

Table 4. Comparative performance of SVM and KNN with and without PCA under different validation strategies

Method	Feature set	KNN accuracy / AUC	SVM accuracy / AUC
66%-34% split	Without PCA	73.4% / 0.935	92.6% / 0.995
66%-34% split	with PCA	72.9% / 0.920	89.3% / 0.990
20-fold	Without PCA	73.4% / 0.925	93.4% / 0.996
20-fold	With PCA	74.2% / 0.928	90.1% / 0.991

The performance of KNN was consistently lower than SVM under all configurations. The model struggled most in the high-dimensional space without PCA, highlighting its sensitivity to the curse of dimensionality. PCA improved KNN performance slightly (with 74.2% accuracy in 20-fold cross-validation), consistent with theoretical expectations that distance-based classifiers benefit from compressed feature spaces and reduced noise.

The use of 20-fold cross-validation provided more reliable performance estimates than the 66%–34% split by allowing every sample to contribute to both training and evaluation. This repeated sampling process reduces variance, mitigates class imbalance effects, and offers a more realistic assessment of the model's generalization capability [33].

3.4. Comparison to referenced methods

Overall, the proposed SVM-based approach provides the best balance of accuracy, robustness, and clinical applicability compared to previously published methods, as shown in Table 5. Unlike earlier studies, the proposed method evaluates eight blood cell types, incorporates PCA-based dimensionality reduction, and demonstrates stable performance even after feature compression.

Table 5. Comparison of the proposed SVM-based method with previously published methods

Reference	Techniques	Accuracy	Limitation
[9]	DL feature extraction + ML	89%	No evaluation under dimensionality reduction; limited multi-class validation
[12]	Attention-based Segmentation	93%	Designed for segmentation, not classification; computationally heavy; no PCA integration
[17]	ML techniques	90%	Binary classification only (WBC vs RBC); no evaluation of subclass accuracy
[7]	CNN	91.3%	No PCA; high dataset of WBC types
[13]	AI models comparison	88%	General disease prediction focus; does not evaluate multi-class imbalance or PCA impact
Proposed work (2025)	InceptionV3 + SVM+PCA/no PCA	92.6% (No PCA), 90.1% (PCA)	Superior resilience to feature reduction; highest robustness across all 8 cell types with less features

4. CONCLUSION

SVM robustly classified all blood cell types, achieving 92.6% accuracy without PCA and 90.1% with PCA. Platelets, eosinophils, and neutrophils showed the highest AUCs 0.998, 0.994, 0.987, respectively. PCA slightly reduced accuracy, particularly for monocytes and IGs, but improved computational efficiency. Overall, full features maximize accuracy, while PCA provides a practical trade-off for faster processing.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Ruaa H. Ali Al-Mallah	✓	✓			✓	✓		✓	✓				✓	✓

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

The data that support the findings of this study are openly available in [Kaggle] at https://www.kaggle.com/datasets/unclesamulus/blood-cells-image-dataset, reference number [22].

REFERENCES

[1] G. Apostolopoulos, S. Tsinopoulos, and E. Dermatas, "Recognition and identification of red blood cell size using angular radial transform and neural networks," in *XII Mediterranean Conference on Medical and Biological Engineering and Computing 2010*, P. D. Bamidis and N. Pallikarakis, Eds., Berlin, Heidelberg: Springer, 2010, pp. 707–710, doi: 10.1007/978-3-642-13039-7_178.

[2] J. A. Alkrimi, L. E. George, A. Suliman, A. R. Ahmad, and K. Al-jashamy, "Isolation and classification of red blood cells in anemic microscopic images," *International Journal of Medical, Health, Pharmaceutical and Biomedical Engineering*, vol. 8, no. 10, pp. 686–689, 2014, doi: 10.5281/ZENODO.1096591.

[3] D. Cruz *et al.*, "Determination of blood components (WBCs, RBCs, and Platelets) count in microscopic images using image processing and analysis," in *2017IEEE 9th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM)*, Dec. 2017, pp. 1–7, doi: 10.1109/HNICEM.2017.8269515.

[4] G. Litjens *et al.*, "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, pp. 60–88, Dec. 2017, doi: 10.1016/j.media.2017.07.005.

[5] A. K. Alhialy, "Optimizing AI for white blood cell analysis: a multi-objective neural architecture search approach," *NTU Journal of Engineering and Technology*, vol. 4, no. 3, pp. 28–37, Sep. 2025, doi: 10.56286/n9amtj50.

[6] S. Q. Hasan, "Shallow model and deep learning model for features extraction of images," *NTU Journal of Engineering and Technology*, vol. 2, no. 3, pp. 1–8, Nov. 2023, doi: 10.56286/ntujet.v2i3.449.

[7] M. Ammar, M. Daho, K. Harrar, and A. Laidi, "Feature extraction using CNN for peripheral blood cells recognition," *EAI Endorsed Transactions on Scalable Information Systems*, vol. 9, no. 34, p. e12, Jul. 2018, doi: 10.4108/eai.20-10-2021.171548.

[8] K. Pearson, "LIII. On lines and planes of closest fit to systems of points in space," *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 2, no. 11, pp. 559–572, Nov. 1901, doi: 10.1080/14786440109462720.

[9] A. Cayir, I. Yenidogan, and H. Dag, "Feature extraction based on deep learning for some traditional machine learning methods," in *2018 3rd International Conference on Computer Science and Engineering (UBMK)*, Sep. 2018, pp. 494–497, doi: 10.1109/UBMK.2018.8566383.

[10] L. Pinto-Coelho, "How artificial intelligence is shaping medical imaging technology: a survey of innovations and applications," *Bioengineering*, vol. 10, no. 12, p. 1435, Dec. 2023, doi: 10.3390/bioengineering10121435.

[11] N. J. Schork, "Artificial intelligence and personalized medicine," in *Precision Medicine in Cancer Therapy*, D. D. Von Hoff and H. Han, Eds., Cham: Springer International Publishing, 2019, pp. 265–283, doi: 10.1007/978-3-030-16391-4_11.

[12] A. Sinha and J. Dolz, "Multi-scale self-guided attention for medical image segmentation," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 1, pp. 121–130, Jan. 2021, doi: 10.1109/JBHI.2020.2986926.

[13] N. Ghaffar Nia, E. Kaplanoglu, and A. Nasab, "Evaluation of artificial intelligence techniques in disease diagnosis and prediction," *Discover Artificial Intelligence*, vol. 3, no. 1, p. 5, Jan. 2023, doi: 10.1007/s44163-023-00049-5.

[14] M. Mahmud, M. S. Kaiser, A. Hussain, and S. Vassanelli, "Applications of deep learning and reinforcement learning to biological data," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 6, pp. 2063–2079, Jun. 2018, doi: 10.1109/TNNLS.2018.2790388.

[15] L. Alzubaidi, M. A. Fadhel, O. Al-Shamma, J. Zhang, and Y. Duan, "Deep learning models for classification of red blood cells in microscopy images to aid in sickle cell anemia diagnosis," *Electronics*, vol. 9, no. 3, p. 427, Mar. 2020, doi: 10.3390/electronics9030427.

[16] I. Singh, N. P. Singh, H. Singh, S. Bawankar, and A. Ngom, "Blood Cell Types Classification Using CNN," *Lecture Notes in Computer Science. Springer International Publishing*, pp. 727–738, 2020. doi: 10.1007/978-3-030-45385-5_65.

[17] K. . Navya, K. Prasad, and B. M. K. Singh, "Classification of blood cells into white blood cells and red blood cells from blood smear images using machine learning techniques," in *2021 2nd Global Conference for Advancement in Technology (GCAT)*, IEEE, Oct. 2021, pp. 1–4, doi: 10.1109/GCAT52182.2021.9587524.

[18] T. A. Elhassan *et al.*, "Classification of atypical white blood cells in acute myeloid leukemia using a two-stage hybrid model based on deep convolutional autoencoder and deep convolutional neural network," *Diagnostics*, vol. 13, no. 2, p. 196, Jan. 2023, doi: 10.3390/diagnostics13020196.

[19] M. A. Ali, F. Dornaika, and I. Arganda-Carreras, "White blood cell classification: convolutional neural network (CNN) and vision transformer (ViT) under medical microscope," *Algorithms*, vol. 16, no. 11, p. 525, Nov. 2023, doi: 10.3390/a16110525.

[20] A. Heni, I. Jdey, and H. Ltifi, "Blood cells classification using deep learning with customized data augmentation and EK-means segmentation," *Journal of Theoretical and Applied Information Technology*, vol. 101, no. 3, pp. 1162–1173, 2023.

[21] H. Sun, X. Wan, S. Tang, and Y. Li, "SSW-YOLO: enhanced blood cell detection with improved feature extraction and multi-scale attention," *Journal of Imaging Informatics in Medicine*, vol. 38, no. 6, pp. 4100–4118, 2025, doi: 10.1007/s10278-025-01460-3.

[22] UNCLESAMULUS, "Blood cells image dataset: a dataset for microscopic peripheral blood cell images," Kaggle. [Online]. Available: https://www.kaggle.com/datasets/unclesamulus/blood-cells-image-dataset (accessed: Jan, 10 2025)

[23] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2016, pp. 2818–2826, doi: 10.1109/CVPR.2016.308.

[24] M. Shoaib and N. Sayed, "YOLO object detector and inception-V3 convolutional neural network for improved brain tumor segmentation," *Traitement du Signal*, vol. 39, no. 1, pp. 371–380, Feb. 2022, doi: 10.18280/ts.390139.

[25] S. Likhitha and R. Baskar, "Skin cancer segmentation using R-CNN comparing with inception V3 for better accuracy," in *2022 11th International Conference on System Modeling & Advancement in Research Trends (SMART)*, IEEE, Dec. 2022, pp. 1293–1297, doi: 10.1109/SMART55829.2022.10047686.

[26] O. Iparraguirre-Villanueva, V. Guevara-Ponce, O. R. Paredes, F. Sierra-Liñan, J. Zapata-Paulini, and M. Cabanillas-Carbonell, "Convolutional neural networks with transfer learning for pneumonia detection," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 9, pp. 544–551, 2022, doi: 10.14569/IJACSA.2022.0130963.

[27] D. Saini, T. Chand, D. K. Chouhan, and M. Prakash, "A comparative analysis of automatic classification and grading methods for knee osteoarthritis focussing on X-ray images," *Biocybernetics and Biomedical Engineering*, vol. 41, no. 2, pp. 419–444, Apr. 2021, doi: 10.1016/j.bbe.2021.03.002.

[28] E. Micheli-Tzanakou, H. Sheikh, and B. Zhu, "Neural networks and blood cell identification," *Journal of Medical Systems*, vol. 21, no. 4, pp. 201–210, Aug. 1997, doi: 10.1023/A:1022899519704.
[29] P. Tabesh, G. Lim, S. Khator, and C. Dacso, "A support vector machine approach for predicting heart conditions," in *IIE Annual Conference and Expo 2010 Proceedings*, A. Johnson and J. Miller, Eds., 2010.
[30] S. H. Shetty, S. Shetty, C. Singh, and A. Rao, "Supervised machine learning: algorithms and applications," in *Fundamentals and Methods of Machine and Deep Learning*, P. Singh, Ed., Wiley, 2022, ch. 1, pp. 1–16, doi: 10.1002/9781119821908.ch1.
[31] 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, p. 20150202, Apr. 2016, doi: 10.1098/rsta.2015.0202.
[32] M. Bhagat and B. Bakariya, "A comprehensive review of cross-validation techniques in machine learning," *International Journal on Science and Technology*, vol. 16, no. 1, Jan. 2025, doi: 10.71097/IJSAT.v16.i1.1305.
[33] M. M. M. Al-Hatab, R. R. Omar Al-Nima, and M. A. Qasim, "Classifying healthy and infected Covid-19 cases by employing CT scan images," *Bulletin of Electrical Engineering and Informatics*, vol. 11, no. 6, pp. 3279–3287, Dec. 2022, doi: 10.11591/eei.v11i6.4344.

BIOGRAPHIES OF AUTHOR

Ruaa H. Ali Al-Mallah received the B.Sc. and M.Sc. degrees in Computer Engineering from the Technical Engineering College, Mosul, Northern Technical University (NTU), Mosul, Iraq. She is currently a lecturer with the Technical Engineering College for Computer and Artificial Intelligence, Northern Technical University, Mosul, Iraq. Her research interests include computer engineering and artificial intelligence. She can be contacted at email: ruaa_almallah@ntu.edu.iq.