IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received December 20, 2021, accepted February 1, 2022, date of publication February 7, 2022, date of current version February 15, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3149637

Feature Extraction of White Blood Cells Using CMYK-Moment Localization and Deep Learning in Acute Myeloid Leukemia Blood Smear Microscopic Images

TUSNEEM AHMED M. ELHASSAN[1], MOHD SHAFRY MOHD RAHIM[1], TAN TIAN SWEE[2], SITI ZAITON MOHD HASHIM[3], AND MAHMOUD ALJURF[4]

[1]School of Computing, Universiti Teknologi Malaysia, Johor Bahru 81310, Malaysia
[2]Bioinspired Device and Tissue Engineering Research Group, Faculty of Engineering, School of Biomedical Engineering and Health Sciences, Universiti Teknologi Malaysia, Skudai, Johor 81300, Malaysia
[3]Department of Data Science, Universiti Malaysia Kelantan, Kota Bharu 16100, Malaysia
[4]King Faisal Specialist Hospital and Research Centre, Riyadh 11211, Saudi Arabia

Corresponding author: Tusneem Ahmed M. Elhassan (amtusneem@graduate.utm.my)

ABSTRACT Artificial intelligence has revolutionized medical diagnosis, particularly for cancers. Acute myeloid leukemia (AML) diagnosis is a tedious protocol that is prone to human and machine errors. In several instances, it is difficult to make an accurate final decision even after careful examination by an experienced pathologist. However, computer-aided diagnosis (CAD) can help reduce the errors and time associated with AML diagnosis. White Blood Cells (WBC) detection is a critical step in AML diagnosis, and deep learning is considered a state-of-the-art approach for WBC detection. However, the accuracy of WBC detection is strongly associated with the quality of the extracted features used in training the pixel-wise classification models. CAD depends on studying the different patterns of changes associated with WBC counts and features. In this study, a new hybrid feature extraction method was developed using image processing and deep learning methods. The proposed method consists of two steps: 1) a region of interest (ROI) is extracted using the CMYK-moment localization method and 2) deep learning-based features are extracted using a CNN-based feature fusion method. Several classification algorithms are used to evaluate the significance of the extracted features. The proposed feature extraction method was evaluated using an external dataset and benchmarked against other feature extraction methods. The proposed method achieved excellent performance, generalization, and stability using all the classifiers, with overall classification accuracies of 97.57% and 96.41% using the primary and secondary datasets, respectively. This method has opened a new alternative to improve the detection of WBCs, which could lead to a better diagnosis of AML.

INDEX TERMS Acute myeloid leukemia (AML), white blood cell (WBC) feature extraction, deep learning, feature fusion, CNN, ROI.

I. INTRODUCTION

Features are data descriptors used to describe data elements such as classification and clustering. A comprehensive understanding of WBC features is critical for differentiating between various types and subtypes of leukemia. Current methods used for WBC detection, segmentation, and classification face several challenges, although they are performed using automatic and manual approaches [1]. Manual detection of WBCs is conducted by pathologists and is typically subject to human error and produces inaccurate results. This process is tedious, time-consuming, and subject to inter- and intra-class variations among pathologists. Only 76.6% of the cases showed agreement between pathologists during leukemia diagnosis [2]. Other challenges are associated with the complex nature of WBCs, including irregular boundaries and the textural similarities between WBCs and other blood components, which cause difficulties in separating

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

WBCs from other blood components [1], [3]. Also, WBCs are complex in terms of shape, texture, color, and intensity diversity [4], [5]. WBCs are heterogeneous and have different subtypes, including normal and abnormal cells. In addition, different staining and illumination variations render WBC recognition more difficult [6], [7]. However, current automated methods of WBC detection used in laboratories primarily focus on quantitative rather than qualitative methods used in image processing and pattern recognition [1], [8], [9]. Therefore, the use of new computer-aided systems for accurate WBC detection can aid the development of stable and generalized learning systems. In this study, a new feature extraction method for WBC detection was proposed. The proposed method is a hybrid CMYK localization method based on image processing and a deep-learning-based feature fusion method using a CNN. The proposed method can also be used to build a semantic segmentation model to help pathologists detect and localize WBCs to improve the diagnosis accuracy.

II. RELATED WORKS

WBC recognition is a challenging task because of the complex nature of cell images, which makes the identification of significant WBC features more difficult. Researchers have attempted to extract and identify significant features of WBCs to discriminate between WBCs and other blood components. WBC features can be categorized into two types: handcrafted features and deep-learning-based features. Handcrafted features are obtained using image processing techniques and are used with traditional machine-learning (ML) algorithms. Conversely, deep learning-based features are automatic features extracted using deep learning models and can be used with fully connected layers (part of a deep learning model) or linked to an external ML classifier. Many researchers have used handcrafted features to perform WBC recognition and segmentation and have shown good performance [10]–[16]. However, these methods exhibit marked limitations in terms of efficiency and generalization for solving complex problems [17], [18]. Therefore, several researchers have focused on investigating deep-learning-based features [6], [19]–[26]. Lu *et al.* [27] extracted and fused multiscale features using a feature encoder with residual blocks to develop a WBC segmentation system. They also used convolution and deconvolution decoder to improve the WBC segmentation mask. Their method was evaluated using four datasets of normal WBCs: neutrophils, eosinophils, basophils, monocytes, and lymphocytes. Their system achieved the best results compared to other benchmark methods.

Roy *et al.* [7] proposed a white blood cell (WBC) feature extraction method using the DeepLabv3+ architecture and a ResNet-50 feature extractor to extract WBC features. The extracted features were then used to build a WBC segmentation system. The system was evaluated using three different public datasets and achieved 96.1% segmentation accuracy. Abdurrazzaq *et al.* [28] used a singular value decomposition approach to localize WBCs using the similarity level of features between WBCs. Their results showed an improvement in WBC nuclei detection as well as WBC detection, particularly for WBCs with light color intensity, compared to other methods. Their method achieves an average segmentation accuracy of 63%. Khomairoh *et al.* [29] used the Haar cascade model for WBC feature extraction to extract ROIs for WBC segmentation. The model was built using a dataset of M4, M5, and M7 AML subtypes, and features were extracted using different convolution kernels including edges, lines, and four-rectangle kernels. Subsequently, a color-based method was used for nuclear and cytoplasmic segmentation. The overall accuracies of nucleus segmentation were 87.5%, 90.4%, and 84.6% for M4, M5, and M7, respectively. However, for cytoplasmic segmentation, the model achieved overall accuracies of 75%, 71.4%, and 80.76% for M4, M5, and M7, respectively. Hegde *et al.* [21] compared the performance of handcrafted features with that of deep-learning-based features using a CNN. The handcrafted features considered included shape, color, and texture, and deep-learning-based features were extracted from the fc6, fc7, and fc8 layers of AlexNet. The two approaches were then compared using a neural network (NN) classifier. Both methods achieved comparable results, with an overall accuracy of 99%. Saleem *et al.* [20] used feature fusion with DarkNet-53 and ShuffleNet to extract WBC features for both segmentation and classification, and achieved 98.6% segmentation accuracy.

Ramya *et al.* [30] extracted a set of image levels and statistical features using segmented WBCs for classification into AML and normal. The extracted features included color, shape, and gray-level co-occurrence matrix (GLCM). Rad *et al.* [31] used statistical and morphological features to develop a new object detection technique to overcome the problem of the initial contour in the level-set segmentation method. They used statistical and morphological features inside and outside the contour to develop an automatic region-based initial contour. Their method achieved an overall accuracy of 96%, was evaluated using two external datasets, and achieved optimal results. Puigdollers *et al.* [16] used a bag-of-words approach to extract the local image descriptors for WBC detection. Their method achieved an overall accuracy of 80% and did not require carefully crafted features to localize WBCs, thus making it simpler and more generalized.

Loddo *et al.* [32] extracted color and statistical features to train a multiclassification system based on an SVM and KNN. Color features were calculated pixel-wise by averaging the color values of each pixel using a 3×3 pixel neighborhood. The model achieved 99% accuracy, and was extended to develop a WBC counting system using a circular Hough transform, which achieved 99.7% accuracy [33].

Literature shows that WBC feature extraction has primarily focused on normal WBCs and acute lymphoid leukemia (ALL). However, owing to several challenges, limited research has been conducted on the recognition of WBCs obtained from AML patients. Therefore, this study focuses on feature extraction using several types of WBCs,

including normal WBCs and WBCs obtained from AML microscopic images.

This study makes the following contributions to the literature:

1. We developed a new WBC localization method based on the CMYK color space transformation and image moments, named CMYK-moment localization.
2. We proposed a new CNN-based feature extraction method based on the feature fusion of pointwise and localized features by combining a shallow layer with a deep stacked layer to extract generalized features without losing pixel information originality.
3. We proposed a hybrid WBC feature extraction framework for CMYK-moment localization and CNN-based feature extraction based on feature fusion.

III. METHODS

A. DATASET

This study used a single-cell morphological dataset of leukocytes from patients with AML and non-malignant controls (AML_Cytomorphology_LMU). The dataset consisted of 18,365 expert-labeled single-cell images obtained from peripheral blood smears of 100 AML patients and 100 controls at Munich University Hospital between 2014 and 2017. The dataset is classified into 15 different types of single-cell images. Four of these were leukemic cells and the other 11 were normal blood cells. Among the 11 types, seven were mature leukocytes and four were immature. Cancerous and noncancerous WBCs were classified by expert pathologists based on standard morphological classification [26], see Figure1.

A secondary dataset of 17,092 normal peripheral blood samples from individual cells was used to evaluate the model performance. The dataset was obtained using CellaVision DM96 in RGB color space. The images were obtained using the RGB color space, jpg format, and 360 × 363 dimensions, and were labeled by an expert pathologist. The dataset consists of eight classes of different types of blood cells, including segmented neutrophils, eosinophils, basophils, lymphocytes, monocytes, erythroblasts, metamyelocytes, myelocytes, promyelocytes, and platelets [34].

B. PROPOSED MODEL

The proposed feature extraction method is a hybrid of the CMYK-moment localization and CNN feature fusion. In this study, the CMYK-moment localization method was used to extract the ROI by using a combination of the color transformation method (CMYK) and image moments. An ROI was used to reduce the amount of irrelevant information to extract context-free WBC features that depend only on WBC cells, which helps identify more generalized WBC features that can be used to detect different types of WBCs [29]. The elimination of unnecessary information also mitigates overfitting and decreases computation time. Extracting features that can successfully identify different types of WBCs and discriminate

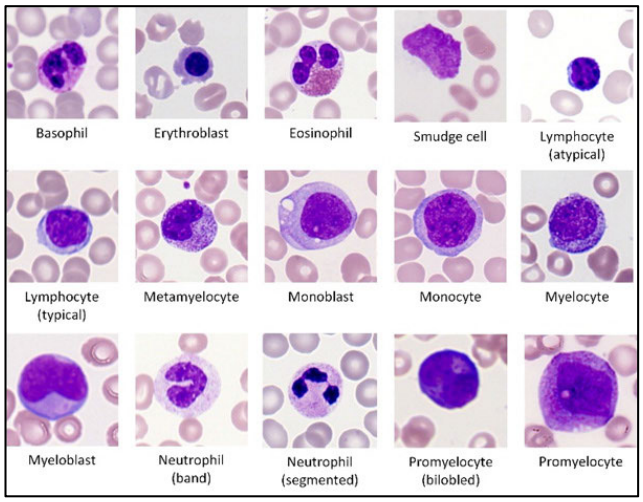

FIGURE 1. Samples of the fifteen different types of WBC presented in the primary dataset including normal and abnormal cells.

between WBCs and other blood components is challenging because of the complex biological nature of WBCs, such as their shape, texture, color, density variations [4], [5], irregular boundaries and textural similarities between WBCs and other blood components. However, deep CNN convolutional filters can extract complex textural patterns compared to other conventional texture feature extraction methods, such as Gabor filters [1], [3]. In general, CNN convolutional filters have been shown to achieve better performance with images than with other types of data [35]. The proposed CNN feature fusion model consisted of four layers. The first layer is the input layer of the RGB images, followed by two convolutional layers. The first layer is a single pointwise layer and a stacked spatial layer consisting of two layers. The first convolutional layer helps extract simple features such as edges, whereas the second layer is used to extract more complex patterns such as texture features. The proposed feature extraction method was divided into four phases: Phase I (ROI localization), Phase II (feature extraction), and Phase III (model evaluation). The proposed method is illustrated in Figure 2.

1) PHASE I: ROI LOCALIZATION

In this phase, an ROI was extracted using the CMYK moment localization method. In this method, RGB images are converted into the CMYK color space. The C channel was then extracted, and the Otsu thresholding method was applied to generate a binary mask to extract the WBC nucleus. Post-processing operations were then applied using morphological opening and maximum connected region (MCR) to remove isolated components and obtain a refined binary mask. The centroid of the nucleus was then calculated using image moments. The nucleus centroid was calculated using the following equation:

$$M_{ij} = \sum_{x} \sum_{y} x^i y^j I(x, y) \tag{1}$$

$$C_X = \frac{M_{10}}{M_{00}} \tag{2}$$

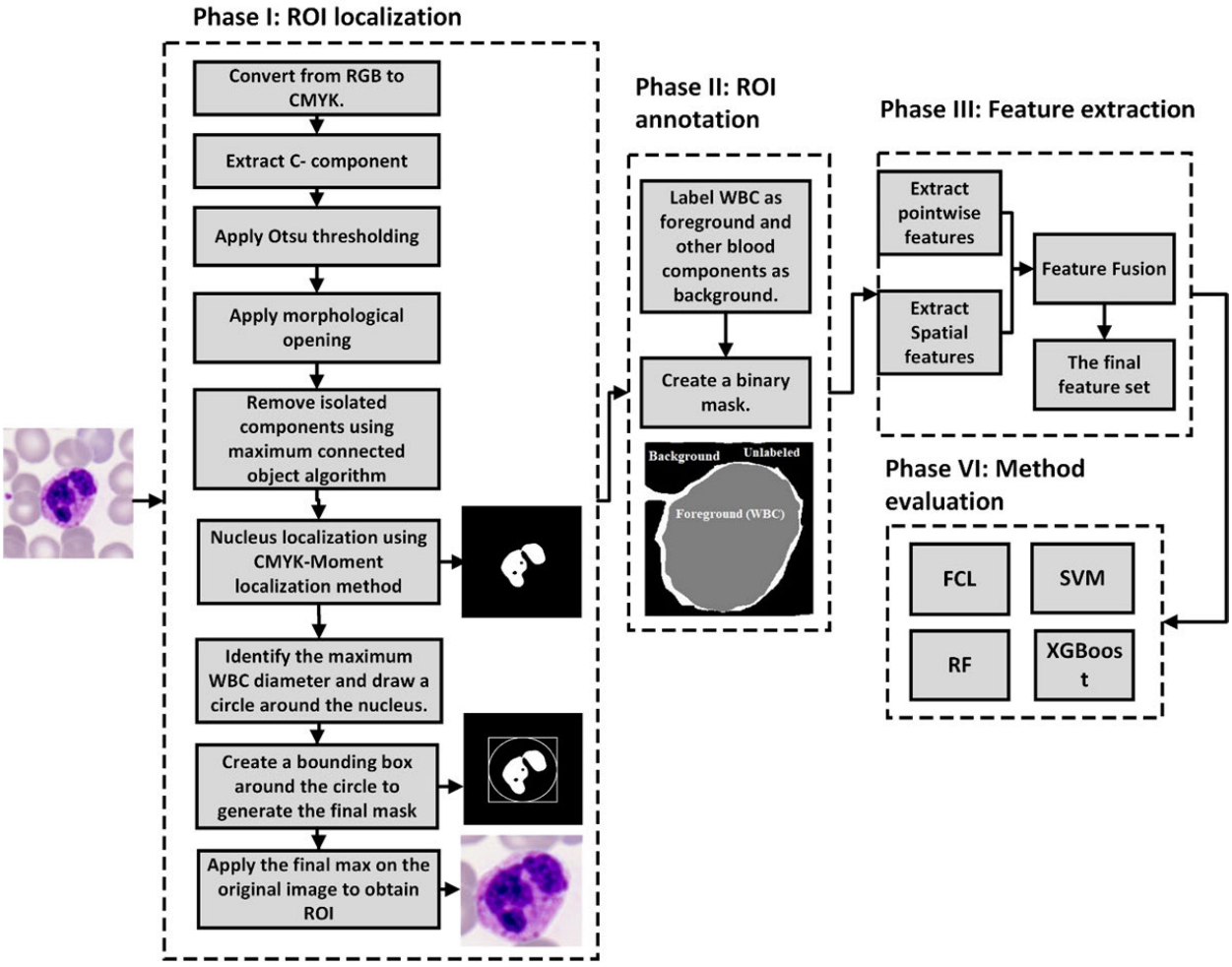

FIGURE 2. The proposed WBC feature extraction method framework.

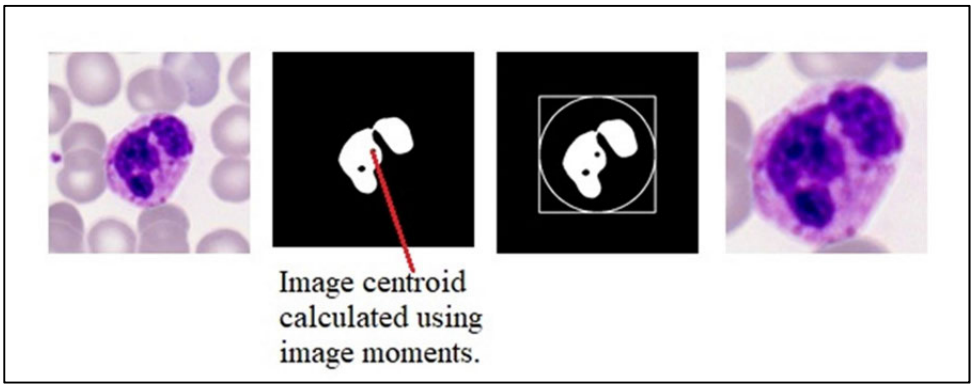

FIGURE 3. Extraction of ROI using the CMYK-Moment's localization method.

$$C_Y = \frac{M_{01}}{M_{00}} \tag{3}$$

where I (x, y) is the image intensity and C_X and C_Y are the x and y coordinates, respectively.

The maximum diameter of WBCs was determined by drawing a circle around the centroid. A square polygon was then drawn around the circle to extract the ROI, Figure 3.

2) PHASE II: FEATURE EXTRACTION

In this phase, the ROI images were annotated into a foreground representing WBCs and a background representing other blood components. Subsequently, 2D CNN convolutional layers were used to extract features using the feature fusion of pointwise and localized features.

Algorithm: Proposed feature extraction method.

1. Read RGB image as I (x, y).
2. **Get the image dimensions of I (x, y):**
   M = width of I (x, y).
   N =height of I (x, y).
3. **Define the output class levels as:**
   WBC=1.
   Non-WBC= 0.
4. **Convert to CMYK:**
   CALL Convert_To_CMYK with current image I (x, y).
5. Extract C channel.
6. Convert the image to grayscale.
7. **Apply OTSU thresholding:**
   Calculate probabilities for each image intensity level using histogram:
   For i = 1 to M
   For j = 1 to N
   Calculate probability intensity p (I (x, y))
   Initialize the output class probabilities.
   Initialize the output class means.
   T= domain of thresholds
   WCV =0
   For t = 1 to T:
   For i = 0 to 1:
   Calculate w_i
   Calculate σ_i
   Calculate within-class variance WCV_i as $WCV_i = w_i \sigma_i$
   WCV [i] ⟵ WCV_i
   End for
   Calculate within-Class variance WCV as $WCV_t = \sum_{i=0}^{1} WCV\ [i]$
   WCV[t] ⟵ WCV_t
   End for
   T_{final}= min (WCV[t])

   If I (x, y) < T_{final}:

   I (x, y) = 0

   Else

   I (x, y) =1.
8. **Apply morphological opening:**
   S (5,5) ⟵ Structural element of 5
   o ⟵ opening operation.
   $\ominus$ ⟵ erosion operation.
   $\oplus$ ⟵ dilation operation.
   $I(i, j) \circ S(5,5) = I(x, y) \ominus (S(5,5) \oplus (S(5,5))$
9. **Calculate the MCR:**
   Label all image blobs after morphological opening.
   Calculate the size of each blob.
   Select the blob with the maximum size as WBC nucleus.
10. Create a new image with the extracted WBC nucleus.
11. **Calculate the nucleus centroid using image moments:**

FIGURE 4. **Algorithm of the proposed feature extraction method (pseudocode).**

$$M_{ij} = \sum_x \sum_y x^i y^j I(x,y)$$

Calculate the x coordinate of the image centroid as $C_X = \frac{M_{10}}{M_{00}}$

Calculate the y-coordinate of the image centroid as $C_Y = \frac{M_{01}}{M_{00}}$

12. Calculate the maximum WBC diameter d.
13. Draw a circle (C) around the nucleus, using $(x - C_X)^2 + (y - C_Y)^2 = d^2$.
14. Draw square (S) with a side length of d around C.
15. ROI ⟵ S.
16. **Extract 64 pointwise CNN features from ROI:**
    $h_j(X)$ ⟵ The jth feature maps.
    $g_{ij}(X)$ ⟵ kernel function of size 1x1.
    f_i ⟵ ROI input channel.
    $h_j(X) = \sum_{i=1}^{64} f_i \otimes g_{ij}(X)$.
17. **Extract 64 special CNN features from ROI:**
    $k_j(X)$ ⟵ The jth feature maps.
    $g_{ij}(X)$ ⟵ kernel function of size 3x3.
    f_i ⟵ ROI input channel.
    $k_j(X) = \sum_{i=1}^{64} f_i \otimes g_{ij}(X)$.
18. **Concatenate the extracted pointwise and special features**

$$c_j(X) = h_j(X) + k_j(X)$$

FIGURE 4. *(Continued.)* **Algorithm of the proposed feature extraction method (pseudocode).**

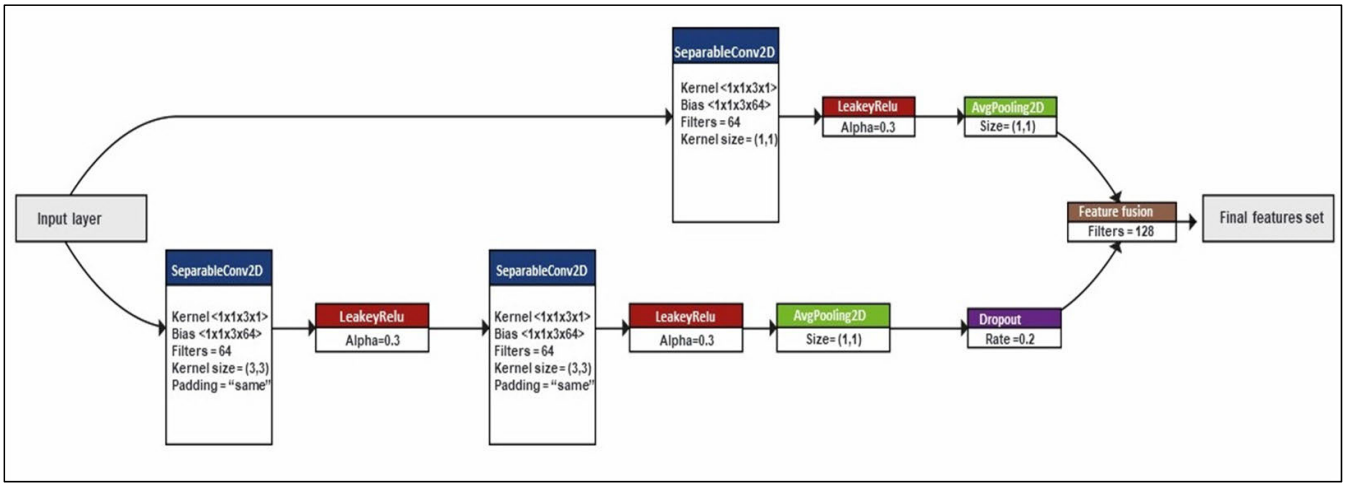

FIGURE 5. **CNN feature extraction model configuration.**

Convolutional layers are key components of a CNN architecture, and consist of a set of kernels used to convolve the input image and forward it to the top layers using the following equation:

$$h_j(X) = \sum_{i=1}^{C} f_i \otimes g_{ij}(X) \quad (4)$$

TABLE 1. Classification results of the proposed feature extraction method using primary dataset.

Primary dataset	Type of classifier	NO. OF Features	Overall Accuracy	Sensitivity	Specificity	Precision	F-score	AUC	Jaccard index	Computation time (Secs)
Proposed method	FCL	128	0.9734	0.9584	0.9489	0.9941	0.9759	0.9756	0.9530	997.5
	SVM	128	0.9715	0.9597	0.9502	0.9892	0.9742	0.9731	0.9498	38.7
	XGBoost	128	0.9739	0.9611	0.9520	0.9921	0.9764	0.9757	0.9539	1153.5
	RF	128	0.9757	0.9716	0.9642	0.9848	0.9782	0.9762	0.9574	3156.5

TABLE 2. Classification results of the proposed feature extraction method using secondary dataset.

Secondary dataset	Type of classifier	NO. OF Features	Overall Accuracy	Sensitivity	Specificity	Precision	F-score	AUC	Jaccard index
Proposed method	FCL	128	0.9492	0.9359	0.9351	0.9635	0.9495	0.9494	0.9039
	SVM	128	0.9641	0.9774	0.9758	0.9534	0.9653	0.9638	0.9330
	XGBoost	128	0.9518	0.9395	0.9385	0.9653	0.9522	0.9521	0.9088
	RF	128	0.9547	0.9603	0.9581	0.9514	0.9558	0.9545	0.9154

where h_j (X) is the j^{th} feature map obtained by the convolutional operation of the input image at the special location X = (x, y); g_{ij} is the kernel defined between the f_i input channel and the h_i feature map; and $\otimes$ is the convolution operation defined as follows:

$$f_i \otimes g_{ij}(x, y) = \sum_m \sum_n f_i(m, n) g_{ij}(x - m)(y - n) \quad (5)$$

Convolutional layers are used to learn the input image features, from simple to more complex features. The bottom layer of the CNN is designed to learn simple features such as edge features, which are similar to features learned by Gabor filters, whereas the following deeper layers are designed to learn more complex patterns such as texture features. However, compared with the Gabor method, convolution layers can be considered as an advanced texture feature extraction technique. The input for the convolution layer is a three-channel image represented as $3 \times W \times H$, whereas the output is a feature map of $F \times W' \times H'$ dimensions, where F is the number of output filters, and W', and H' are the width and height of the output filters, respectively. Assuming that k_w and k_h are the width and height of the convolution kernel, respectively, W' and H' are calculated as $W' = W\text{-}k_w + 1$ and $H' = H\text{-}k_h + 1$. Each filter in the output feature map was obtained by adding the individual filter weights of the three input channels (R, G, and B) [36], [37].

In this study, the features were extracted using 2D separable convolution, where the features were extracted from each channel separately. The extracted features were then averaged over the three channels using pointwise convolution. The model consisted of two parallel layers. The first layer is a single pointwise convolution layer of 64 filters using a 1×1 kernel size and LeakyReLU activation function with alpha = 0.3, followed by an average pooling layer of size two. The

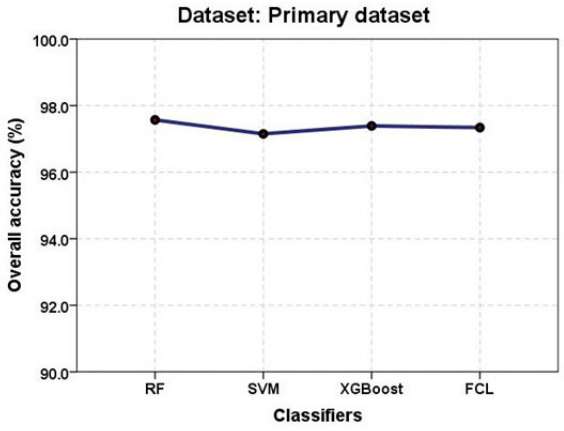

FIGURE 6. Overall accuracy of the proposed feature extraction method using RF, SVM, and XGBoost, and FCL on primary dataset.

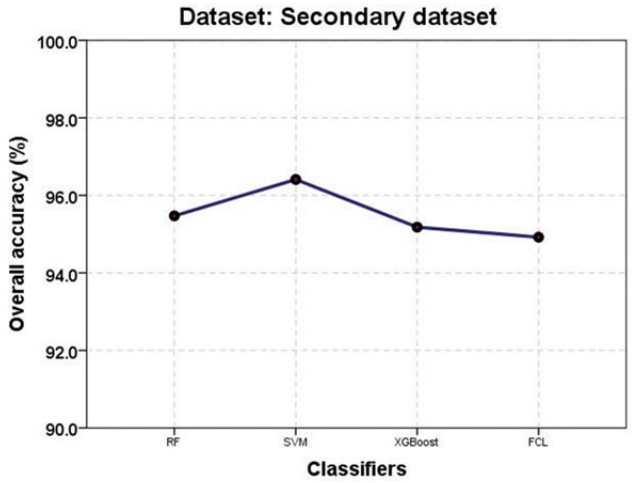

FIGURE 7. Overall accuracy of the proposed feature extraction method using RF, SVM, and XGBoost, and FCL on secondary dataset.

leakyReLU layer is represented by the following equation:

$$P_c^v(y) = \max\left(L_c^v(y), L_c^v(y)\, x0.3\right) \tag{6}$$

where $P_c^v(y)$ is the pixel value at location y of the c^{th} channel and v^{th} convolution layer after applying the LeakyReLU activation function. The average pooling can be represented by the following equation:

$$A_c^v(y) = \frac{1}{p_h p_w} \sum_{x=1}^{x=p_h p_w} P_c^v(x) \tag{7}$$

where $A_c^v(y)$ represents the pixel at location y after applying the average pooling operation; p_h *and* p_w represent the image height and width, respectively; and $P_c^v(x)$ represents the pixel at location x after applying the LeakyReLU activation function [24].

The pointwise layer was used to extract low-level features without losing pixel information owing to the multiple convolutional operations (Huang *et al.*, 2017). The second layer was a stacked layer consisting of two convolutional layers of 64 filters using a 3 × 3 kernel and a LeakyReLU activation function with an alpha coefficient of 0.3. The convolution layers were followed by average pooling and a dropout layer at a rate of 20% to avoid overfitting [38]. A zero-padding technique was applied to the input images to maintain the same output fitters' size as input images. Feature fusion is used to combine the features obtained from the two parallel branches to construct the final set of features. The algorithm for the proposed feature extraction method is illustrated in Figure 4. The extracted features were then mapped to the corresponding binary labels after removing the unlabeled data to obtain the final dataset for model training. Figure 5 shows the network configuration of the proposed method.

3) PHASE III: METHOD EVALUATION

The proposed feature extraction method was evaluated using several classification algorithms on two datasets: primary and secondary. The primary dataset was used for model training and validation, whereas the secondary dataset was used only for testing. The classification algorithms applied for method evaluation included fully connected layers (FCL) [39], random forest (RF) [40], support vector machine (SVM) [41], and XGBoost [42]. Classification performance was measured using seven evaluation metrics: overall accuracy, sensitivity, precision, specificity, F-score, area under the receiver operating characteristic (ROC) curve (AUC), and intersection over union (IoU) [43]–[45]. The overall accuracy measures the rate of correctly classified pixels; the sensitivity measures the rate of correctly classified WBC pixels and is also known as the true positive (TP) rate; the specificity measures the correctly classified non-WBC pixels and is also known as the true negative (TN) rate. Precision measures the positive prediction value (PPV) of the model, and the F-score measures the harmonic mean of both precision and sensitivity. Therefore, the F-score provides a single measure of both precision and sensitivity, which is particularly useful for imbalanced data classification problems. The AUC measures model performance using several thresholds. The similarity between a predicted object and its corresponding ground truth is measured using IoU, also known as the Jaccard index [43], and is commonly used in the field of object detection. Equations (8) – (14) were used to calculate the performance measures, as follows:

$$\text{Overall accuracy} = \frac{\mathbf{TP} + \mathbf{TN}}{\mathbf{TP} + \mathbf{FN} + \mathbf{FP} + \mathbf{TN}} \tag{8}$$

$$\text{Precision} = \frac{\mathbf{TP}}{\mathbf{TP} + \mathbf{FP}} \tag{9}$$

$$\text{Sensitivity} = \frac{\mathbf{TP}}{\mathbf{TP} + \mathbf{FN}} \tag{10}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \tag{11}$$

$$\text{F1-score} = 2 * \frac{\text{precision} * \text{sensitivity}}{\text{precision} + \text{sensitivity}} \tag{12}$$

$$\text{AUC} = \frac{\text{sensitively} + \text{precision}}{2} \tag{13}$$

$$IoU = \frac{TP}{FP + FN + TP} \tag{14}$$

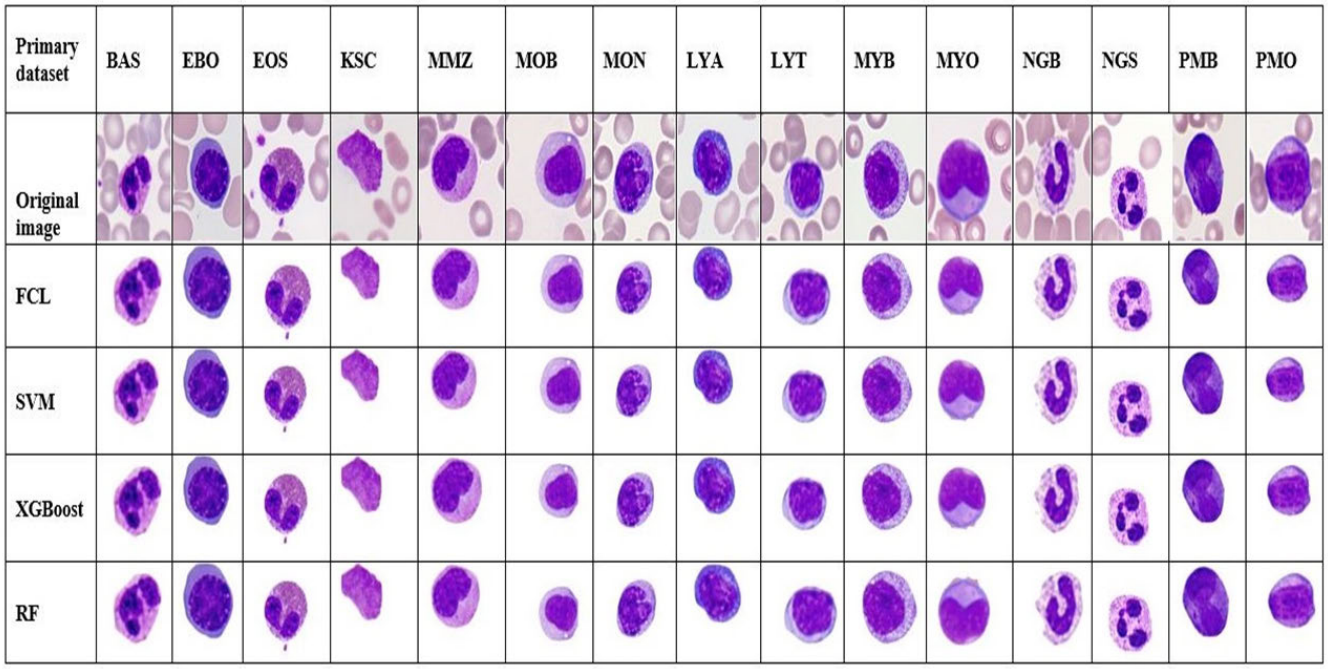

FIGURE 8. Results of WBCs segmentation using the proposed feature extraction method and SVM algorithms (primary dataset).

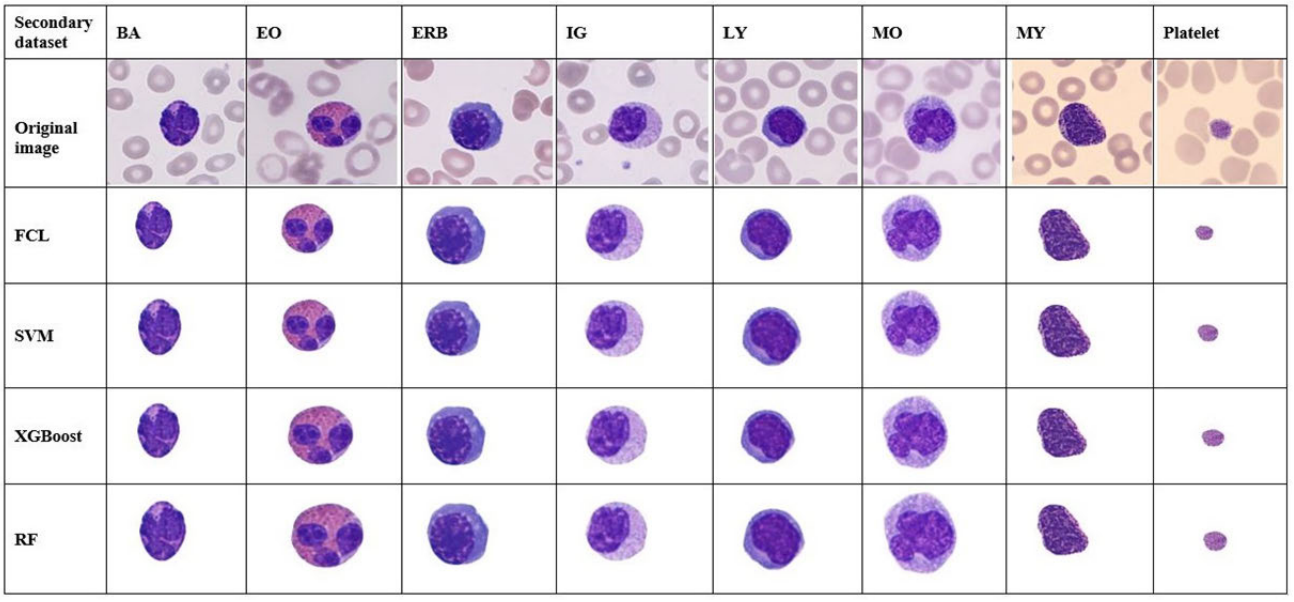

FIGURE 9. Results of different blood-cell segmentation using the proposed feature extraction method with the SVM algorithm (secondary dataset).

The model was developed and implemented using an Intel® Core™ i7-9750 h @ 2.60 GHz 192 with a 64-bit operating system, an x64-based processor, 16 GB of RAM, and an NVIDIA GeForce RTX 2070 with a max-design. The algorithm was written in Python using the Keras deep learning package and other image-processing packages to extract handcrafted features.

IV. RESULTS AND DISCUSSION

A set of 128 features was obtained, and each pixel was labeled as foreground (WBC) or background (other blood components). The dataset consisted of 3,192,550 pixels; 1,795,988 (56.3%) pixels were labeled as foreground and 1,396,562 (43.7%) pixels were labeled as background. The dataset was then divided into 80% and 20% for training and testing,

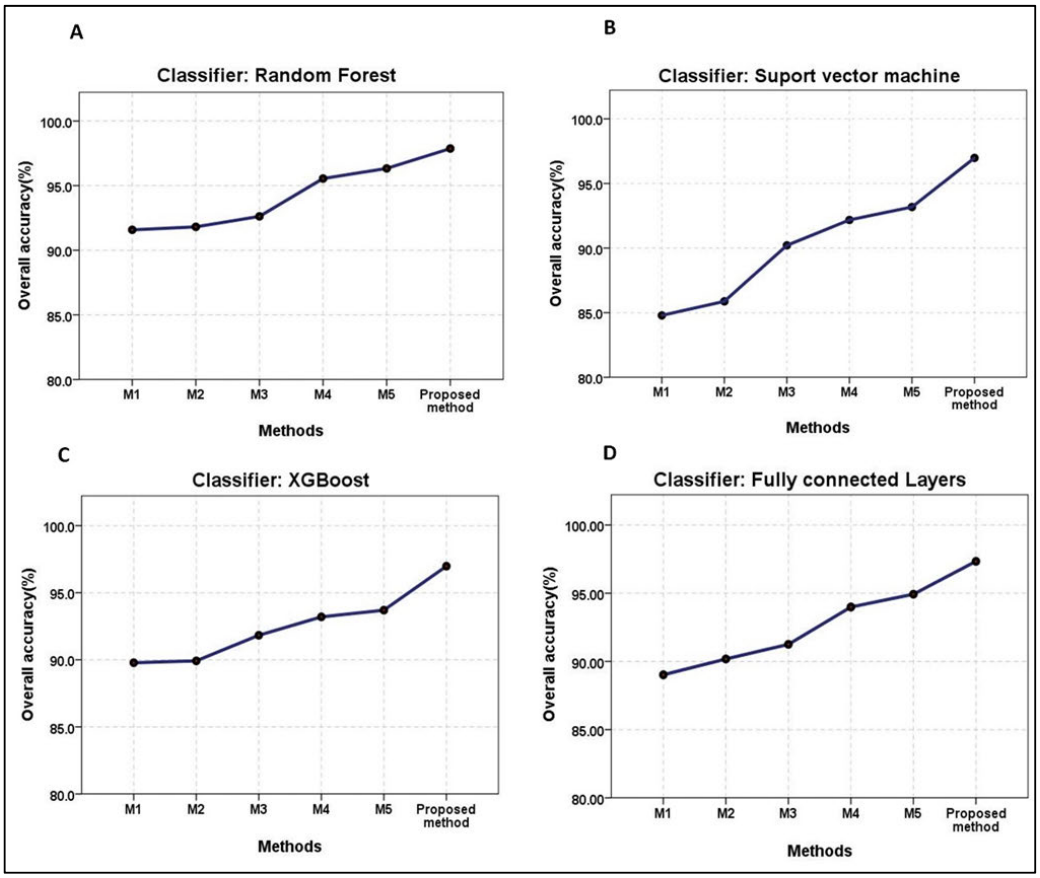

FIGURE 10. (A–D) Overall accuracy of the proposed method compared to other benchmark methods.

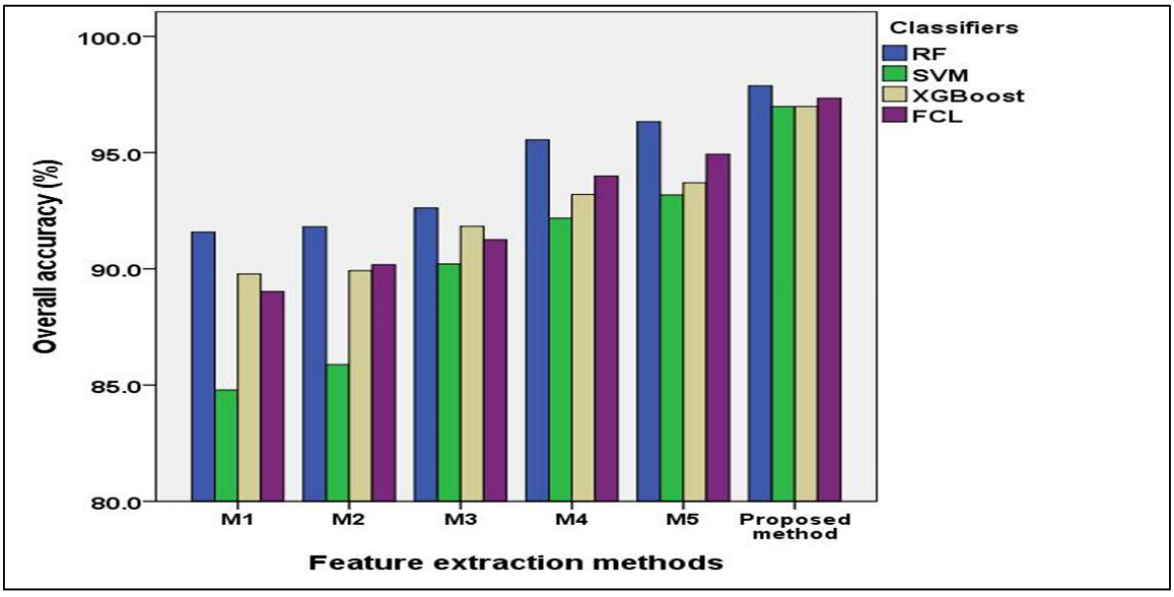

FIGURE 11. Overall accuracy of the proposed feature extraction methods compared to other feature extraction methods.

respectively. The following are the results of the four evaluation methods mentioned in Section III.

A. PRIMARY DATASET

The proposed feature extraction method achieved an overall accuracy of 97.34%, 97.57%, 97.15%, 97.39% using FCL, RF, SVM, and XGBoost, respectively. Table 1 shows that the proposed feature extraction method achieved comparable results for all classifiers. However, SVM showed more efficacy in terms of computation time (computation time = 38.7 s) and comparable overall accuracy using the same hardware facilities (see Figure 6).

TABLE 3. Benchmarking the proposed feature extraction method with other feature of extraction methods using a primary dataset.

	Type of classifier	NO. OF Features	Overall Accuracy	Sensitivity	Specificity	Precision	F-score	AUC	Jaccard index	Computation time (Secs)
M1	FCL	97	0.8902	0.9423	0.9061	0.8725	0.9061	0.8828	0.8283	2233.9
	RF	97	0.9158	0.9074	0.8901	0.9407	0.9237	0.9170	0.8583	1239.6
	SVM	97	0.8479	0.8623	0.8243	0.8664	0.8644	0.8445	0.7612	15.2
	XGBoost	97	0.8978	0.8675	0.8463	0.9463	0.9052	0.9020	0.8268	819.5
M2	FCL	98	0.9018	0.9398	0.9170	0.8914	0.9150	0.8964	0.8433	1389.0
	RF	98	0.9181	0.9108	0.8901	0.9416	0.9260	0.9191	0.8621	1284.8
	SVM	98	0.8588	0.8654	0.8311	0.8814	0.8733	0.8579	0.7750	17.0
	XGBoost	98	0.8992	0.8673	0.8466	0.9491	0.9063	0.9038	0.8288	825.0
M3	FCL	103	0.9125	0.9481	0.9286	0.9014	0.9242	0.9075	0.8590	2384.1
	RF	103	0.9262	0.9216	0.9026	0.9464	0.9338	0.9273	0.8756	1374.6
	SVM	103	0.9021	0.8783	0.8565	0.9436	0.9098	0.9054	0.9345	17.1
	XGBoost	103	0.9183	0.8990	0.8791	0.9529	0.9252	0.9210	0.8608	743.0
M4	FCL	107	0.9399	0.9389	0.9231	0.9534	0.9461	0.9400	0.8978	2873.6
	RF	107	0.9555	0.9522	0.9399	0.9681	0.9601	0.9560	0.9233	1253.6
	SVM	107	0.9218	0.9053	0.8858	0.9532	0.9286	0.9241	0.8668	17.4
	XGBoost	107	0.9320	0.9194	0.9016	0.9580	0.9383	0.9338	0.8838	732.7
M5	FCL	108	0.9493	0.9420	0.9279	0.9670	0.9543	0.9504	0.9127	3489.9
	RF	108	0.9633	0.9611	0.9509	0.9732	0.9671	0.9636	0.9364	1227.3
	SVM	108	0.9318	0.9159	0.8982	0.9609	0.9378	0.9340	0.8830	24.0
	XGBoost	108	0.9370	0.9188	0.9022	0.9684	94.3050	0.9402	0.8921	746.9
Proposed method	FCL	128	0.9734	0.9584	0.9489	0.9941	0.9759	0.9756	0.9530	997.5
	RF	128	0.9715	0.9597	0.9502	0.9892	0.9742	0.9731	0.9498	38.7
	SVM	128	0.9739	0.9611	0.9520	0.9921	0.9764	0.9757	0.9539	1153.5
	XGBoost	128	0.9757	0.9716	0.9642	0.9848	0.9782	0.9762	0.9574	3156.5

M1: feature bank of Gabor filters; M2: feature bank of Gabor filters and local binary pattern (LBP); M3: feature bank of Gabor, LBP, and edge detection filters; M4: feature bank of Gabor, LBP, edge detection filters, and K-means; M5: feature bank of Gabor filters, LBP, edge detection, Gaussian filters, and k-means.

B. SECONDARY DATASET

The secondary dataset consisted of 343752 and 175551 (51.1%) blood cells and 168201 (49.9%) non-blood cells, respectively. Using the secondary dataset, the proposed feature extraction method achieved overall accuracies of 94.92%, 95.47%, 96.41%, and 95.18% using FCL, RF, SVM, XGBoost, respectively. Table 2 shows that the proposed feature extraction method exhibited stability and produced

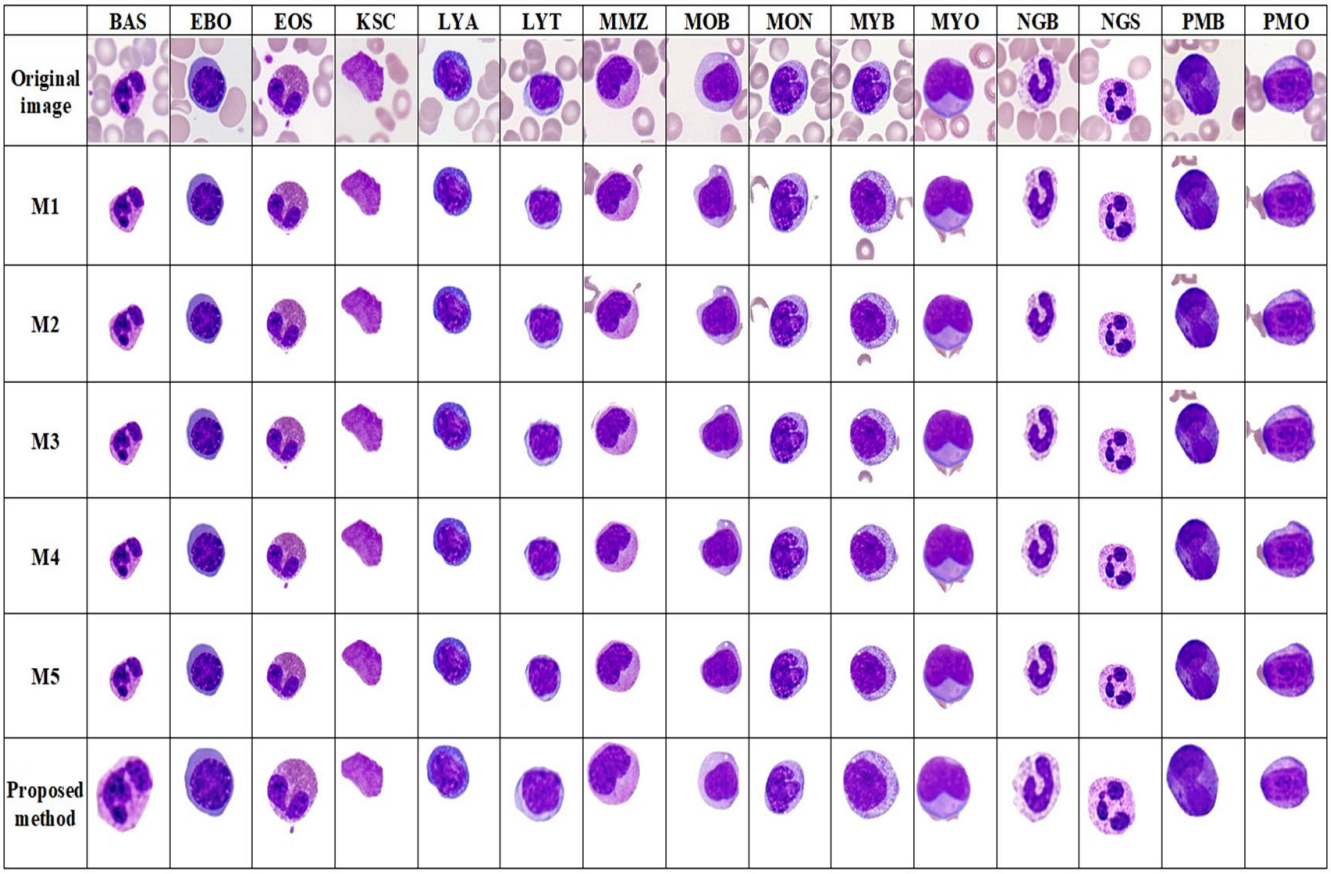

FIGURE 12. Results of 15 different types of WBC segmentation using the proposed segmentation method compared to other methods.

comparable results among the classifiers. However, the SVM achieved better performance compared to the other classifiers and mitigated overfitting compared to the other classifiers (see Figure 7).

Figures 8 and 9 show the results of applying the proposed feature extraction method to primary and secondary datasets, respectively. The proposed method was able to detect all types of blood cells present in the datasets and accurately detect platelet cells that were not present during training with the primary dataset.

V. BENCHMARKING THE PROPOSED FEATURE EXTRACTION METHOD WITH OTHER METHODS

The selected benchmarking methods were chosen based on many experiments, starting with conventional methods and ending with advanced deep learning methods. However, deep learning-based features using CNN achieved the highest accuracy compared with the other methods. The proposed feature extraction method was benchmarked using several other feature extraction methods. These methods include a feature bank of texture features that uses Gabor filters, local binary pattern (LBP), edge detection filters, K-means clusters, and Gaussian filters. The first method used a feature bank of Gabor texture filters (M1). The second method used Gabor filters and an LBP filter (M2). The third method uses Gabor filters, LBP features, and edge-detection features (M3). The fourth method applied a combination of Gabor filters, LBP features, edge detection features, and K-means clusters (M4). The fifth method applies Gabor filters, LBP, edge-detection features, K-means clusters, and a Gaussian filter (M5).

The benchmark feature extraction methods were evaluated using multiple classifiers, including FCL, RF, SVM, and XGBoost, based on the overall accuracy, sensitivity, specificity, precision, F-score, AUC, and Jaccard index, which were used to evaluate each classifier (Table 3).

As shown in Table 3, using only Gabor filters, the model achieved overall accuracies of 89.02, 91.58, 84.79, and 89.78% using FCL, RF, SVM, and XGBoost, respectively. However, adding the LBP texture feature did not markedly improve the performance, achieving 90.18%, 91.81%, 85.88%, and 89.92% overall accuracy using FCL, RF, SVM, and XGBoost, respectively. Similarly, the addition of edge detection filters, including the Canny, Robert, Sobel, Scharr, and Prewitt filters, did not markedly improve performance. The overall accuracies achieved using the FCL, RF, SVM, and XGBoost were 91.25%, 92.62%, 90.21%, and 91.83%, respectively. However, adding k-means clusters as a

new feature improved the performance to 93.99%, 95.55%, 92.18%, and 93.20% when FCL, RF, SVM, and XGBoost were used, respectively. The addition of a Gaussian filter increased the overall accuracy to 94.93%, 96.33%, 93.18%, and 93.7% using FCL, RF, SVM, and XGBoost, respectively. Finally, the results obtained by the benchmark methods were compared with the results obtained using the proposed feature extraction method. The proposed method achieved the best results compared with all feature extraction methods, with overall accuracies of 97.34%, 97.15%, 97.39%, and 97.57% using FCL, RF, SVM, and XGBoost, respectively. Figure 10 (A–D) compare the different feature extraction methods and the proposed method using FCL, RF, SVM, and XGBoost. Figure 10 shows that the proposed feature extraction method consistently improves performance compared to the other benchmark methods. Figure 11 summarizes the performance of the proposed feature extraction method compared to other benchmark methods using FCL, RF, SVM, and XGBoost.

Figure 12 shows the results of applying the benchmark feature extraction methods for WBC recognition compared with those of the proposed method. Figure 12 shows that the application of the feature bank using Gabor features (M1) achieved good overall performance but exhibited poor performance in MOB and NGB detection. The addition of the LBP texture feature (M2) did not improve the performance and exhibited an inferior performance in NGB recognition. The addition of edge detection features (M3) marginally improved the NGB detection performance but exhibited worse MOB detection performance. The addition of k-means (M4) to the feature bank improved both the NGB and MOB performance. Adding a Gaussian filter (M5) produced inferior results for both the MOB and NGB. However, the proposed feather extraction method exhibited the best performance in recognizing different types of WBCs, including MOBs and NGBs.

This study highlights one of the primary limitations of handcrafted features: they are strongly associated with the type of selected features that require extensive domain knowledge and experience [17]. They are also difficult to automate, require more training time (see Table 1), cannot be generalized, and cannot be used to detect other types of blood cells that are not used during training. Figures 10 and 11 show that handcrafted features (M1 to M5) produced inferior results compared with the proposed methods, which might be because only one type of WBC was used for model training. Using handcrafted features, Abdurrazzaq *et al.* segmented the entire WBC only in cases where WBCs have high WBC intensity colors, such as basophils and monocytes; however, in cases where WBCs were presented with light intensity colors, their method was able to segment the WBC nucleus and failed in cytoplasm segmentation [28].

The proposed feature extraction method produced the best results compared to the other comparative feature extraction methods. As shown in Table 3, despite the complexity and heterogeneity of WBCs found in this study, the use of CMYK-moment localization combined with deep-learning-based feature fusion improved the detection of WBCs with different shapes, structures, colors, illumination, and staining variabilities. In this study, features were extracted, and the model was trained using only one morphological class (BAS). However, the model was able to classify all the other 14 different types of WBC morphological classes with a high level of accuracy. The proposed feature extraction method also achieved a good performance when identifying different types of blood cells using an external dataset of different types of normal WBCs and platelets. Thus, the results demonstrate that the proposed method is generalizable and invariant to shape, structure, color, and other differential factors. However, these characteristics may be attributed to the use of an ROI, which helps establish a context-free learning environment by extracting only features that are related to WBCs, while eliminating other irrelevant features. Using 2D convolution layers improves learning by extracting more complex features compared with other conventional methods, such as Gabor filters, which can extract only simple features. Feature fusion of different kernel sizes and stacked layers of different depth levels helps to increase the generalizability of the proposed model.

Figure 11 shows that most models fail to detect WBCs in some situations, particularly for cells with low contrast light color intensity [28], and cells that showed similarities of cytoplasm and background color. NGB and MOB are the most difficult WBCs to detect, which may be due to their light cytoplasm colors and smooth cytoplasm textures, which make it difficult to discriminate between the cytoplasm and image background. However, despite these difficulties, the proposed feature extraction method achieved excellent performance when detecting all types of WBCs, including MOB and NGB.

In this study, a hybrid method of CMYK-moment localization and a deep CNN-based feature fusion model were used for feature extraction. The model consists of two parallel layers: 1) a shallow convolution layer that extracts pointwise WBCs without losing pixel-wise information originality and 2) a deep stacked coevolution of two layers that extracts localized WBC features. The first layer was used to extract simple features such as edges, and the second layer was used to extract more complex WBC patterns. In this study, we hypothesized that using multiple convolutional operations might lose original pixel information; therefore, the first layer was applied as a shallow layer with a kernel size of one.

VI. CONCLUSION

Using a hybrid model of conventional image processing and deep learning methods using feature fusion, we developed a generalized WBC feature extraction method. Using ROIs, as opposed to the entire image, helped reduce noise and omit irrelevant features, thereby aiding the generation of a stabilized and context-free learning environment to extract the WBC features. 2D convolution layers also create strong feature extraction methods for WBCs because they are invariant to morphological structure, color, and staining variations. Feature fusion of single and stacked convolutional layers of different kernel sizes and depth levels helped establish

a generalized feature extraction framework. The proposed method detected 15 types of WBCs, including both normal and AML cancer cells. The proposed method also harnessed the benefits of conventional image processing techniques for extracting ROIs and the strengths of deep learning methods for feature extraction. In the future, the proposed model could be further improved by hybridization and parallelization methods to reduce the computation time and increase model accuracy.

ACKNOWLEDGMENT

The authors would like to thank Dr. Ameera Gaafar and Dr. Taqwa Alhaj for critically reviewing this manuscript.

REFERENCES

[1] K. AL-Dulaimi, J. Banks, K. Nugyen, A. Al-Sabaawi, I. Tomeo-Reyes, and V. Chandran, "Segmentation of white blood cell, nucleus and cytoplasm in digital haematology microscope images: A Review–challenges, current and future potential techniques," *IEEE Rev. Biomed. Eng.*, vol. 14, pp. 290–306, 2021.

[2] L. Bigorra, A. Merino, S. Alférez, and J. Rodellar, "Feature analysis and automatic identification of leukemic lineage blast cells and reactive lymphoid cells from peripheral blood cell images," *J. Clin. Lab. Anal.*, vol. 31, no. 2, Mar. 2017, Art. no. e22024.

[3] J. W. Choi, Y. Ku, B. W. Yoo, J.-A. Kim, D. S. Lee, Y. J. Chai, H.-J. Kong, and H. C. Kim, "White blood cell differential count of maturation stages in bone marrow smear using dual-stage convolutional neural networks," *PLoS ONE*, vol. 12, no. 12, Dec. 2017, Art. no. e0189259.

[4] R. B. Hegde, K. Prasad, H. Hebbar, and B. M. K. Singh, "Comparison of traditional image processing and deep learning approaches for classification of white blood cells in peripheral blood smear images," *Biocybernetics Biomed. Eng.*, vol. 39, no. 2, pp. 382–392, Apr. 2019.

[5] A. Tareef, Y. Song, D. Feng, M. Chen, and W. Cai, "Automated multi-stage segmentation of white blood cells via optimizing color processing," in *Proc. IEEE 14th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2017, pp. 565–568.

[6] N. Baghel, U. Verma, and K. K. Nagwanshi, "WBCs-Net: Type identification of white blood cells using convolutional neural network," *Multimedia Tools Appl.*, vol. 162, pp. 1–17, Sep. 2021.

[7] R. M. Roy and A. P. M., "Segmentation of leukocyte by semantic segmentation model: A deep learning approach," *Biomed. Signal Process. Control*, vol. 65, Mar. 2021, Art. no. 102385.

[8] H. K. Walker, W. D. Hall, and J. W. Hurst, *Clinical Methods: The History, Physical, and Laboratory Examinations*, Boston, MA, USA: Butterworths, 1990.

[9] K. A. K. Al-Dulaimi, J. Banks, V. Chandran, I. Tomeo-Reyes, and K. N. Thanh, "Classification of white blood cell types from microscope images: Techniques and challenges," Tech. Rep., 2018.

[10] R. I. Agustin, A. Arif, and U. Sukorini, "Classification of immature white blood cells in acute lymphoblastic leukemia l1 using neural networks particle swarm optimization," *Neural Comput. Appl.*, vol. 33, no. 17, pp. 10869–10880, Sep. 2021.

[11] Rakhmadi, "Connected component labeling using components neighbors-scan labeling approach," *J. Comput. Sci.*, vol. 6, no. 10, pp. 1099–1107, Oct. 2010.

[12] F. Cao, M. Cai, J. Chu, J. Zhao, and Z. Zhou, "A novel segmentation algorithm for nucleus in white blood cells based on low-rank representation," *Neural Comput. Appl.*, vol. 28, no. S1, pp. 503–511, Dec. 2017.

[13] S. H. Shirazi, A. I. Umar, N. Haq, S. Naz, M. I. Razzak, and A. Zaib, "Extreme learning machine based microscopic red blood cells classification," *Cluster Comput.*, vol. 21, no. 1, pp. 691–701, Mar. 2018.

[14] H. A. Elsalamony, "Detection of anaemia disease in human red blood cells using cell signature, neural networks and SVM," *Multimedia Tools Appl.*, vol. 77, no. 12, pp. 15047–15074, 2018.

[15] M. A. Parab and N. D. Mehendale, "Red blood cell classification using image processing and CNN," *Social Netw. Comput. Sci.*, vol. 2, no. 2, pp. 1–10, Apr. 2021.

[16] D. López-Puigdollers, V. Javier Traver, and F. Pla, "Recognizing white blood cells with local image descriptors," *Expert Syst. Appl.*, vol. 115, pp. 695–708, Jan. 2019.

[17] T. Pansombut, S. Wikaisuksakul, K. Khongkraphan, and A. Phon-on, "Convolutional neural networks for recognition of lymphoblast cell images," *Comput. Intell. Neurosci.*, vol. 2019, pp. 1–12, Jun. 2019.

[18] D. T. Nguyen, T. D. Pham, N. R. Baek, and K. R. Park, "Combining deep and handcrafted image features for presentation attack detection in face recognition systems using visible-light camera sensors," *Sensors*, vol. 18, no. 3, p. 699, Feb. 2018.

[19] Y. Lu, X. Qin, H. Fan, T. Lai, and Z. Li, "WBC-Net: A white blood cell segmentation network based on UNet++ and Resnet," *Appl. Soft Comput.*, vol. 101, Mar. 2021, Art. no. 107006.

[20] S. Saleem, J. Amin, M. Sharif, M. A. Anjum, M. Iqbal, and S.-H. Wang, "A deep network designed for segmentation and classification of leukemia using fusion of the transfer learning models," *Complex Intell. Syst.*, Jul. 2021.

[21] R. B. Hegde, K. Prasad, H. Hebbar, and B. M. K. Singh, "Feature extraction using traditional image processing and convolutional neural network methods to classify white blood cells: A study," *Australas. Phys. Eng. Sci. Med.*, vol. 42, no. 2, pp. 627–638, Jun. 2019.

[22] S. Purwar, R. K. Tripathi, R. Ranjan, and R. Saxena, "Detection of microcytic hypochromia using cbc and blood film features extracted from convolution neural network by different classifiers," *Multimedia Tools Appl.*, vol. 79, nos. 7–8, pp. 4573–4595, Feb. 2020.

[23] H. Yan, X. Mao, X. Yang, Y. Xia, C. Wang, J. Wang, R. Xia, X. Xu, Z. Wang, Z. Li, X. Zhao, Y. Li, G. Liu, L. He, Z. Wang, Z. Wang, Z. Li, W. Cai, H. Shen, and H. Chang, "Development and validation of an unsupervised feature learning system for leukocyte characterization and classification: A multi-hospital study," *Int. J. Comput. Vis.*, vol. 129, no. 6, pp. 1837–1856, Jun. 2021.

[24] P. P. Banik, R. Saha, and K.-D. Kim, "An automatic nucleus segmentation and CNN model based classification method of white blood cell," *Expert Syst. Appl.*, vol. 149, Jul. 2020, Art. no. 113211.

[25] C. Matek, S. Krappe, C. Münzenmayer, T. Haferlach, and C. Marr, "Highly accurate differentiation of bone marrow cell morphologies using deep neural networks on a large image data set," *Blood*, vol. 138, no. 20, pp. 1917–1927, Nov. 2021.

[26] C. Matek, "A single-cell morphological dataset of leukocytes from AML patients and non-malignant controls 2019: The cancer imaging archive," Tech. Rep., 2019, doi: 10.7937/tcia.2019.36f5o9ld.

[27] Y. Lu, X. Qin, H. Fan, T. Lai, and Z. Li, "WBC-Net: A white blood cell segmentation network based on UNet++ and Resnet," *Appl. Soft Comput.*, vol. 101, Mar. 2021, Art. no. 107006.

[28] A. Abdurrazzaq, A. K. Junoh, Z. Yahya, and I. Mohd, "New white blood cell detection technique by using singular value decomposition concept," *Multimedia Tools Appl.*, vol. 80, no. 3, pp. 4627–4638, Jan. 2021.

[29] N. Khomairoh, R. Sigit, T. Harsono, Y. Hernaningsih, and A. Anwar, "Segmentation system of acute myeloid leukemia (AML) subtypes on microscopic blood smear image," in *Proc. Int. Electron. Symp. (IES)*, Sep. 2020, pp. 565–570.

[30] V. J. Ramya and S. Lakshmi, "Acute myelogenous leukemia detection using optimal neural network based on fractional black-widow model," *Signal, Image Video Process.*, vol. 16, no. 1, pp. 229–238, Feb. 2022.

[31] A. E. Rad, M. S. M. Rahim, H. Kolivand, and I. B. M. Amin, "Morphological region-based initial contour algorithm for level set methods in image segmentation," *Multimedia Tools Appl.*, vol. 76, no. 2, pp. 2185–2201, Jan. 2017.

[32] C. D. Ruberto, A. Loddo, and L. Putzu. "A multiple classifier learning by sampling system for white blood cells segmentation," in *Computer Analysis of Images and Patterns*. Cham, Switzerland: Springer, 2015.

[33] C. Di Ruberto, A. Loddo, and L. Putzu, "A leucocytes count system from blood smear images," *Mach. Vis. Appl.*, vol. 27, no. 8, pp. 1151–1160, 2016.

[34] A. Acevedo, A. Merino, S. Alférez, Á. Molina, L. Boldú, and J. Rodellar, "A dataset of microscopic peripheral blood cell images for development of automatic recognition systems," *Data Brief*, vol. 30, Jun. 2020, Art. no. 105474.

[35] V. Andrearczyk and P. F. Whelan, "Chapter 4—Deep learning in texture analysis and its application to tissue image classification, in *Biomedical Texture Analysis*, A. Depeursinge, O. S. Al-Kadi, and J.R. Mitchell, Eds. New York, NY, USA: Academic, 2017, pp. 95–129.

[36] C. Karabağ, J. Verhoeven, N. R. Miller, and C. C. Reyes-Aldasoro, "Texture segmentation: An objective comparison between five traditional algorithms and a deep-learning U-Net architecture," *Appl. Sci.*, vol. 9, no. 18, p. 3900, Sep. 2019.

[37] A. Zhang, "Chapter 6: Convolutional neural networks, section 6.4: Multiple input and multiple output channels," in *Dive into Deep Learning*. 2021.
[38] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 56, pp. 1929–1958, 2014.
[39] F. Rosenblatt, "Principles of neurodynamics. Perceptrons and the theory of brain mechanisms," Cornell Aeronautical Lab Inc, Buffalo, NY, USA, Tech. Rep., 1961.
[40] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
[41] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.
[42] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2016, pp. 785–794.
[43] H. Rezatofighi, N. Tsoi, J. Gwak, A. Sadeghian, I. Reid, and S. Savarese, "Generalized intersection over union: A metric and a loss for bounding box regression," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 658–666.
[44] M. Hossin and M. N. Sulaiman, "A review on evaluation metrics for data classification evaluations," *Int. J. Data Mining Knowl. Manage. Process*, vol. 5, no. 2, pp. 1–11, Mar. 2015.
[45] O. O. Koyejo, N. Natarajan, P. K. Ravikumar, and I. S. Dhillon, "Consistent binary classification with generalized performance metrics," in *Proc. NIPS*, 2014, pp. 1–9.

TUSNEEM AHMED M. ELHASSAN received the B.Sc. and M.Sc. degrees in computer science and (statistics and computer science) from the Faculty of Mathematical Science, University of Khartoum, in 2002 and 2007, respectively. She is currently pursuing the Ph.D. degree with the University of Technology Malaysia (UTM). Her Ph.D. research focuses on computer vision using deep learning. She has been a Data Scientist working at Cancer Research at King Faisal Specialist Hospital and Research Centre, since 2007. She has over 13 years of accomplishments in medical research with special focus on modern statistical techniques, including data modeling and survival analysis. She is the author and coauthor of more than 50 publications and 26 proceedings.

MOHD SHAFRY MOHD RAHIM was born in Malaysia. He received the bachelor's degree in computer science and the M.S. and Ph.D. degrees in computing from UTM, Malaysia. He is currently a Professor at UTM, where he leads a research group comprising of a faculty members and a graduate students. He is also the Chair of undergraduate students at UTM. He has published research articles in several renowned international journals and conference papers and has been a keynote speaker in many international conferences.

TAN TIAN SWEE received the M.S. and Ph.D. degrees from UTM, in 2004 and 2008, respectively. He has spearheaded numerous projects and acquired prestigious grants from various sources. One of his notable milestones is his collaboration between the UTM and IJN. He is currently a Senior Lecturer with the Faculty of Biomedical Engineering, UTM, and works as a Program Coordinator with the Faculty of Biomedical Engineering. He has published numerous high-impact-factor journals, thus establishing his expertise in this domain. His research interest includes digital signal processing. He is also a member of the Medical Device and Technology Group and Frontier Materials Research Alliances.

SITI ZAITON MOHD HASHIM received the B.Sc. degree in computer science from the University of Hartford, USA, the M.Sc. degree in computing from the University of Bradford, U.K., and the Ph.D. degree in soft computing from The University of Sheffield, U.K. She is currently a Professor with the Department of Data Science, Universiti Malaysia Kelantan (UMK). She previously held several administrative posts with the School of Computing, UTM, Johor, from 2007 to 2018, including the Head of the Department, the Deputy Dean of Postgraduate Studies, and the Deputy Dean of Academics. She was also the Director of the Big Data Centre (Centre of Excellence), UTM, from February 2019 to February 2020. She has supervised and co-supervised more than 20 master's and 20 Ph.D. students. She has authored and coauthored more than 150 publications: 80 proceedings and 57 journals, with 19 H-indices and more than 1000 citations. Her research interests include soft computing, machine learning, and data analytics.

MAHMOUD ALJURF received the M.D. degree in 1985, the master's degree in public health (MPH) from Johns Hopkins University, and FRCPath, U.K. He completed his residency training in internal medicine with Brigham and Women's Hospital, Harvard Medical School, Boston, and a Hematology and Oncology Fellowship with the Stanford University Medical Center. He is an American board certified in internal medicine, medical oncology, hematology, medical management, and quality assurance. He is currently the Director of the Adult Hematology/Oncology/Stem Cell Transplant Program and the Deputy Director of the Oncology Centre, King Faisal Specialist Hospital and Research Centre, Riyadh, Saudi Arabia. In addition, he works as the Scientific Director of the Eastern Mediterranean Blood and Marrow Transplantation Group (EMBMT) and as the Editor-in-Chief for Elsevier journals, *Hematology/Oncology*, and *Stem Cell Therapy*. He received the American College of Physicians Master Ship Award, in 2012.