Improved White Blood Cells Classification Based on Pre-trained Deep Learning Models

Ensaf H. Mohamed, Wessam H. El-Behaidy[*], Ghada Khoriba, and Jie Li

Abstract—Leukocytes, or white blood cells (WBCs), are microscopic organisms that fight against infectious disease, bacteria, viruses and others. The manual method to classify and count WBCs is tedious, time-consuming and may have inaccurate results, whereas the automated methods are costly. This research aims to automatically identify and classify WBCs in a microscopic image into four types with higher accuracy. BCCD is the used dataset in this study, which is a scaled-down blood cell detection dataset. BCCD is firstly preprocessed by passing through various processes such as segmentation and augmentation; then, it is passed to the proposed model. Our model combines the advantage of deep models in automatically extracting features with the higher classification accuracy of traditional machine learning classifiers. The proposed model consists of two main stages: a shallow tuning pre-trained model and a traditional machine learning classifier on top of it. In this study, ten different pre-trained models with six types of machine learning are used. Moreover, the fully connected network (FCN) of pre-trained models is used as a baseline classifier for comparison. The evaluation process shows that the hybrid of MobileNet-224 as a feature extractor and logistic regression as classifier has a higher rank-1 accuracy of 97.03%. Besides, the proposed hybrid model outperformed the baseline FCN by 25.78% on average.

Index Terms—Deep learning, feature extraction, classification, white blood cells (WBCs).

I. INTRODUCTION

WHite blood cells (WBCs), or leukocyte cells, serve as the armed forces of our body against infectious disease. WBCs have five consequential types that could be differentiated according to their shape and size, the presence of granules in their cytoplasm, and the number of lobes in their nucleus. WBCs can be categorized as cells that contain granules, including neutrophils, basophils and eosinophils, and cells without granules, including monocytes and lymphocytes [1], [2].

The density of WBCs in the bloodstream contributes an insight into the state of our immune system and any possible risks. A significant change in the WBC count comparable
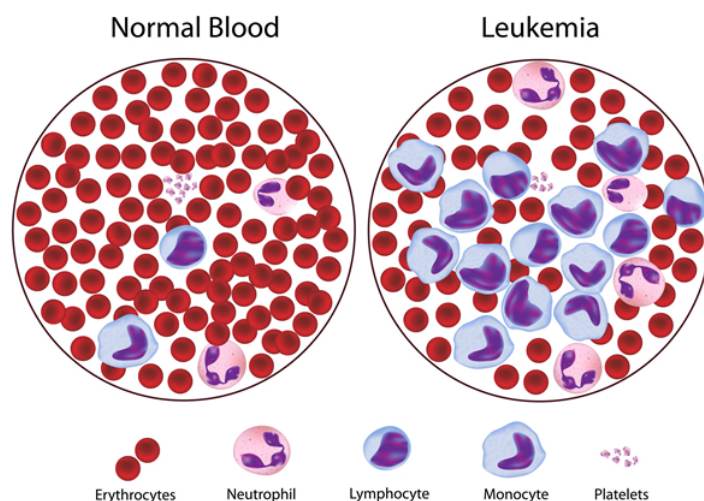
Fig. 1. WBC analysis example shows the variation in the number of white blood cells between normal blood and blood with leukemia [3].

to the baseline is commonly a sign that the human body is currently being affected by an antigen. Furthermore, a change in a particular type of WBC usually relates to a particular type of antigen. For example, the malfunctioning of the immune system could be indicated in leukemia patients by a higher level of lymphocytes in the bloodstream, as shown in Fig. 1. Moreover, an increase in the eosinophil counts is a sign of people fighting allergies as these WBCs are key to fighting allergens. Consequently, the count of WBCs in the bloodstream can provide a person with a compelling quantitative picture of his/her health.

Therefore, considerable research has been done on the classification and counting of WBCs due to their significance for the medical analysis field. A typical image classification system includes multiple phases: image preprocessing, segmentation, feature extraction, and finally, the classifier.

The feature extraction phase is one of the significant steps that strongly affect the performance of the classification system. The extracted features from an image could be global or local [4]. The global features represent the whole aspects of an image by one-dimensional vectors such as color, texture, or shape. Moreover, local binary patterns (LBP), histogram of oriented gradients (HOG) and color histograms could be used as global feature descriptors. Conversely, the local features only extract features from a set of regions of interest (ROI). Scale-invariant feature transform (SIFT), Speeded-Up Robust
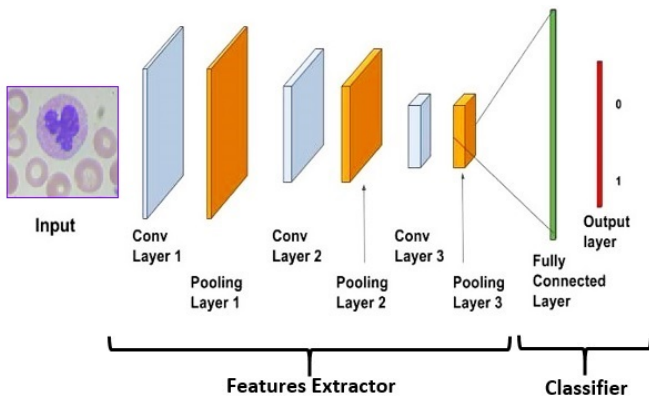
Fig. 2. Deep models sections: The convolutional layers act as feature extractors, and the fully connected layers act as classifiers [7].

Features (SURF), and Oriented FAST and Rotated BRIEF (ORB) could be used as local feature descriptors.

The commonly extracted features, for the WBCs, can be categorized into geometrical, textural, intensity and color based features [5]. The geometrical features are related to the shape of the cell and nucleus. They may include area, perimeter, number of nuclei, orientation and compactness.

Over the last few decades, the extraction of features was hand-crafted, requiring a high level of expertise and taking a lot of time and effort. Nowadays, deep learning models use end-to-end feature extraction and classification, which outperform the hand-crafted methods. Instead of depending on domain expertise to extract features, deep neural networks extract these features automatically from images in hierarchical-based techniques. Lower layers learn low-level features such as edges and corners, whereas the middle layers learn features like color and shape. Higher layers learn high-level features describing the object in the image [6]. These deep models consist of convolutional layers that act as feature extractors and fully connected layers that act as classifiers, as shown in Fig. 2.

The main contributions of the paper can be summarized as the following:

- A two-stage hybrid classification system is proposed. The first stage uses a pre-trained model to produce the feature vector to avoid the artifacts of manual feature extraction. This feature vector is fed to the next stage, which consists of a traditional classifier to enhance the classification accuracy.
- We have studied the performance of the proposed two-stage model by varying ten pre-trained models as feature extractors in the first stage, and six traditional classifiers in the second stage. The results of different classifiers are compared to a baseline classifier (i.e., pre-trained convolutional neural network with a fully connected layer).
- The proposed hybrid model always outperforms, on average, the classical pre-trained models with a fully connected layer with four classifiers: logistic regression, random forest, k-nearest neighbors (KNN) and linear discriminant analysis (LDA).

This paper is organized as follows: Section II reviews some of the recent related research that has been done on WBCs with different datasets and medical images. The proposed approach and dataset are introduced in Section III. Section IV states all experiments and results that have been done on WBC dataset and the explanation and discussion of these results. The conclusion of the paper can be found in Section V.

II. RELATED WORK

Many related works have been conducted over the decades to classify WBCs. A wide variety of related work has been considered. Firstly, the traditional methods for extracting features with machine learning techniques are introduced. Then, the new attempts with deep neural models are clarified.

Yampri et al. [8] extracted the features by combining eigenvalue and parametric feature detection. They reached 92% in their classification using 50 images for testing. Falcon et al. [9] extracted from manually segmented nucleus the shape features (contour-based and region-based). The first descriptor concentrates on the boundary lines of the nucleus, whereas the other considers the whole area of the segmented nucleus. Five classifiers – multi-layer perceptron, pair-wise support vector machine (SVM), KNN, PART, and C4.5 with five-fold cross-validation – were used. All classifiers achieved a classification rate of over 96%. Habibzadeh et al. [10] segmented low-resolution images, extracted three different sets of features, and applied SVM as a classifier. The first set is composed of a feature vector of shape, intensity and texture features. The second set is developed by applying the dual-tree complex wavelet transform (DT-CWT) to have robust and invariant characteristics of the segmented image. The third set builds a lower-dimensional feature vector by applying the kernel principal component analysis (K-PCA) to the feature vector of the first set. The best classification rate was achieved by the second set, i.e., DT-CWT, with SVM with 84% classification rate. Su et al. [5] extracted geometrical, color and local-directional-pattern (LDP)-based texture features from the segmented cell. These features were fed into traditional neural networks. They reached 99.11% overall correctness over 450 images for testing. Gautam et al. [11] extracted only the geometric shape-based features from the segmented images and applied a classification rule to classify the type of WBC. They reached 73% overall performance on a set of 63 images for testing. Prinyakupt et al. [12] used a greedy search algorithm called sequential forward selection to extract several features from the segmented nucleus and cytoplasm regions. They reached 98% and 94% for classification using linear and naive Bayes (NB) classifiers, respectively. Ravikumar et al. [13] classified the WBCs using relevance vector machine (RVM) with a testing efficiency of 91%. They applied this method on three feature vectors based on 62 images for training and 23 images for testing. Ravikumar [14] extended his work and used the fast relevance vector machine (Fast-RVM) to detect the WBCs. This technique reduces the effects of illumination and staining effectively. Furthermore, it reduces the computational time. Sarrafzadeh et al. [15] worked on seven categories of texture features with LDA, KNN and NB

classifiers. The best classification rate was 85.53%. It was achieved with a combination of RICLBP features for the LDA classifier.

Nowadays, deep learning models obtain promising results in classifying and recognizing objects in the medical image field. Fu *et al.* [16] applied a multi-scale and multi-level convolutional neural network (CNN) with a side-output layer to learn a rich hierarchical design and utilized a conditional random field (CRF) to model the long-range interactions between pixels. Datasets provide two manual segmentations generated by two different experts for each image. Their proposed technique accuracy reached 95%. Shahin *et al.* [17] proposed an identification system for WBCs based on deep CNN. Two approaches based on transfer learning were used. Transfer learning based on deep activation features and fine-tuning of existing deep networks were combined. Two thousand five hundred fifty-one images from three different public WBC datasets were used. The dataset contains five healthy WBC types. The best overall accuracy achieved was 96.1%. Rajaraman *et al.* [18] used deep learning techniques to diagnose malaria in blood cell images. The researchers used the CNN-based deep learning model as a feature extractor. To improve disease screening, the researchers classified uninfected and parasitized cells. Experimentally, the optimal model layers for feature extraction were determined from the underlying data. They prove that the use of pre-trained CNNs is an auspicious tool for feature extraction. The researchers conclude that the best result was 95.7% accuracy using ResNet-50. Mehdi *et al.* [19] classify WBCs into four primary types – neutrophils, eosinophils, lymphocytes and monocytes – by consecutive deep learning framework. Using ResNet V1 50, their framework detects, on average, 100%, whereas alternative ResNet V1 152 and ResNet got promising results with 99.84% and 99.46% accuracy rate, respectively. Their experiment used 3,000 epochs to fine-tune all layers, trained on 11,200 samples, and evaluated 1,244 WBCs. Liang *et al.* [20] propose a framework that combines the CNN with the recurrent neural network (RNN) to deeply understand the image content and learn the structured features of images. The best performance of this framework reaches 90.97% when combined with Xception and long short-term memory (LSTM).

III. Materials and Methods

In this section, the dataset and the proposed model configuration are introduced.

A. Dataset

The dataset used in this paper is BCCD, which is publicly offered by Kaggle [21]. It contains 12,500 images with dimensions of 320x240 pixels each. The images are classified into only four classes: eosinophil, lymphocyte, monocyte and neutrophil. Each class has approximately 3,120 images –2,500 for training, and 620 for testing.

B. Proposed Model Configuration

Fig. 3 shows the proposed model architecture. It consists of four main building blocks: image preprocessing, feature extraction, classification and performance evaluation.

Firstly, the WBC images are segmented as in Sec. III-B1. Then, the segmented images are passed to our hybrid model. The *proposed model* consists of two stages: feature extraction using a pre-trained model as in Sec. III-B2 and, on top of it, the classification as in Sec. III-B3. Finally, the results are evaluated using multiple metrics as in Sec. III-B4.

1) Image Preprocessing: The used WBC images are pigmented with a distinct color to segment them from other components. Thresholding is the simplest technique of segmentation, which is computationally inexpensive and fast [22]. Firstly, WBCs are located using global thresholding technique. Then, WBCs are centered within the cropped 200x200 images, as shown in Fig. 4. The correctly segmented images are then split into 90% for training and 10% for testing before they are passed to the next step.

2) Features Extraction Using Pre-trained Model: The complex feature extraction process needs domain expertise. In this paper, pre-trained models are used to learn these features in a hierarchical structure. However, instead of building and training these models from scratch, the pre-trained weights trained on the ImageNet dataset are initially used, but with all layers left unfrozen in the training of our models since medical images need different weights. This method's effectiveness is already demonstrated by [23] in medical image analysis.

The last fully connected layer of each pre-trained model is removed to reach the feature vector, then the rest of the model is treated as a fixed feature extractor from the pre-trained model for the new dataset [24]. These extracted features will be used as input for the next stage in our hybrid model.

In this paper, *ten pre-trained deep learning models* are independently used as feature extractors on correctly segmented images. *Eight main models*: VGG-16, VGG-19 [25], ResNet-50 [26], DenseNet-121, DenseNet-169 [27], Inception-V3 [28], Inception-ResNet-V2 [29] and Xception [30]. *Two recent models with lightweight architecture*: 1.0 MobileNet-224 [31] and Mobile NASNet-A [32].

3) Classification: The previously extracted feature vector from each pre-trained model is then passed to one of these classifiers: traditional classifiers and fully connected network (FCN) classifiers.

a) Traditional classifiers: Six traditional machine learning classifiers have been chosen to cover popular algorithms of classification [33] and that exist in scikit-learn tool [34] for machine learning in Python. Logistic regression (LR), and LDA are used as linear classifiers. KNN, NB and decision tree (DT) are used as nonlinear classifiers. Finally, random forest is used as an ensemble classifier. The default parameters are used for the six classifiers. LR has L2 norm penalty, 1.0 for inverse regularization strength (C), L-BFGS optimizer for multiclass problems, and 0.0001 tolerance (tol) for stopping criteria. LDA has a singular value decomposition (svd) solver, which is recommended for data with a large number of features and 0.0001 (tol). KNN has 5 neighbors to use, 'uniform' weights, 30 leaf size, and the distance 'Minkowski' metric with p = 2, which is equivalent to the standard Euclidean metric. The Gaussian NB is used with the 1e-9 portion, of the
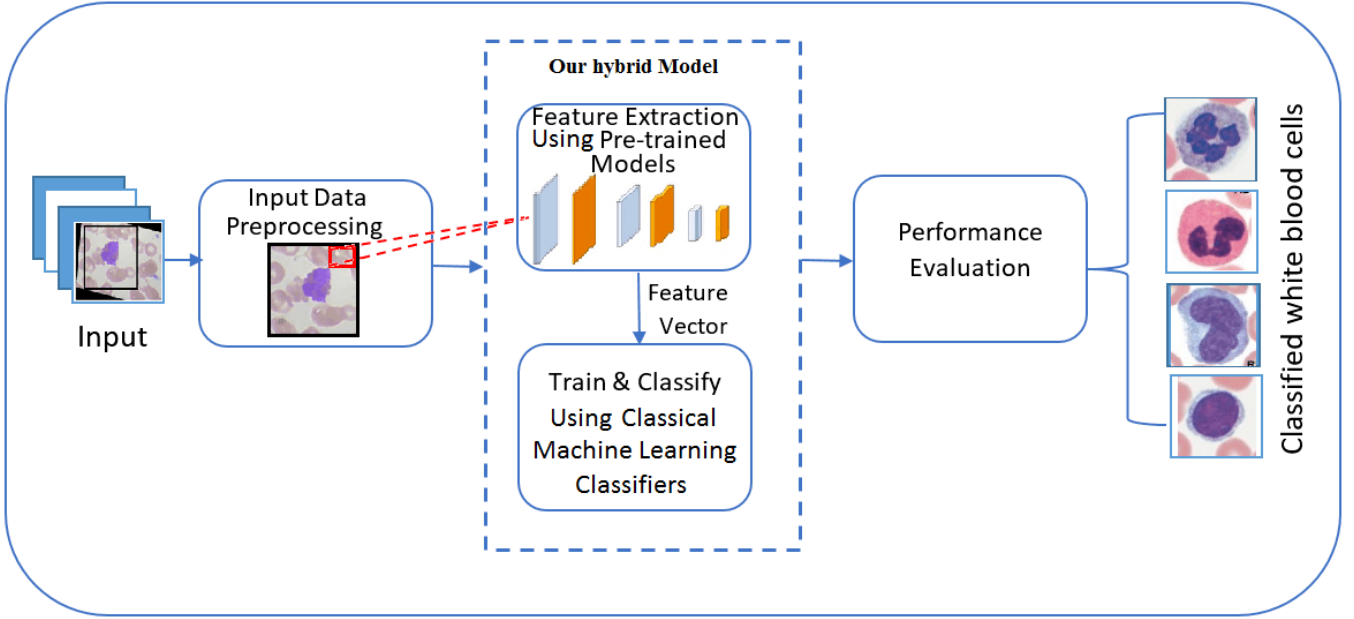
Fig. 3.   The proposed model architecture: The main phases used for classification and evaluation of WBC augmented BCCD dataset.
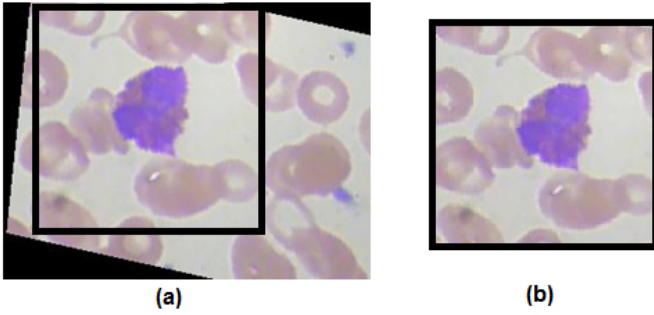

Fig. 4.   (a) Original image with dimensions 320x240. (b) Segmented image with dimensions 200x200.

largest variance of all features, that is added to variances for calculation stability (var_smoothing). Both DT and RF have 'Gini impurity' criterion to measure the quality of a split, whereas DT has 'best' splitter to choose the best split and RF has 100 trees in the forest.

b) Fully connected classifier: FCN is used as our baseline classifier. The FCN is put on top of the pre-trained extraction stage, and finally, a softmax classifier is added. Many experiments are performed in this step to find the FCN, which has the highest accuracy. The used fully connected classifier stage consists of these consecutive layers: dense layer with 512 neurons with 'relu' activation function, dense layer with 256 neurons with 'relu' activation function, dropout layer with 0.5, and finally a dense layer with four neurons with softmax activation function. Then, the model is trained for 20 epochs using 'Adam' optimizer as there is not more improvement in the training and validation accuracies. The batch size and learning rate are, respectively, eight and 1e-7 with decay 1e-7/20, which are empirically determined for the FCN to get the best results for classification with the majority

of pre-trained models.

4) Performance Evaluation: In these experiments, the performance of each pre-trained model with the six classifiers is evaluated using multiple metrics: accuracy, precision, sensitivity (recall), specificity, F1-score, rank-1 accuracy, receiver operating characteristic (ROC) curve, and area under the ROC curve (AUC). For each WBC type, the corresponding predictions are classified into four categories: true positive (TP), false positive (FP), true negative (TN) and false negative (FN).

The accuracy, calculated by using (1), is equivalent to the entirety of TP and TN separated by the aggregate number of leukocytes. The precision, calculated by using (2), is the ratio of correctly predicted true positive to the total predicted positive, i.e., TP and FP. The sensitivity refers to the true positive rate and is equivalent to the proportion of TP to the entirety of TP and FN. It is calculated by using (3). In contrast, the specificity refers to the true negative rate and is equivalent to the proportion of TN to the entirety of TN and FP. It is calculated by using (4). F1-score, calculated by using (5), determines the accuracy based on both precision and sensitivity.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \qquad (1)$$

$$Precision = \frac{TP}{TP + FP} \qquad (2)$$

$$Sensitivity = \frac{TP}{TP + FN} \qquad (3)$$

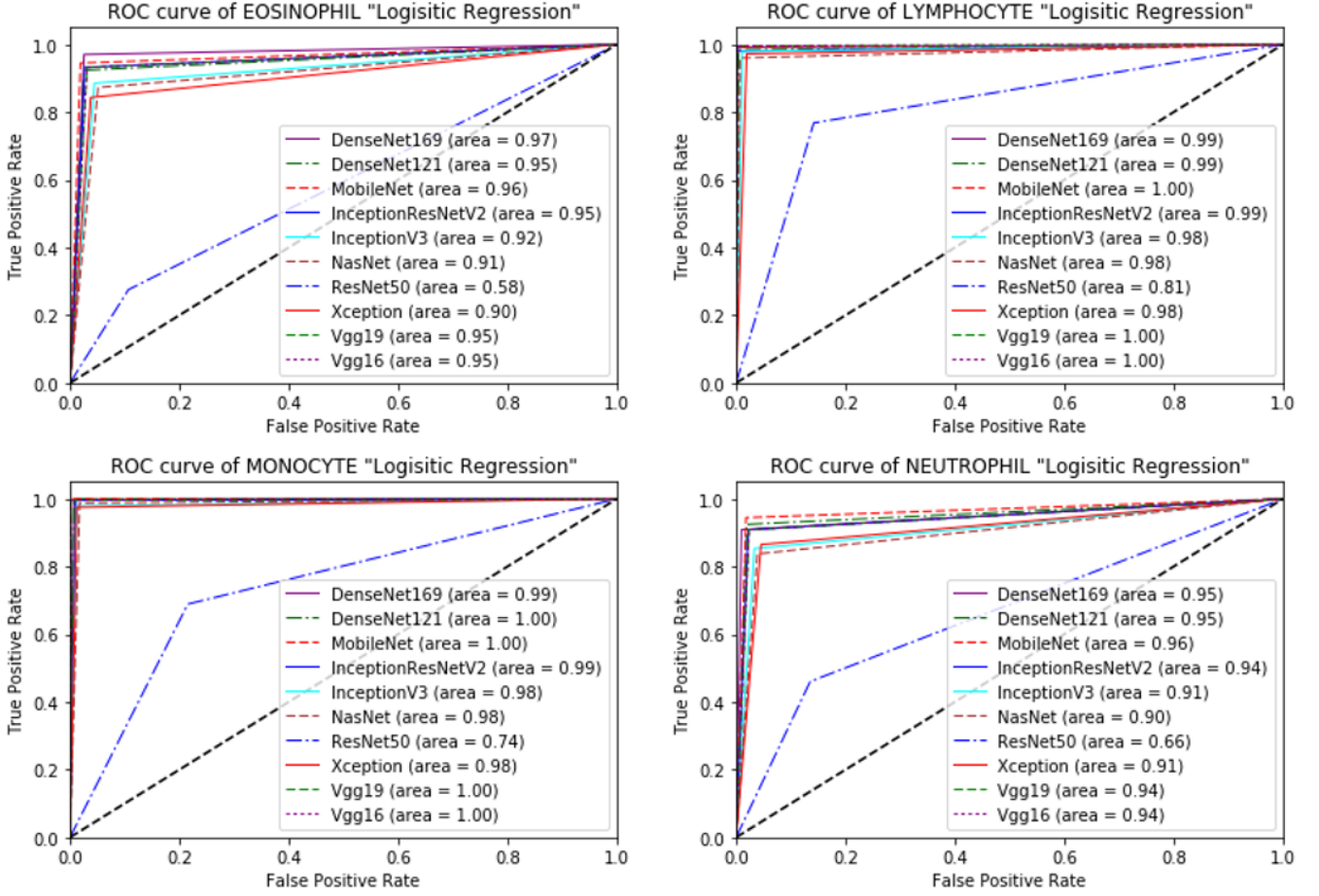$$Specificity = \frac{TN}{FP + TN} \qquad (4)$$

Fig. 5. ROC curve analysis of WBCs classifications into (eosinophil, lymphocyte, monocyte, neutrophil). Proposed model used 10 pre-trained models as a feature extractor and logistic regression as a classifier. AUC values are included.

$$F1 - score = 2\frac{Recall \times Precision}{Recall + Precision} = \frac{2TP}{2TP + FP + FN} \tag{5}$$

The ROC curve is created by plotting the true positive rate (TPR) against the false positive rate (FPR) at various threshold settings, whereas AUC measures the entire two-dimensional area underneath the ROC. A test with no overlap in the two distributions has a ROC curve that passes within the upper left corner (100% sensitivity, 100% specificity). Therefore, the closer the ROC curve is to the upper left corner, the higher the overall accuracy of the test [35].

IV. RESULTS AND DISCUSSION

As shown in Fig. 3, there are multiple phases that our WBC augmented dataset passed through for classification and evaluation.

a) Preprocessing and overall results: Firstly, the input images of the used dataset are preprocessed, as in Sec. III-B1. The used technique for segmentation is based on considering the highest connected regions with pigmented color as the region of interest (i.e., the wanted WBCs). Sometimes, the highest connected regions are not the required cells, as there are other pigmented parts in images as the unsegmented parts

in Fig. 4. In that case, these images are excluded. The correctly segmented 9,760 images are then split into 8,784 images for training and 976 images for testing. Secondly, one of the ten chosen pre-trained deep models, mentioned in Sec. III-B2, is applied separately on training images as a feature extractor. The extracted feature vector differs in length between pre-trained models, as shown in Table I. Then, each extracted feature vector from a pre-trained model is passed to one of the chosen classifiers, mentioned in Sec. III-B3. Six of these classifiers are traditional machine learning classifiers that used the extracted features for training the classifiers themselves, whereas the seventh classifier is the FCN of deep models, which is used as a baseline. Thirdly, the 976 testing images are also passed to the pre-trained model for feature extraction then tested over trained classifiers. Finally, different metrics are used for evaluation. The rank-1 accuracy of each classifier with the different models is calculated, as shown in Table II. Furthermore, each classifier has been evaluated with different performance metrics (see Sec. III-B4). These results are shown in Table V to Table VII and Fig. 5 and 6.

b) Our model results: The rows of Table II represent the pre-trained models used as features extractors. The columns represent the classification accuracy (i.e., rank-1 accuracy)
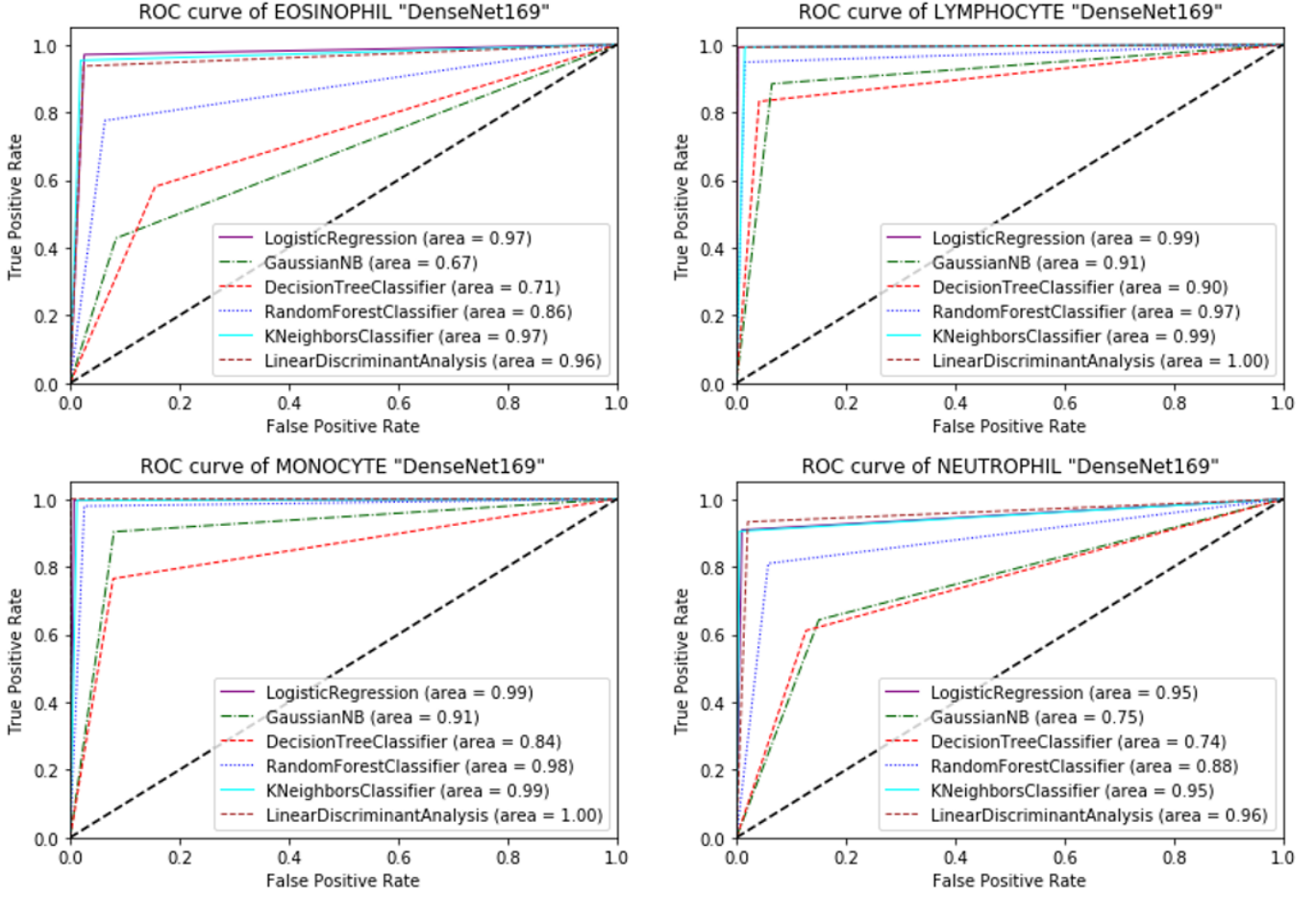
Fig. 6. ROC curve analysis of WBCs classifications into (eosinophil, lymphocyte, monocyte, neutrophil). Proposed model used a pre-trained model DenseNet-169 as a feature extractor and six classifiers. AUC values are included.

TABLE I
FEATURES VECTOR OF PRE-TRAINED MODELS

Model	Features vector length	Model	Features vector length
VGG-16	18432	Inception-V3	32768
VGG-19	18432	Inception ResNet-V2	24576
ResNet-50	2048	Xception	2048
DenseNet-121	50176	1.0 MobileNet-224	50176
DenseNet-169	81536	Mobile NasNet	51744

when that classifier is applied on top of the pre-trained model. The hybrid of pre-trained MobileNet-224 model and logistic regression classifier achieves the highest classification accuracy (i.e., bold underlined value), which is 97.03%. Comparing the results of the baseline classifier (i.e., FCN) with other classifiers shows that our hybrid model always outperforms classical pre-trained model with three classifiers: logistic regression, random forest and KNN.

Table III takes the average of classifications accuracy (i.e., rank-1) of a pre-trained model all over the different classifiers of Table II. In other words, it takes the average of rows of Table II. The DenseNet-169 has the highest average rank-1, which is 86.04%. It outperformed all other pre-trained models by between 2.51% and 33.62%. For that, DenseNet-169 is

considered to be the best feature extractor pre-trained model.

Table IV takes the average of classification accuracy (i.e., rank-1) of a classifier all over the different pre-trained feature extractor models of Table II. In other words, it takes the average of columns of Table II. Four classifiers have a higher average than the baseline FCN: logistic regression, KNN, random forest and LDA. Logistic regression has the highest average at 0.90758, which outperformed the baseline FCN by 25.78%. In addition, logistic regression outperformed all other classifiers between by 0.92% and 32.7%. Therefore, logistic regression is considered to be the best classifier. Table V represents the average of the five metrics used for performance evaluation over each pre-trained model. In this table, DenseNet-169 has the highest average in all five metrics

TABLE II
RANK-1 ACCURACY FOR TEN PRE-TRAINED DEEP MODELS WITH SIX DIFFERENT CLASSIFIERS AND FULLY CONNECTED NETWORK (FCN) USED AS THE BASELINE

Model	FCN (Baseline)	Logistic Regr.	Decision Tree	Rand. Forest	Naive Bayes	KNN	LDA
VGG-16	69.26%	95.80%	72.23%	92.42%	64.14%	94.57%	27.05%
VGG-19	63.42%	96.72%	67.73%	91.91%	63.83%	93.44%	26.43%
ResNet-50	27.36%	55.02%	44.16 %	64.96%	38.83%	67.73%	68.85 %
DenseNet-121	76.84%	95.80%	65.68%	86.99%	67.73%	94.77%	96.93%
DenseNet-169	82.83%	96.62%	69.77%	88.01%	71.62%	96.93%	96.52%
Inception-V3	71.93%	92.42%	55.12%	78.38%	62.30%	89.65%	89.96%
Inception ResNet-V2	69.57%	95.39%	60.35%	85.04%	59.02%	87.09%	92.93 %
Xception	45.90%	91.39%	53.07%	75.1 %	53.18%	89.45%	90.06 %
1.0 MobileNet-224	73.26%	**97.03%**	61.68%	86.07%	53.28%	94.67%	92.42%
Mobile NasNet	69.47%	91.39%	54.82%	80.23%	46.62%	90.06%	85.86%

TABLE III
THE AVERAGE OF CLASSIFICATION ACCURACY (I.E. RANK-1) OVER THE TEN PRE-TRAINED MODELS

Model	Average of Rank-1	Model	Average of Rank-1
VGG-16	73.64 %	Inception-V3	77.11 %
VGG-19	71.93 %	Inception ResNet-V2	78.48 %
ResNet-50	52.42 %	Xception	71.16 %
DenseNet-121	83.53 %	1.0 MobileNet-224	79.77 %
DenseNet-169	**86.04 %**	Mobile NasNet	74.06 %

TABLE IV
THE AVERAGE OF CLASSIFICATION ACCURACY (I.E. RANK-1) OVER THE SEVEN CLASSIFIERS

Classifier	Average of Rank-1	Classifier	Average of Rank-1
Logistic Regression	**0.90758**	Naive Bayes	0.58055
Decision Tree	0.60461	KNN	**0.89836**
Random Forest	**0.82911**	LDA	**0.76701**
FCN (Baseline)	0.64984		

TABLE V
THE AVERAGE OF FIVE METRICS (PRECISION, SENSITIVITY, SPECIFICITY, ACCURACY AND F-SCORE) OVER EACH PRE-TRAINED MODEL

Model	Precision	Sensitivity	Specificity	Accuracy	F-score
VGG-16	0.74	0.74	0.87	0.83	0.74
VGG-19	0.71	0.71	0.86	0.81	0.7
ResNet-50	0.52	0.53	0.77	0.68	0.5
DenseNet-121	0.83	0.83	0.93	0.91	0.83
DenseNet-169	**0.86**	**0.86**	**0.95**	**0.92**	**0.86**
Inception-V3	0.77	0.7	0.9	0.83	0.71
Inception ResNet-V2	0.79	0.79	0.91	0.88	0.78
Xception	0.7	0.71	0.86	0.82	0.7
1.0 MobileNet-224	0.8	0.71	0.91	0.84	0.74
Mobile NasNet	0.74	0.74	0.88	0.84	0.74

used. It scored, on average, 0.86 in precision, sensitivity and F-score, and 0.92 in accuracy and 0.95 in specificity.

Based on previous results, the best model for feature extraction is DenseNet-169, and the best classifier is Logistic Regression. For that, ROC and AUC are used to visualizing these results.

Table VI and Fig. 5 show the ROC and AUC of the best classifier, logistic regression, with the ten pre-trained models.

It is found that the hybrid of logistic regression and the feature extractor MobileNet-224 gives the best results. Table VII and Fig. 6 show the ROC and AUC of the best feature extractor model, DenseNet-169, with the six classifiers. It is found that the hybrid of DenseNet-169 and LDA classifier gives the best results. From Table VI and Table VII, both AUC value of logistic regression with 1.0 MobileNet-224 and LDA with DenseNet-169 have 98% as the highest value. For that, these

TABLE VI
AUC VALUE FOR WBC, USING 10 PRE-TRAINED MODELS AND LOGISTIC REGRESSION

Model	Eosinophil	Lymphocyte	Monocyte	Neutrophil	Avg
DenseNet169	0.97	0.99	0.99	0.95	0.975
DensenNet121	0.95	0.99	1	0.95	0.9725
MobileNet	**0.96**	**1**	**1**	**0.96**	**0.98**
InceptionResnetV2	0.95	0.99	0.99	0.94	0.9675
InceptionV3	0.92	0.98	0.98	0.91	0.9475
NasNet	0.91	0.98	0.98	0.9	0.9425
ResNet50	0.58	0.81	0.74	0.66	0.6975
Xception	0.9	0.98	0.98	0.91	0.9425
VGG19	0.95	1	1	0.94	0.9725
VGG16	0.95	1	1	0.94	0.9725

TABLE VII
AUC VALUE FOR WBC, USING DENSENET-169 AND 6 DIFFERENT CLASSIFIERS

Classifier	Eosinophil	Lymphocyte	Monocyte	Neutrophil	Avg
Logisitic Regression	0.97	0.99	0.99	0.95	0.975
Gaussian NB	0.67	0.91	0.91	0.75	0.81
Decision Tree	0.71	0.9	0.84	0.74	0.7975
Random Forest	0.86	0.97	0.98	0.88	0.9225
KNN	0.97	0.99	0.99	0.95	0.975
LDA	**0.96**	**1**	**1**	**0.96**	**0.98**

two hybrids are considered as best hybrids for our dataset.

c) Comparison results: Compared with previous research in the literature that manually extracts features, our study shares a similar range of results as studies such as [9], but has higher results than [15]. However, when comparing our results with deep learning models, our results are higher than [17], but lower than [19]. However, these variations in performance could be due to using different datasets and preprocessing steps.

V. CONCLUSION

Classifying WBCs in the microscopic image with high accuracy is our main objective. In this research, ten pre-trained models were used as automatic feature extractors, and six classifiers were used for classification. The FCN was used as a comparison baseline for classification. Our hybrid model proves its effectiveness in increasing the classification accuracy over classical pre-trained models. The hybrid of 1.0 MobileNet-224 model and a logistic regression classifier has reached 97.03% classification accuracy. This accuracy outperformed the baseline FCN with 25.78% on average.

REFERENCES

[1] L. Putzu and C. Di Ruberto, "White blood cells identification and counting from microscopic blood image," *International Journal of Medical, Health, Biomedical and Pharmaceutical Engineering*, vol. 7, no. 1, pp. 15–22, 2013.

[2] L. A. Frederick, *Chapter 19 Blood*. University of Texas at Arlington: Pearson Education, Inc., 2015.

[3] "Laboratoryinfo," https://laboratoryinfo.com/mpv-blood-test/, accessed: 2019-11-11.

[4] M. Hassaballah, A. A. Abdelmgeid, and H. A. Alshazly, "Image features detection, description and matching," in *Image Feature Detectors and Descriptors*. Springer, 2016, pp. 11–45, doi:10.1007/978-3-319-28854-3_2.

[5] M.-C. Su, C.-Y. Cheng, and P.-C. Wang, "A neural-network-based approach to white blood cell classification," in *The Scientific World Journal*, vol. 2014, 2014, doi:10.1155/2014/796371.

[6] K. Suzuki, "Overview of deep learning in medical imaging," *Radiological Physics and Technology*, vol. 10, no. 3, pp. 257–273, Sep 2017, doi:10.1007/s12194-017-0406-5.

[7] "Learn convolution neural network," https://learn-neural-networks.com/convolutional-neural-networks/, accessed: 2019-11-11.

[8] P. Yampri, C. Pintavirooj, S. Daochai, and S. Teartulakarn, "White blood cell classification based on the combination of eigen cell and parametric feature detection," in *1ST IEEE Conference on Industrial Electronics and Applications*, May 2006, pp. 1–4, doi:10.1109/ICIEA.2006.257341.

[9] A. Falcon, A. Taboada-Crispi, M. Orozco Monteagudo, M. Aliosha-Perez, and H. Sahli, "Classification of white blood cells using morphometric features of nucleus," *Cuba-Flanders Workshop on Machine Learning and Knowledge Discovery*, 02 2010.

[10] M. Habibzadeh, A. Krzyzak, and T. Fevens, "Comparative study of shape, intensity and texture features and support vector machine for white blood cell classification," *Journal of Theoretical and Applied Computer Science*, vol. 7, no. 1, pp. 20–35, 2013, doi:10.1007/978-3-319-11656-3_20.

[11] A. Gautam and H. Bhadauria, "Classification of white blood cells based on morphological features," *2014 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pp. 2363–2368, 2014, doi:10.1109/ICACCI.2014.6968362.

[12] J. Prinyakupt and C. Pluempitiwiriyawej, "Segmentation of white blood cells and comparison of cell morphology by linear and naive bayes classifiers," *Biomedical engineering online*, vol. 14, p. 63, 06 2015, doi: 10.1186/s12938-015-0037-1.

[13] S. Ravikumar and A. Shanmugam, "Wbc image segmentation and classification using rvm," *Applied Mathematical Sciences*, pp. 2227–2237, 01 2014, doi: 10.12988/ams.2014.43191.

[14] S. Ravikumar, "Image segmentation and classification of white blood cells with the extreme learning machine and the fast relevance vector machine," *Artificial Cells, Nanomedicine, and Biotechnology*, vol. 44, no. 3, pp. 985–989, 2016, doi:10.3109/21691401.2015.1008506.

[15] O. Sarrafzadeh, A. M. Dehnavi, H. Y. Banaem, A. Talebi, and A. Gharibi, "The best texture features for leukocytes recognition," *Journal of medical signals and sensors*, vol. 7, no. 4, pp. 220–227, 2017, doi:10.4103/jmss.JMSS_7_17.

[16] H. Fu, Y. Xu, S. Lin, D. W. Kee Wong, and J. Liu, "Deepvessel: Retinal vessel segmentation via deep learning and conditional random

field," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016*. Springer International Publishing, 2016, pp. 132–139, doi:10.1007/978-3-319-46723-8_16.

[17] A. Shahin, Y. Guo, K. Amin, and A. A. Sharawi, "White blood cells identification system based on convolutional deep neural learning networks," *Computer Methods and Programs in Biomedicine*, vol. 168, pp. 69 – 80, 2019, doi: 10.1016/j.cmpb.2017.11.015.

[18] S. Rajaraman, S. K. Antani, M. Poostchi, K. Silamut, M. A. Hossain, R. J. Maude, S. Jaeger, and G. R. Thoma, "Pre-trained convolutional neural networks as feature extractors toward improved malaria parasite detection in thin blood smear images," in *PeerJ 6:e4568*, 2018, pp. 271–281, doi:10.7717/peerj.4568.

[19] M. Habibzadeh, M. Jannesari, Z. Rezaei, H. Baharvand, and M. Totonchi, "Automatic white blood cell classification using pre-trained deep learning models: Resnet and inception," in *Proc. of SPIE 10696, Tenth International Conference on Machine Vision (ICMV 2017)*, 2018, pp. 1 069 612–1, doi: 10.1117/12.2311282.

[20] G. Liang, H. Hong, W. Xie, and L. Zheng, "Combining convolutional neural network with recursive neural network for blood cell image classification," *IEEE Access*, vol. PP, pp. 1–1, 07 2018, doi:10.1109/ACCESS.2018.2846685.

[21] Kaggle. (2018) Blood cell images 12,500 images: 4 different cell types. Https://www.kaggle.com/paultimothymooney/blood-cells/home.

[22] D. Kaur and Y. Kaur, "Various image segmentation techniques : A review," in *International Journal of Computer Science and Mobile Computing (IJCSMC)*, vol. 3, no. 5, 2014, pp. 809–814.

[23] N. Tajbakhsh, J. Y. Shin, S. R. Gurudu, R. T. Hurst, C. B. Kendall, M. B. Gotway, and J. Liang, "Convolutional neural networks for medical image analysis: Full training or fine tuning?" *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1299–1312, May 2016, doi:10.1109/TMI.2016.2535302.

[24] F. Chollet, *Deep Learning with Python*, 1st ed. Manning Publications Co., 2017.

[25] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv 1409.1556*, 09 2014.

[26] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016, pp. 770–778, doi:10.1109/CVPR.2016.90.

[27] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," July 2017.

[28] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," pp. 2818–2826, June 2016, doi:10.1109/CVPR.2016.308.

[29] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," p. 4278–4284, 2017.

[30] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," pp. 1800–1807, July 2017, doi:10.1109/CVPR.2017.195.

[31] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," *ArXiv*, vol. abs/1704.04861, 2017.

[32] B. Zoph, V. Vasudevan, J. Shlens, and Q. V. Le, "Learning transferable architectures for scalable image recognition," *CoRR*, vol. abs/1707.07012, 2017, doi:10.1109/CVPR.2018.00907.

[33] J. Brownlee, *Master Machine Learning Algorithms: Discover how They Work and Implement Them from Scratch*, 2016. [Online]. Available: https://books.google.com.eg/books?id=PdZBnQAACAAJ

[34] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[35] M. H. Zweig and G. Campbell, "Receiver-operating characteristic (roc) plots: a fundamental evaluation tool in clinical medicine." *Clinical Chemistry*, vol. 39, no. 4, pp. 561–577, 1993. [Online]. Available: http://clinchem.aaccjnls.org/content/39/4/561

Ensaf H. Mohamed received her Ph.D. in the Computer Science department, Helwan University, Cairo, Egypt, 2013. Her recent research focuses on Medical Image Analysis, Machine Learning, Natural Language Processing, Text Mining, and Social Media Mining. Currently, she is an assistant professor, faculty of computers and information, Helwan University, Cairo, Egypt.

Wessam H. El-Behaidy earned her M.Sc. and Ph.D. in the Computer Science department, Helwan University in 2004 and 2012 respectively. She is assistant professor in faculty of Computers and Information, Computer Science department since 2012. Her research interests also include medical image analysis, protein structure prediction, machine learning, and pattern recognition.

Ghada Khoriba assistant professor in Computer Science department, Helwan University, Egypt since 2011. She had got her Doctor of Philosophy in Engineering, Computer Science, from Graduate School of Systems and Information Engineering, University of Tsukuba, Japan in 2010. She received the B.S. and M.S. degrees from the Department of Computer Science of Helwan University, Egypt, in 2000 and 2004, respectively. Her research interests include geometric graphs, Distributed systems, Neural Networks, and Pattern Recognition.

Jie Li is a professor at Department of Computer Science and Engineering, School of Electronic Information and Electrical Engineering, Shanghai Jiaotong University, Shanghai, China. Dr. Li received the B.E. degree in computer science from Zhejiang University, Hangzhou, China, the M.E. degree in electronic engineering and communication systems from China Academy of Posts and Telecommunications, Beijing, China. He received the Dr. Eng. degree from the University of Electro-Communications, Tokyo, Japan. He has been a visiting Professor in Yale University, USA, Inria Sophia Antipolis and Inria Grenoble–Rhone-Aples, France, during the sabbatical year in September 2014 through August 2015. His current research interests are in Big Data, Cloud, wireless and mobile networking, network secuirty, IoT, OS, modeling and performance evaluation of information systems. His research has been supported by many national and international projects sponsord by JSPS (Japan Society for the Promotion of Science), NICT (National Institute of Communications Technology), NII (National Institute of Informatics), and other foundations such as Okawa Foundation, Casio Science Promotion Foundation, International Information Science Foundation, TEPCO Memorial Foundation, and Japanese-Isreali Research Grant, etc. Dr. Li has served as an editor and a guest editor for many international journals such as IEEE TBD, IEEE JSAC, IEEE IoTJ, IEEE TVT, IEEE TCC, IEEE TNSE, IEEE Access, and IEEE Network, etc. He has served on several editorial boards for the IPSJ Journal and many international professional journals. He has also served on the program committees for several international professional conferences.