Computer Methods and Programs in Biomedicine 168 (2019) 69–80

Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

journal homepage: www.elsevier.com/locate/cmpb

White blood cells identification system based on convolutional deep neural learning networks

A.I. Shahin[a,b], Yanhui Guo[d,*], K.M. Amin[c], Amr A. Sharawi[a]

[a] Department of Biomedical Engineering, Cairo University, Egypt
[b] Department of Biomedical Engineering, HTI, Egypt
[c] Department of Information Technology, Menoufia University, Egypt
[d] Department of Computer Science, University of Illinois at Springfield, Springfield, IL, USA

ARTICLE INFO

Article history:
Received 30 March 2017
Revised 24 August 2017
Accepted 14 November 2017

Keywords:
Blood smear image
Deep learning
Transfer deep learning
WBCs identification
Deep features visualization

ABSTRACT

Background and objectives: White blood cells (WBCs) differential counting yields valued information about human health and disease. The current developed automated cell morphology equipments perform differential count which is based on blood smear image analysis. Previous identification systems for WBCs consist of successive dependent stages; pre-processing, segmentation, feature extraction, feature selection, and classification. There is a real need to employ deep learning methodologies so that the performance of previous WBCs identification systems can be increased. Classifying small limited datasets through deep learning systems is a major challenge and should be investigated.

Methods: In this paper, we propose a novel identification system for WBCs based on deep convolutional neural networks. Two methodologies based on transfer learning are followed: transfer learning based on deep activation features and fine-tuning of existed deep networks. Deep acrivation featues are extracted from several pre-trained networks and employed in a traditional identification system. Moreover, a novel end-to-end convolutional deep architecture called "WBCsNet" is proposed and built from scratch. Finally, a limited balanced WBCs dataset classification is performed through the WBCsNet as a pre-trained network.

Results: During our experiments, three different public WBCs datasets (2551 images) have been used which contain 5 healthy WBCs types. The overall system accuracy achieved by the proposed WBCsNet is (96.1%) which is more than different transfer learning approaches or even the previous traditional identification system. We also present features visualization for the WBCsNet activation which reflects higher response than the pre-trained activated one.

Conclusion: a novel WBCs identification system based on deep learning theory is proposed and a high performance WBCsNet can be employed as a pre-trained network.

1. Introduction

The blood smear under a microscope contains useful information for diagnosis of many diseases. Blood components are divided into three categories: red blood cells (RBCs), white blood cells (WBCs) and platelets [1]. WBCs are divided into five types (shown in Fig. 1) by percentage as: basophil (0–1%), eosinophil (1–5%), lymphocyte (20–45%), monocyte (2–10%) and neutrophil (50–70%). Whereas each WBC type has its own shape of a nucleus and cytoplasm, RBCs have no nuclei [2].

In the industry, there are many automated cell morphology (ACM) systems like Cella-Vision [3], Hema-CAM [4] and MEDICA EasyCell® Assistant [5]. Such systems employ the traditional identification system theory as shown in Fig. 2. Traditional systems consist of different stages such as: pre-processing, segmentation, feature extraction, feature selection (FS), and classification. The pre-processing stage includes de-noising and color correction processes. General image enhancement algorithms are utilized for blood smear image enhancement (e.g. histogram matching, averaging filtering). The segmentation stage plays the most important role in the identification system and affects the overall system ac-

Abbreviations: ACM, Automated cell morphology; RBCs, Red blood cells; WBCs, White blood cells; FS, Feature selection; DL, Deep learning; GPU, Graphic processor unit; TLA, Transfer learning approach; DeCA, Deep convolutional activation; DeCAF, Deep convolutional activation features; ANN, Artificial neural network; CNN, Convolutional neural network; DCNNs, Deep convolutional neural networks; PS, Pooling stride; ReLU, Rectified linear unit.

* Corresponding author.
E-mail addresses: ahmed.esmail@hti.edu.eg (A.I. Shahin), yguo56@uis.edu (Y. Guo).

https://doi.org/10.1016/j.cmpb.2017.11.015
0169-2607/

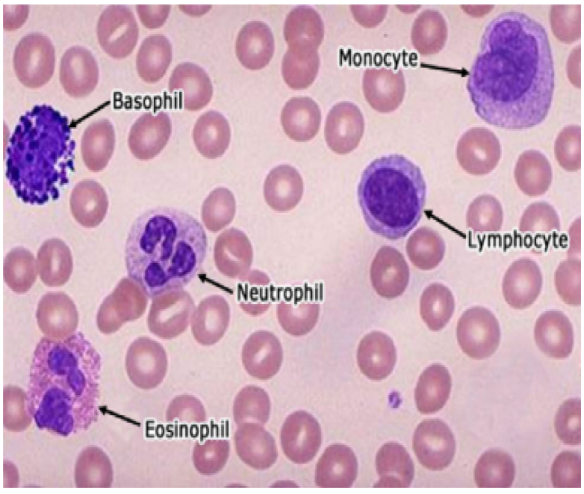

Fig. 1. Different WBCs types (basophil, eosinophil, lymphocyte, monocyte and neutrophil).

curacy [6] and a lot of segmentation algorithms are available in the literature. The feature extraction stage extracts different features (morphologic, color, texture ...etc.) for both nuclei and cytoplasm regions. More complex features (spectral [7], wavelet [8]) may be computed. The classification stage uses multi-step classification or majority-voting classification.

In the traditional WBCs identification system, we observed the following; there is no pre-processing algorithms specialized in enhancement of blood smear image and cannot be generalized. The previous employed algorithms do not take into consideration the nature of blood smear images such as their different components appearance, light distribution and variation of staining intensities. The segmentation algorithms cannot be generalized for all WBCs as their different shapes and colors. The WBCs segmentation also requires the separation of each WBC components (nuclei, cytoplasm) and since each WBC has its own shape and color nature (as shown in Fig. 1), it is difficult to have a general segmentation algorithm for all WBCs. FS reduces the feature vector dimension, and decreases the training and testing times. However, classification accuracy may be lowered due to FS process [9]. The multi-stage classification is also increasing the overall system complexity and consumes a lot of processing time. For all of the above reasons, the traditional method does not achieve the real sense of expertise pathologist in WBCs identification, leaks robustness, and works on limited size datasets under strict conditions. These previous drawbacks create real motivations to employ the deep learning (DL) methodology because of its applicable advantages in our problem-solving. These advantages are as follows; there is no need to perform an enhancement stage for the input image since DL is insensitive to image quality [10], there is no need for neither segmentation stage nor hand-crafted features extraction stage since the features are already extracted through the convolutional concept, and finally the classification stage in DL systems is more simple and does not require a multi-stage of classification.

DL is a recent machine learning theory which exponentially grows up in the last years. DL is an extension of artificial neural network (ANN). Convolutional neural network (CNN) was proposed by LeCun [11], which aims to generate learned filters by performing multi-convolution of the same input, through the multilayers network. CNNs have demonstrated as a powerful tool for image recognition, segmentation, detection and retrieval [12–16].

To apply the DL theory for image classification, a large dataset is needed. In medical fields, it is very difficult to have a large dataset for training and testing procedure. In [17], the author proved that it is possible to apply DL with enough dataset to train CNN network architecture and to obtain the high accuracy performance. In DL, the training of CNN consumes a lot of time. To overcome this problem, the recent hardware technology has been employed to accelerate the computations. The graphic processing unit (GPU) is used for the implementation of CNNs as a fast way for parallel computations.

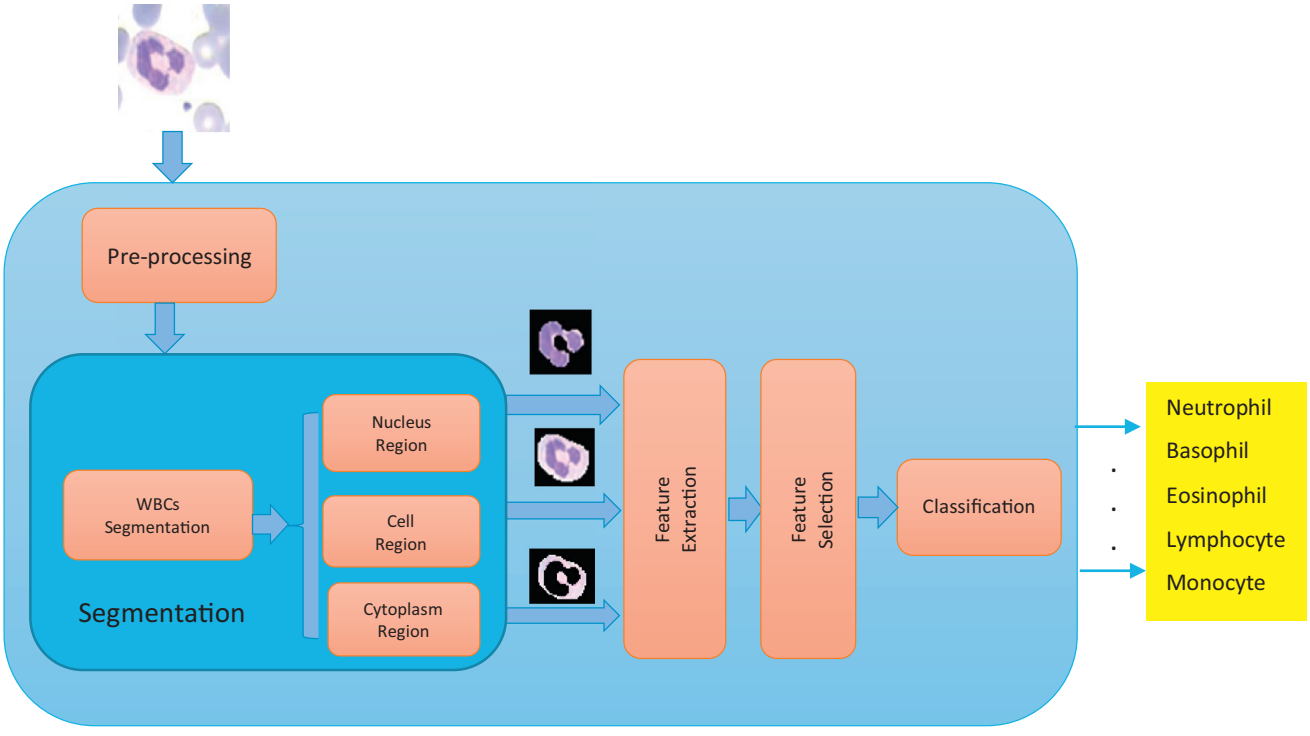

Fig. 2. A traditional WBCs identification system.

Table 1
Survey of recent traditional WBCs identification systems.

Reference	Dataset (size/sources)	Used features/Total number	Classifier/1or 2 steps of classifier	Feature reduction	Classifier accuracy
[7]	60/1	(Morph.+Stat.+Spect.)/(4 Morph.+Spect. Coeff:ND.)	SVM/1	None	90%
[9]	254/1	(Morph.+Text.)/(112)	(ANN+SVM)/2	SFS	96%
[18]	1078/2	(Morph.+Stat.) /(15)	Linear and Naïve Bayes classifier/1	None	96%
[19]	2172/1	(Morph.+Stat.) /(19)	LDA/2	None	93.9%
[20]	33/1	(Morph.+Stat.+Text.)/(131)	SVM/2	None	93%
[21]	12/1	Cell sub-image /(14,400)	MLP/2	PCA	95%
[22]	267/1	(Morph.+Text.)/(7 Morph., Text.:ND.)	Naive Bayes classifier	None	92.7%
[23]	487/1	(Morph.+Stat.+Text.)/(ND)	Fuzzy classifier/2	None	96%
[24]	450/2	(Morph.+Stat.+Text.)/(20)	MLP/1	None	99.1%

Used abbreviations; T: Total No. of features, Text.: Texture, Morph.: Morphological, Stat.: Statistical, Spect.: Spectral, Spect. Coeff.: Spectral Coefficients, ND.: Not Defined. SFS.: Sequential forward selection, Class.: Classification, LDA.: Linear Discriminate Analysis, MLP.: Multi-Layer Perceptron, ANN.: Artificial Neural Network, SVM.: Support Vector Machine.

The rest of the paper is organized as the following. In the next section, the background for both traditional identification systems and DL systems is presented. Then, in Section 3, the proposed identification system is described. In Section 4, the experimental results are presented and discussions are given in Section 5. Finally, the conclusion is presented.

2. Background

In this section, we present a background for the recent works on traditional WBCs identification system, the problem definition in such systems that create the motivation to employ the DL theory. Then, the recent medical applications that employ the DL technique as classification method.

2.1. Traditional WBCs identification systems

In the last decade, many research articles were interested in identification of WBCs. A brief summary of the previous pathological identification systems is given in Table 1. The following details are mentioned: size and sources of datasets used in the experiments, feature extraction method, size of feature pool (if available), feature reduction/selection method, classifier type, and classification accuracy.

From Table 1, the previous WBCs identification systems only achieve high classification accuracy under strictly controlled conditions as in [24] or by using small datasets as in [7,21,22]. In [9], the author performed color adjustment as a pre-processing step to the input image in order to increase the classification accuracy. However, there is still a challenge to be considered in classification accuracy. The main scope of previous WBCs identification systems was increasing the efficiency of segmentation algorithms. However, any over/under segmentation errors have a negative effect on the overall system accuracy. The feature extraction is also a very crucial step. In [20], it was reported that the morphological features are susceptible to any errors in segmentation stage. In [18], the authors extracted the WBCs features from RGB color components individually. Moreover, the ratios of statistical features between these color components are also derived which increase the complexity of the features extraction stage. In [9,19,23], the authors used multi-step classification which also increases the system complexity. All of the above drawbacks of previous traditional identification systems lead to a real need to develop more adaptive and high-performance identification system.

2.2. DL in medical applications

DL today is employed to increase the efficiency of many identification systems. Many researchers applied the DL technique as a tool for classification problems in medical applications. In [17], the author used the DL to build an automated skeletal bone age assessment using 1391 images. In [25], they employed the DL for classifying echocardiography into 8-classes using 432 video images. They fused the extracted features from two different CNNs architectures, and then fed such features to the classification layer. In [26], authors employed the DL for shear-wave elastography images to classify the breast cancer into 2-classes using 227 images. Moreover, they built a limited deep belief network for feature extraction and FS processes. In [27], the DL was used in the prediction of the survival based on MRI images in amyotrophic lateral sclerosis disease. Their system output was classified into 3-classes (short, medium and long) using 135 patients' data. They built three different CNNs architectures; the first CNN process the patient clinical characteristics, the second process the structural connectivity and the third process the brain morphology features. In [28], they proposed a novel DCNN architecture for heterogeneous iris recognition system using 7000 images, however, the system output was binary classification result which makes the classification less difficult.

All of previous works used gray-scale medical images such as ultrasound, CT, MRI or X-ray as input for the deep CNNs. In our case, the WBCs in pathology images are presented in RGB color space, where the color information is the very important identification of each WBCs type as shown in Fig. 1. One of the big challenges in employing the DL in the medical applications is the availability of large datasets which are not available in the medical field. The transfer learning methods are proposed as an acceptable solution to process limited datasets of medical applications [17].

The suggested scenarios to process limited datasets in DL are: deep convolutional activation (DeCA) method [29,30], fine-tuned pre-trained network method [31], and fused features method [32]. In [17], the author employed the DeCA method which achieved

acceptable results and the performance of the proposed system was enhanced through for the fine-tuning process. In [17,33], the fine-tuned pre-trained network with its previous knowledge and learned filters weights were employed to train the target dataset. The last method to process limited dataset in DL is fusing the CNN features with other hand-crafted features [34], or even fusing features extracted from multi CNN architectures [25,27]. On the other hand, with a suitable dataset relative to each classification problem, the CNN can be built and trained from scratch [17,28].

In this paper, our contributions are as follows. A robust automatic identification system is proposed for specific medical application in the pathology area, which employs transfer learning methodology in complete identification system to process the limited dataset. Usage of different dataset sources - acquired from different ACM equipments, make the system more adaptive with various parameters (image resolution, lighting conditions, staining artifacts…etc.). The proposed system process unbalanced datasets because of the nature of WBCs percentage in the blood. This makes the classification problem more difficult. A novel end-to-end CNN structure is built to train all available datasets, and achieve more accuracy than the previous traditional identification system. The proposed WBCsNet considered the first pre-trained network for WBCs classification. Finally, the WBCsNet is tested as a pre-trained network to extract the *off-the-shelf* features for a very limited balanced dataset size and works well.

3. Proposed methods

3.1. Overview

Deep learning is a powerful tool for image category classification and is employed in many medical applications [25–28, 33–35]. The big challenge here is how to employ the DL technique for identifying the different 5-WBCs classes using limited dataset. In this section, we propose three different approaches for WBCs identification system: DeCA approach, fine-tuned network approach, and finally constructing a new optimized low depth DL architecture called WBCsNet.

3.2. DeCA approach

DeCA approach is one of the transfer learning approaches (TLA) that uses the pre-trained CNN models to extract what is known as *off-the-shelf* features. In the case of limited size dataset, TLA reflects significant performance increasements [17,25,26]. In TLA approach, the network was previously trained on general imagery. There are several public pre-trained networks for image category classification such as LeNet [11], OverFeatNet [15], AlexNet [36], GoogLeNet [37], OxfordNet [38], ImageNet [39] and VGGNet [40]. In [16], TLA was used for remote sensing applications in high-resolution images. In [17], they employed the previous pre-trained networks as a feature extractor in skeletal bone age assessment. This means that the features that are extracted from the pre-trained networks can be used to train an external classifier like SVM, Ad boost, KNN or MLP.

In this paper, in order to get the highest possible classification accuracy, we employ different public pre-trained networks: Overfeat, Alex and VGG [17]. Our proposed identification system (shown in Fig. 3) consists of 4-steps: Pre-processing, features extraction, features selection, and classification. The details of each stage are discussed briefly as follows;

3.2.1. Pre-processing

Preparing images is done by resizing each one to be compatible with each pre-trained CNNs image input layer. Each of the previous pre-trained networks has its own input layer size: OverfeatNet ($221 \times 221 \times 3$), AlexNet ($227 \times 227 \times 3$) and VGGNet ($224 \times 224 \times 3$). Since WBCs images are collected from different datasets, any image should be exposed to either down-sampling or up-sampling according to its resolution.

3.2.2. Feature extraction

After the supervised training of CNN on general images categories using state-of-the-art method, the pre-trained network has powerful features on generic vision tasks and specifically on image category classification. The learned weights are adapted to shape, texture, complex features of the general images used. Then, the target dataset is passed through the pre-trained network with its previous weights and learned filters. Finally, "*off-the-shelf*" features will be presented at a certain fully-connected layer according to the CNN structure.

3.2.3. Feature selection (FS)

After the information is extracted from the CNN architecture, the size of features pool reaches to 4096 features (using either AlexNet and VGGNet) or 2048 features (using OverfeatNet). In our proposed method, we evaluate the performance of FS stage either chi-squared or PCA technique. The chi-squared has been proven to be more efficient technique than PCA in image medical FS [41]. However, we also investigate such two FS techniques with the deep features.

3.2.4. Classification

In [26], the support vector machine (SVM) achieved the highest classification accuracy. In this paper, we employ the SVM as a classifier for original feature pool without reduction, then for the reduced feature pool after FS procedure. This help to get the effect of FS procedure on the proposed system performance. The output of SVM classifier has different 5 WBCs classes.

3.3. Fine-tuned network approach

One of the TLA approaches in DL theory is fine-tuning of the pre-trained networks. Many of previous methods employed this approach in medical applications [17,34,35]. The tuning is implemented through little modifications in the CNN architecture such as applying different dataset, modifying either number of output classes or training parameters. In [17], the author proved that fine tuning of the pre-trained networks increase significantly the performance of identification system more than DeCA approach. For that, we perform fine-tuning for both AlexNet and LENet-5 architectures to be suitable for our classification target. The number of classification outputs are tuned in AlexNet from 1000 to 5 classes and in LeNet-5 from10 to 5 classes. The WBCs dataset is fed through the previous fine-tuned networks, so that new weights and learned filters are obtained.

3.4. WBCsNet architecture

In this approach, a novel CNN architecture "WBCsNet" is proposed. The WBCsNet architecture, shown in Fig. 4, consists of image input layer, three main convolutional layers, two pooling layers, four activation rectified linear unit (ReLU), two fully-connected layers and classification (Softmax) layer.

The pre-processing is a preparing stage for the input image which is fed through the CNN. In this paper, we employ three different datasets, however, most of images in the utilized datasets have a resolution of $70 \times 70 \times 3$. In [28], the author fixed the input image width and resize the input image height. He evaluated the performance of his proposed CNN architecture with three input image heights, the lowest down sampling value achieved the best performance. In our proposed WBCsNet, we investigate how the

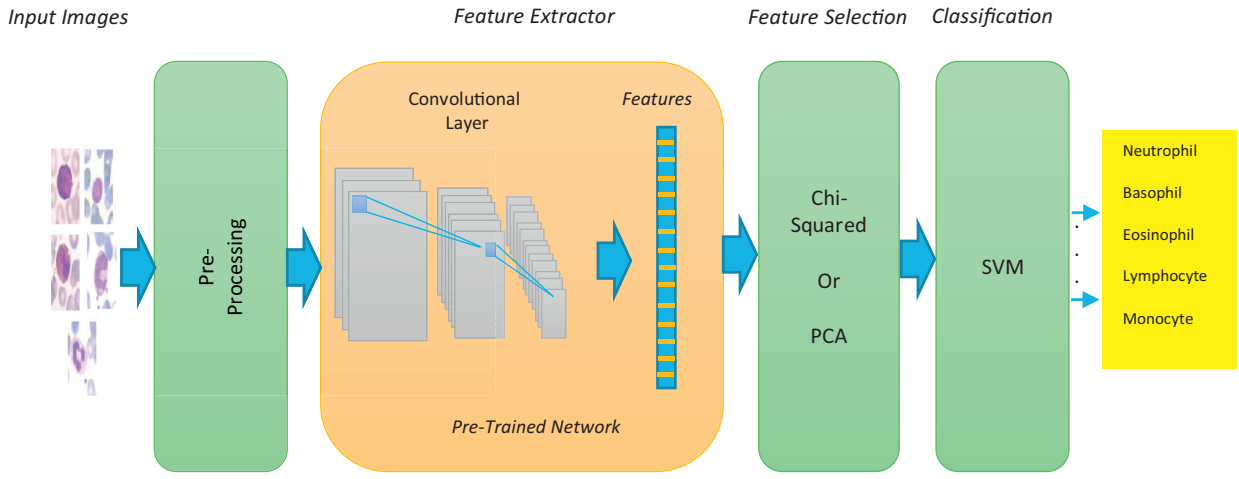

Fig. 3. Our proposed system using TLA approach.

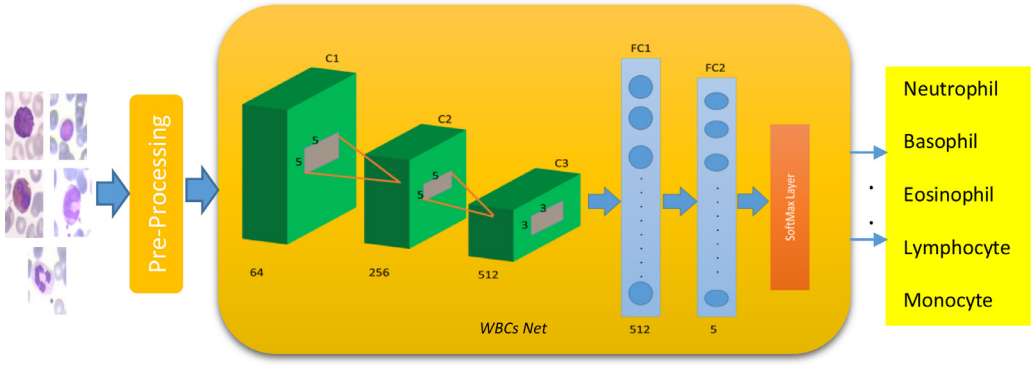

Fig. 4. The proposed WBCsNet architecture.

system performance varies with different image input layer sizes. We perform several experiments to optimize the image input layer size which is selected to be $50 \times 50 \times 3$.

The state of art of the CNN architecture had been introduced by LeCun [11], involved two layers: convolutional layer and pooling layer. Moreover, ReLU activation function is also involved in most of modern CNN architectures. The CNN mathematical model can be explained using different equations. For a given training set of RGB images: x^i and y^i vector (indicates the affiliated classes of x^i), the image features maps will be learnt based on the solution of the following [34];

$$\text{argmin}_{w_1,\ \ldots,\ w_L} \frac{1}{n} \sum_{i=1}^{n} \ell \left(\text{f} \left(x^i;\ w_1,\ \ldots,\ w_L \right), y^i \right) \tag{1}$$

where $w1 \ldots w_L$ are the learnt weights which represent the feature maps based on CNN theory, ℓ refers to the loss function and f is the classification function.

In Eq. (2), the new features value V_{ij}^{xy} at position (x, y) on the jth feature map which is located at the ith network layer can be extracted by convolving over the local neighborhood on feature maps of the previous layer (ith−1). These features map is biased by b_{ij} and the result is passed through an activation function (hyperbolic tangent function).

$$V_{ij}^{xy} = \tanh(\text{b}_{\text{ij}} + \sum_{m} \sum_{\text{p}=0}^{\text{p}_\text{i}} \sum_{\text{q}=0}^{\text{Q}_\text{i}} \text{W}_{ijm}^{\text{pq}}\ V_{(\text{i}-1)\text{m}}^{((x+p)(\text{y}+\text{q}))} \tag{2}$$

where m refers to index over the set of feature maps in the (ith−1) layer, $w_{\text{ijm}}^{\text{pq}}$ refers to value at the position (p, q) of the kernel connected to the kth feature map and (p, q) is 2D position of a kernel P_i, Q_i refers to the height and width of the kernel.

The pooling layer is constructed to reduce the feature maps dimensions. The reduction is done by pooling over the local neighborhood on the feature maps in the previous layer. The downsize rate is controlled by pooling stride (PS). The size of feature maps x_{i+1} follows the rule that is set out in Eq. (3).

$$x_{i+1}^{size} = \left(\frac{x_i^{size} - F_i + 2 \times Pad}{Stride} + 1 \right) \tag{3}$$

where x_i^{size} : represent the input resolution layer, F: filter size, *Pad:* is the padding pixels, Stride: is the pooling stride size.

The ReLU activation function thresholds the data with zero. The advantage of using ReLU over traditional activation functions like sigmoid or hyperbolic tangent is that it performs a constantly switching off subset of output unit by enforcing sparse activation in addition to achieving a type of intensity invariance. The ReLU function can be defined as follows:

$$y_{ijk} = \max(0,\ x_{ijk}) \tag{4}$$

In our proposed WBCsNet, the first convolutional layer (C1) size is selected to be $5 \times 5 \times 64$. The neighborhood filter size is 5×5, where 64 filter banks are chosen to apply the sub-sampling. PS controls the downsize rate, PS at the C1 is set to 1 and padding size is set to 0. C1 is followed by ReLU activation function, then a pooling layer with maximum values are applied with neighborhood filter size 2×2 and zero padding.

The size of the second convolutional layer (C2) is $5 \times 5 \times 256$. Neighborhood filter size is 5×5, where 256 filter banks are chosen to apply the sub-sampling. PS controls the downsize rate, PS at C2

Table 2
Different types of WBCs found in different datasets.

WBCs Type/Dataset	Neutrophil	Eosinophil	lymphocyte	Monocyte	Basophil	Total WBCs in each dataset
Dataset1	25	2	85	12	1	125
Dataset2	55	43	55	48	53	254
Dataset3	1412	83	525	142	10	2172
Total WBCs (Dataset_ALL)	1492	128	665	202	64	2551

is set by 1 and padding size is set to 0. C2 is followed by ReLU activation function, then a pooling layer with maximum values is applied with neighborhood filter size 2×2 and zero padding.

The third convolutional layer (C3) size is $3 \times 3 \times 512$. Neighborhood filter size is 3×3, where 512 filter banks are chosen to apply the sub-sampling. PS controls the downsize rate, PS at C3 is set by 1 and padding size is set to 0. C3 is followed by ReLU activation function.

Since AlexNet established a successful visual recognition concept based on three fully-connected layers [36] before the classification layer, we follow the similar approach. In WBCsNet, the first fully-connected layer (FC1) reduce the size of features to 512. FC1 layer is followed by ReLU activation function. The final fully-connected layer (FC2) contains the 5-classes of healthy WBCs, so, the final feature pool dimension is 5×512.

Finally, the classification layer is performed using Softmax classifier [42] that determines a score of normalized class probabilities. Eq. (5) defines mathematically the Softmax function.

$$f_i(z) = \frac{e^{z_j}}{\sum_k e^{z_k}} \tag{5}$$

where the function takes a vector of arbitrary real-valued scores (in z) and compresses it to a vector of values between zero and one that sums to one. Obtaining class scores f involves the calculation of cross-entropy loss that is formulated in Eq. (6).

$$L_i = f_{y_i} + log \sum_j e^{f_j} \tag{6}$$

where the f_j refers to the jth element of the vector of class scores f [42].

The feature pool resulted from the WBCsNet has low dimensions relative to the previous pre-trained networks. The cross-validation technique with 10-folds has been employed here easily in training and testing to evaluate the proposed WBCsNet. This leads to an independent dataset without over-fitting. This also generates a model which can accurately be a predictive model and has more reality in practice.

4. Experimental results

In this work, we employ three different previously public published datasets called; “Dataset1”, “Dataset2” and “Dataset3. Dataset1 is extracted from ALL_DB1 and ALL_DB2 datasets [43]. The dataset contains healthy WBCs and saved in JPG images with 24 bits depths. The ALL_DB1 dataset had been captured using a digital microscopy (300×–500× objective lens) which attached to a camera (2592×1944 pixels resolution). The ALL_DB2 dataset was manually cropped with a resolution is 257×257. Dataset2 contains the 5-WBCs healthy classes. It had been acquired from 100 microscopic slides with magnification lens 100X [9]. Such images were acquired with a resolution 720×576 pixels and saved in BMP format, where each WBC image resolution is 150×150. Dataset3 [19], contains the 5-WBCs healthy classes. It consists of 320 images at $20 \times$ magnification. Each WBC image resolution is 70×70 and saved in TIF image format. The differential counts of these different WBCs classes for all datasets used are shown in Table 2.

The percentage of each WBC type in both Dataset1 and Dataset3 is relative to each WBC percentage found in the blood. As an example, the percentage of basophils is 0–1% in the blood. However, Dataset2 is the most balanced used dataset. Therefore, the Dataset_ ALL (combination of Dataset1, Dataset2, and Dataset3) suffer from unbalancing between the 5 classes. Unbalancing in training data is considered a real challenge and should be overcome. Usage of different datasets acquired from many sources with total 2551 cases is one of our strategies to overcome such problem.

During our experiments, algorithms are implemented using the Matlab 2016b [44]. The system platform containing Quad-Core 2 GHz Intel i7 with 8GB RAM. The GPU computation is done through NVidia-tesla with compute capability 3.5 and 12GB RAM.

The performance of traditional identification systems is evaluated using several evaluation parameters such as accuracy, specificity, sensitivity…..etc. For the proposed system, its performance is evaluated using similar metrics as in [45]. The overall system accuracy is obtained based on 10-folds cross-validation technique. We use the confusion matrix and its influenced results: TP ≡ True positive, TN ≡ True negative, FP ≡ False positive, FN ≡ False negative. In [23], the average of several evaluation parameters are computed to evaluate the system performance, such parameters can be calculated as follows:

$$Sensitivity = \frac{TP}{TP + FN} \tag{7}$$

$$Specifity = \frac{TN}{TN + FP} \tag{8}$$

$$Positive\ predictve\ value\ (PPV) = \frac{TP}{TP + FP} \tag{9}$$

$$Negative\ predictive\ value\ (NPV) = \frac{TN}{TN + FN} \tag{10}$$

Our proposed WBCsNet is also evaluated from CNN features visualization. The features extracted from any input image may be good (feature has a power of discrimination) or bad (features has a weak discrimination power), and this will be an important evaluation metric of biomedical image classification tasks [33].

4.1. Classification results

In this section, we present the experiments for our proposed methodologies; TLA approach based on DeCAF and fine-tuned network, WBCsNet architecture approach, and WBCsNet as a pre-trained network. In TLA experiments, we investigate which pre-trained network may achieve higher accuracy, and then we investigate how network accuracy is affected by using additional FS stage. For the previous experiments, we employ Dataset_ALL (2551 images) for training and testing. In WBCsNet architecture experiments, we investigate its superior performance. For that, we utilize each dataset separately with different sizes and then Dataset_ALL with different resolutions. To evaluate the performance of WBCsNet with respect to both fine-tuned networks and the traditional identification system, we employ Dataset3. As the CNN features visualization is considered as an evaluation parameter, we visualize the extracted deep features of WBCsNet at different depths inside the network and compare with other pre-trained network. Finally, the WBCsNet are employed to identify limited size dataset as a

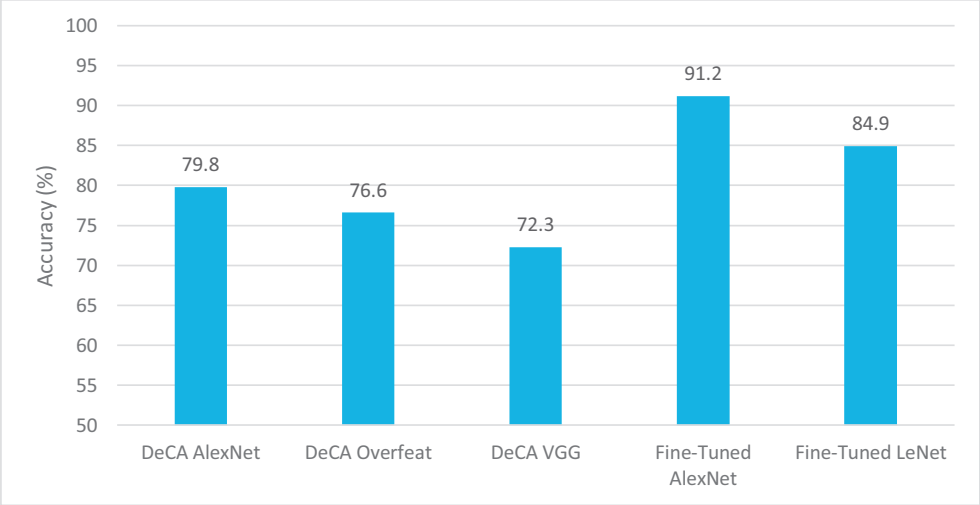

Fig. 5. Classification accuracy obtained from different networks using TLA approach.

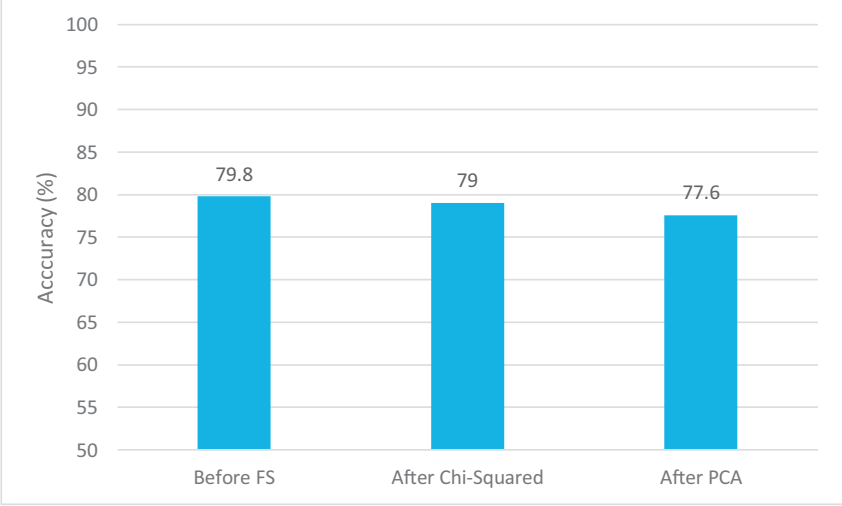

Fig. 6. Classification accuracies before and after employing different FS algorithms for DeCAF of AlexNet.

specialized pre-trained network. We utilize Dataset2 which is the only available small size and balanced dataset in the literature.

In our first experiment, general pre-trained networks for image classification and visual recognition applications are employed. The classification accuracy over different types of pre-trained CNNs is compared. Such networks are AlexNet, OverfeatNet, and VGGNet. As shown in Fig. 5, the highest accuracy (79.8%) is obtained using AlexNet. The OverfeatNet has a lower accuracy (76.6%), and the VGGNet achieved the lowest accuracy score with 72.3%.

In the second experiment, the effect of using FS on the final features pool extracted from AlexNet is investigated. Using the chi-squared algorithm reduces the features pool size from 4096 to only 187 features, using PCA algorithm reduces the feature pool size from 4096 to 127 features. The classification accuracies before and after employing different FS algorithms are shown in Fig. 6. Classification accuracy is found slightly degraded after reduction using chi-squared technique from 79.8% to 79% and using PCA technique from 79.8% to 77.6%.

For fine-tuning experiments, we test the performance of both fine-tuned AlexNet and LeNet-5. Fig. 5 illustrates that the accuracy achieved by fine-tuning of AlexNet (91.2%) is found more than DeCA approach (79.8%). The fine-tuning of LeNet-5 achieved lower accuracy (84.89%) than fine-tuned AlexNet.

For the image input layer size selection experiment, we investigate the effect of image input layer resizing as a pre-processing stage on the performance of WBCsNet. For this experiment, we employ Dataset3 which has the majority of images inside the utilized dataset with a resolution $70 \times 70 \times 3$. The WBCsNet with image input layer size $(50 \times 50 \times 3)$ achieves the highest accuracy (96.1%), however, the high down-sampling to $(30 \times 30 \times 3)$ decreases the classification accuracy to (89.4%) as shown in Fig. 7. According to this experiment, the image input layer size $(50 \times 50 \times 3)$ is selected to be the image input layer size of WBCsNet.

In the next experiment, we present the performance of our proposed system WBCsNet. We investigate both the effect of each dataset size and the image resolution scaling effect on the system performance. The classification accuracy of the WBCs using different datasets are shown in Fig. 8. Firstly, we utilize each dataset separately Dataset1 (125 images), Dataset2 (254 images), and Dataset3 (2172 images) to train and test WBCsNet. Dataset3 achieved the highest accuracy (96.1%), however, the Dataset1 achieved the lowest accuracy (29.2%) and Dataset2 achieve (62.6%). Fig. 8 also illustrates the effect of image scaling on the WBCsNet using Dataset_ALL which has different image resolutions from different datasets, the classification accuracy is little degraded with 92.9%.

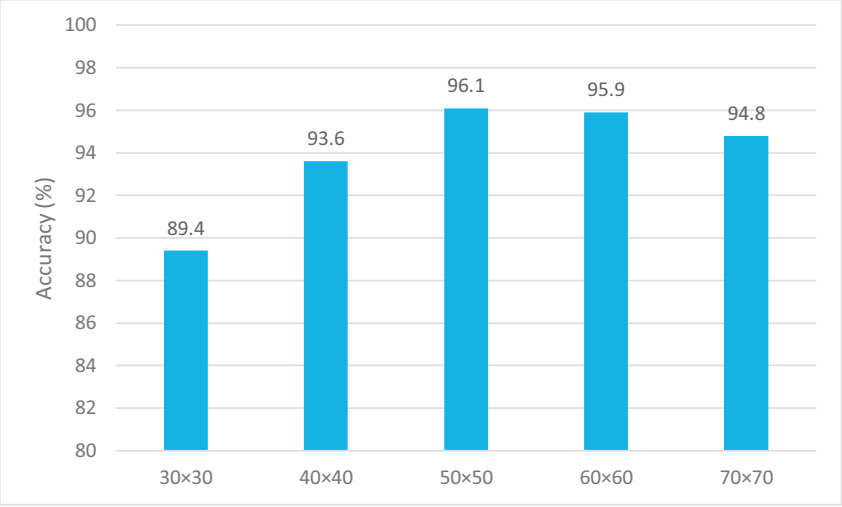

Fig. 7. Classification accuracy of WBCsNet with different image input layer sizes.

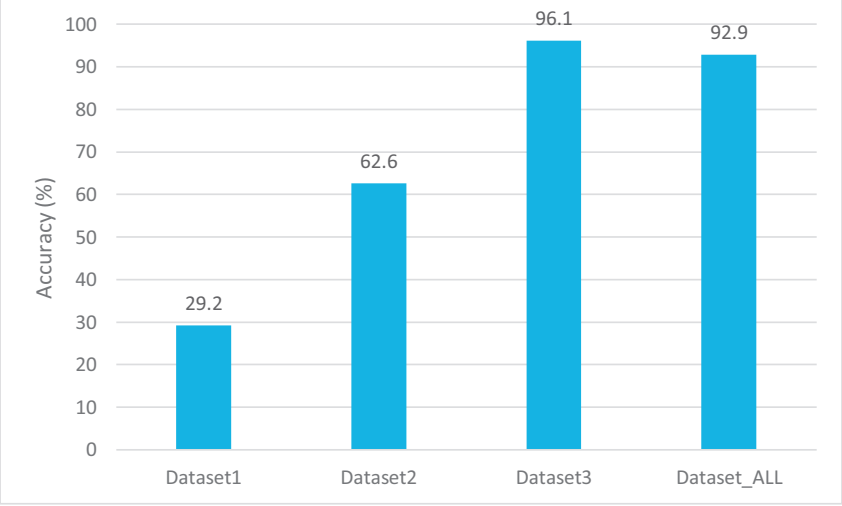

Fig. 8. Classification accuracy of WBCsNet using different datasets.

In the following experiment, we compare between different DL approaches and the traditional WBCs identification system which utilized the same dataset (Dataset3) as shown in Fig. 9. DL approaches are as follow; fine-tuned LeNet-5, fine-tuned AlexNet and the proposed WBCsNet. The traditional system was presented in [19] and consisted of pre-processing stage which contains many color space transformations, two steps of WBCs isolation and segmentation, and finally two steps of classification. The previous traditional system achieved 93.9%, however, our proposed WBCsNet achieved the highest accuracy 96.1%. The WBCsNet has also achieved higher accuracy than both fine-tuned LeNet-5 (88.7%) and AlexNet (92.5%).

One of our contributions in this paper, that we re-employed the WBCsNet to identify small WBCs size dataset (Dataset2) as a pre-trained network. On the other hand, the pre-trained AlexNet is also employed as a feature extractor for the same dataset. The classification accuracy is shown in Fig. 10 before and after FS stage. The FS procedure using chi-squared FS technique is also investigated for WBCsNet as a pre-trained network as shown in Fig. 10. The WBCsNet performance as a pre-trained network achieved higher accuracy (93.4%) than the pre-trained AlexNet (77.6%). After FS procedure using chi-squared technique for the WBCsNet features, the accuracy increase significantly from 93.4% to 94.6%. On the other hand, the pre-trained AlexNet after FS procedure achieve lower accuracy (75.2%) than the WBCsNet as a pre-trained network.

5. Discussion

In this section, we present a concise analysis and contrast our DL approaches. Firstly, we discuss the quantitative measurements which achieved by different DL proposed systems such as classification accuracy, confusion matrix and several general identification system evaluation parameters. Secondly, we discuss the qualitative measurements which are related to the visualization of CNN features.

5.1. Classification results discussion

In our first experiment that investigates the DeCA approach; the highest accuracy (79.8%) is obtained using AlexNet. This may be due to the nature of previous images during the pre-training. In the second experiment, the used FS algorithms success in reducing the features with about 95% which decrease the training and testing times. On the other hand, the classification accuracy after the chi-squared FS technique is higher than PCA technique. Our study

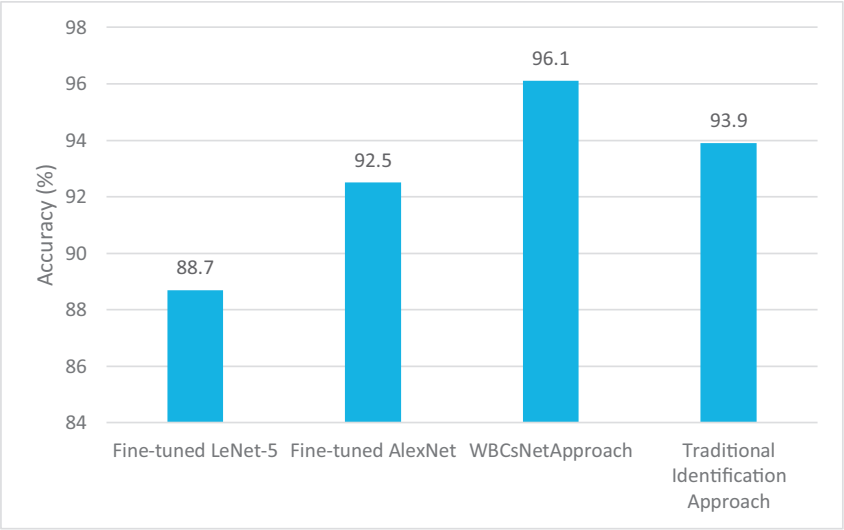

Fig. 9. Classification accuracy of DL approaches vs. the traditional identification approach.

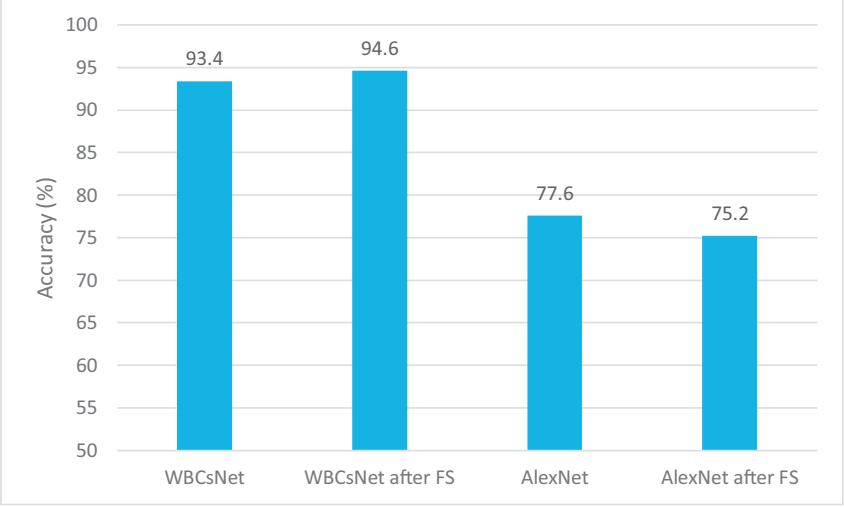

Fig. 10. Accuracy obtained before and after FS using AlexNet and WBCsNet as feature extractors.

provides additional support for the power of chi-squared feature selector.

The performance of fine-tuned AlexNet is illustrated in (Fig. 5). It achieves higher accuracy than the DeCA approach. Such response is logic and expected since in fine tuning approach, the network structure parameters are adjusted to suit the target dataset. On the other hand, the fine-tuned LeNet-5 achieves lower accuracy than the fine-tuned AlexNet. The fine-tuned LeNet-5 performance is reduced according to several reasons such as; its low size of image input layer and characteristics of each convolution layer (no .of filter banks, size of filters, padding size…..etc.). We observed that, the performance of each deep network architecture can be adapted with each problem solving.

From the results shown in Fig. 7, the classification accuracy severely degrades with high down-sampling. We conclude that we cannot generalize that the high down-sampling will always keep its benefits. On the other hand, for small size datasets, the system performance is sensitive to the image input layer size.

Fig. 8 illustrates the performance of the proposed WBCsNet when tested using different datasets. Dataset3 which has images with resolution 70×70 achieves the highest accuracy score with 96.1%. The low classification accuracy resulted from using Dataset1 confirms the fact that CNN network training cannot be established with small unbalanced datasets. On the other hand, Dataset2 achieves higher accuracy than Dataset1 since Dataset2 is more balanced and contains more images. Fig. 8 also illustrates that down-sampling of both Dataset1 (257×257) and Dataset2 (150×150) to size of image input layer (50×50) degrades the classification accuracy.

In Fig. 9, the WBCsNet achieves accuracy value with 96.1% higher than both fine-tuning approaches (LeNet-5, AlexNet) and the traditional identification system. This result reflects the super capabilities of our proposed network over traditional method. Our proposed WBCsNet has the advantages of low depth like LeNet-5 and its final layers are well-arranged like AlexNet.

Table 3 shows the confusion matrix of the proposed system for WBCs identification. The proposed system reflects high true positive value for identifying Neutrophil between different cells. Although Basophil has low sample size in Dataset3, the proposed system achieves acceptable true positive value for identifying it. For the others cells (eosinophil, lymphocyte, monocyte), it's noticed that the error percentage decreases with the increase of sample size. We observed that there is misclassification between some of WBCs classes as seen in Table 3. The misclassification between

Table 3
Confusion matrix for WBCsNet performance.

	Rec.Neutrophil	Rec.Eosinophil	Rec.Basophil	Rec.lymphocyte	Rec.Monocyte
Neutrophil	1382	2	0	28	0
Eosinophil	8	74	0	1	0
Basophil	0	0	9	1	0
Lymphocyte	0	15	1	502	7
Monocyte	2	0	0	19	121

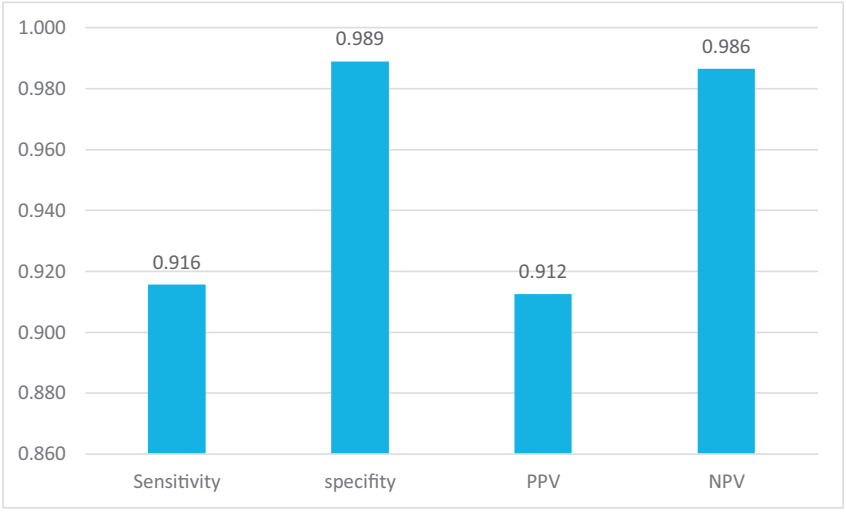

Fig. 11. Several evaluation parameters for the WBCsNet.

monocyte and lymphocyte is related to the similarity of their nucleus' texture and the weak intensities of monocytes' cytoplasm. This also represents a challenge for the expertise pathologist. However, the misclassification between eosinophil and neutrophil can be explained by the similarity of their segmented nuclei. We suppose that the overall system performance can be enhanced with more balanced dataset.

In Fig. 11, the calculated parameters for our proposed system between WBCs classes are illustrated. It achieves high score of specificity 98.9%, NPV 98.6% and PPV 91.2%. The average sensitivity value between WBCs classes is 91.6%, this low percentage can be explained by the lower sensitivity achieved by Basophil which is related to its very low percentage in the dataset.

Fig. 10 shows that the feature extractor based on WBCsNet can be utilized for limited WBCs dataset. The pre-trained WBCsNet achieved a classification accuracy of 93.4%, which is higher than the classification accuracy achieved by the general imagery pre-trained AlexNet. Our previous trials of using the general imagery as a feature extractor do not achieve high accuracy. The specific pre-trained network which is trained on the same type of data achieves more accuracy score. Moreover, the effect of using FS on the final features pool extracted from WBCsNet is studied. The chi-squared technique reduces the feature pool dimensions from 512 to 93 features and the classification accuracy increase with 1.2%.This is opposite to the response of AlexNet to FS procedure. This reflects that FS works better with features extracted from WBCsNet than AlexNet, and FS from specific kind dataset will be more efficient than the general images datasets.

5.2. CNN features discussion

One of the most interest evaluation parameters in DL is the CNN features visualization. CNN features are the learned filters which are created during the training process. Good appearance of such features reflects the performance of DL network [33]. Each input image (WBC type) activate the network layers and then produces a different response. The first convolutional layer contains the basic image features (colors, blobs, edges…etc). CNN features of the different DL networks are shown in Fig. 12. Before the activation of the pre-trained AlexNet network, CNN features are shown in Fig. 12(a) which represent the generally images features. After the activation of AlexNet with WBCs images, CNN features are shown in Fig. 12(b). The change in the appearance of CNN features is noticed, however, the activation of the pre-trained network with target dataset does not reflect a high response to WBCs features (shapes, colors….etc.). In Fig. 12(c), the proposed WBCsNet learned filters are shown. It is noticed that our proposed WBCsNet has lower dimension than the other pre-trained network. Furthermore, our filters describe WBCs appearance with various color descriptors, this is clearly defined through the color of each WBC type appeared in WBCsNet kernels.

In deeper convolutional layers, the learned filters features are more complicated. As the AlexNet and WBCsNet have different network depths, the comparison between each similar convolutional layer is not fair. However, the last convolutional layer should contain the final kernel responses in the network, and it should contain the clearest features that the network had been learned. We choose the Neutrophil cell to study its response to both networks. In Fig. 13, we choose the highest positive and negative activated kernels on both networks. It's noticed that the proposed WBCsNet features have reasonable responses for cell and nuclei of the input WBC as shown in Fig. 13(a). On the other hand, the pre-trained activated AlexNet does not obtain a reasonable meaning of the input WBC as shown in Fig. 13(b).

6. Conclusion

In this paper, we employed DL for WBCs identification in the blood smear images based on multi-approaches. Firstly, a complete TLA framework for WBCs identification is proposed based

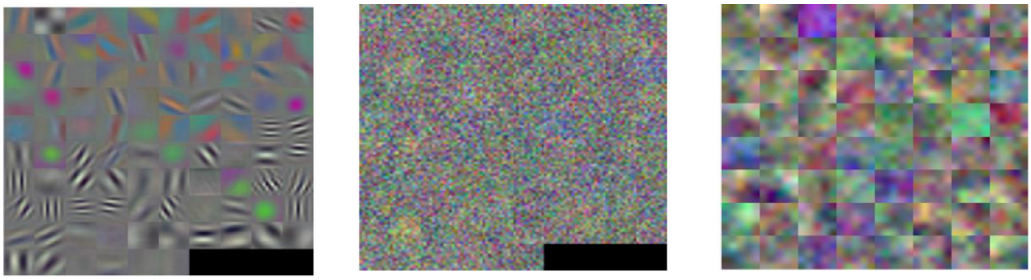

Fig. 12. A comparison of the deep convolutional kernels from the first convolutional layers of the deep models. Both the (a) pre-trained AlexNet and (b) Activated AlexNet has 96 kernels, (c) our proposed WBCsNet has 64 kernels.

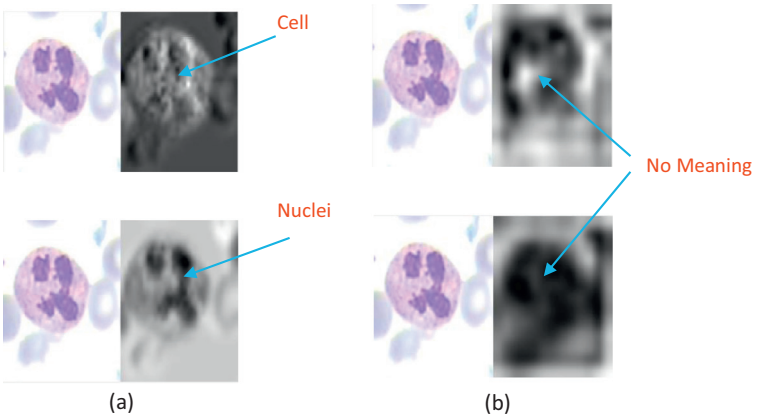

Fig. 13. A comparison of the deep convolutional kernels from the last layer. (a) (top-left) the highest positive activated kernel in WBCsNet, (bottom-left) the highest negative activated kernel in WBCsNet. (b) (top-right) the highest positive activated kernel in the pre-trained AlexNet, (bottom-right) the highest negative activated kernel in the pre-trained AlexNet.

on DeCA and fine-tuned approaches. In DeCA approach, the pre-trained ALexNet achieved the high classification accuracy with 79.8% as a feature extractor. The classification accuracy slightly decreased (79%) after the feature reduction. On the other hand, the chi-squared FS technique works better than PCA FS technique with the deep features. Secondly, fined-tuned networks are employed for WBCs identification and achieve higher accuracy than the DeCA approach. Finally, WBCsNet architecture is proposed which is the first trained CNNs for WBCs identification. It achieves the highest accuracy score with 96.1% for Dataset3 and 92.9% for Dataset_All. The WBCsNet has also achieved higher accuracy than both fine-tuned LeNet-5 (88.7%) and AlexNet (92.5%). The WBCsNet successes as a pre-trained network for WBCs identification to classify a balanced limited WBCs dataset. It achieves a classification accuracy (93.4%) which is higher than the previous general imagery pre-trained network. The WBCsNet activation features are visualized and reflect higher reasonable meaning than the pre-trained AlexNet activated network. The experimental results are very promising for applying WBCsNet in WBCs disorders like leukemia. The system performance can be increased with increasing the dataset size. The proposed system is able to be applied in a fully-automated system for WBCs identification.

Conflict of interest

We declare that we have no financial and personal relationships with other people or organizations that can inappropriately influence our work, there is no professional or other personal interest of any nature or kind in any product, service and/or company that could be construed as influencing the position presented in, or the review of, the manuscript entitled, “White Blood Cells Identification System Based on Convolutional Deep Neural Learning Networks”.

References

[1] M.A. Lichtman, Williams Manual of Hematology, Eighth ed., McGraw-Hill Education, 2011.
[2] “labtestsonline,” [Online]. Available: https://labtestsonline.org/understanding/analytes/blood-smear/tab/test/. [Accessed 28 12 2016].
[3] “CellaVision.com,”; CellaVision, [Online]. Available: http://www.cellavision.com. [Accessed 25 12 2016].
[4] “VisionHema.com,”; Vision Hema, [Online]. Available: http://wm-vision.com/en/. [Accessed 25 12 2016].
[5] “medicacorp.com”; Sysmex, [Online]. Available: http://www.medicacorp.com/products/hematology-imaging-analyzers/. [Accessed 25 12 2016].
[6] J. Wu, P. Zeng, Y. Zhou, C. Olivier, A novel color image segmentation method and its application to white blood cell image analysis, 8th International Conference on Signal Processing, 2006.
[7] Q. Wang, L. Chang, M. Zhou, Q. Li, H. Liu, F. Guo, A spectral and morphologic method for white blood cell classification, Opt. Laser Technol. 84 (2016) 144–148.
[8] G. Abedini, M. Firouzmand, A. Razavi, Recognition and counting of WBCs using wavelet transform, Int. J. Emerg. Trends Technol. Comput. Sci. 2 (5) (2013).
[9] S.H Rezatofighi, H. Soltanian-Zadeh, Automatic recognition of five types of white blood cells in peripheral blood, Comput. Med. Imag. Graph. 35 (4) (2011) 333–343.
[10] S. Dodge, L. Karam, Understanding how image quality affects deep neural networks, Quality of Multimedia Experience (QoMEX), 2016 Eighth International Conference on, IEEE, 2016.
[11] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, Proc. IEEE 86 (18) (1998) 2278–2324.
[12] J. Redmon, S. Divvala, R. Girshick, A. Farhadi, You only look once: unified, real-time object detection, Comput. Vis. Pattern Recognit. (2015).
[13] C. Szegedy, W. Liu and Y. Jia, "Going deeper with convolutions," arXiv [cs.CV], 2014.
[14] D. Hoiem, Y. Chodpathumwan, Q. Dai, Diagnosing error in object detectors, Computer Vision–ECCV 2012, 2012.
[15] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, Y. LeCun, Overfeat: integrated recognition, localization and detection using convolutional networks, International Conference on Learning Representations (ICLR 2014), 2014.

[16] F. Hu, G.S. Xia, J. Hu, L. Zhang, Transferring deep convolutional neural networks for the scene classification of high-resolution remote sensing imagery, Remote Sens. 7 (11) (2015) 14680–14707.
[17] C. Spampinato, S. Palazzoe, D. Giordanoe, M. Aldinuccie, R. Leonardie, Deep learning for automated skeletal bone age assessment in X-ray images, Med. Image Anal. 36 (2017) 41–51.
[18] J. Prinyakupt, C. Pluempitiwiriyawej, Segmentation of white blood cells and comparison of cell morphology by linear and naïve Bayes classifiers, Biomed. Eng. Online 14 (1) (2015) 1:63.
[19] N. Ramesh, B. Dangott, M.E. Salama, T. Tasdizen, Isolation and two-step classification of normal white blood cells in peripheral blood smears, J. Pathol. Inform. 3 (1) (2012) 3–13.
[20] L. Putzu, Caocci, C. Di Ruberto, Leucocyte classification for leukaemia detection using image processing techniques, Artif. Intell. Med. 3 (62) (2014) 179–191.
[21] S. Nazlibilek, D. Karacor, T. Ercan, M.H. Sazlı, O. Kalender, Y. Ege, Automatic segmentation, counting, size determination and classification of of white blood cells, Measurement 55 (2014) 58–65.
[22] A. Mathur, A.S. Tripathi, M. Kuse, Scalable system for classifcation of white blood cells from Leishman stained blood stain images, J. Pathol. Inform. 4 (2013).
[23] P. Ghosh, D. Bhattacharjee, M. Nasipuri, Blood smear analyzer for white blood cell counting: a hybrid microscopic image analyzing technique, Appl. Soft Comput. 46 (2016) 629–638.
[24] M.C. Su, C.Y. Cheng, P.C. Wang, A neural-network-based approach to white blood cell classification, Sci. World J. (2014) 1–9.
[25] X. Gao, W. Li, M. Loomes, L. Wang, A fused deep learning architecture for viewpoint classification of echocardiography, Inform. Fus. 36 (2016) 103–113.
[26] Q. Zhang, Y. Xiao, W. Dai, J. Suo, C. Wang, J. Shi, H. Zheng, Deep learning based classification of breast tumors with shear-wave elastography, Ultrasonics 72 (2016) 150–157.
[27] H.K.V.D. Burgha, R. Schmidta, H.J. Westenenga, M.A. de Reusb, L.H.V.D. Berga, M.P. Heuvel, Deep learning predictions of survival based on MRI in amyotrophic lateral sclerosis, NeuroImage 13 (2017) 361–369.
[28] N. Liu, M. Zhang, H. Li, Z. Sun, T. Tan, DeepIris: learning pairwise filter bank for heterogeneous iris verification, Pattern Recognit. Lett. 82 (2016) 154–161.
[29] J. Donahue, Y. Jia, O. Vinyals, J. Hoffman, N. Zhang, E. Tzeng, T. Darrell, DeCAF: a deep convolutional activation feature for generic visual recognition, Icml 2014, 2014.
[30] A.S. Razavian, H. Azizpour, J. Sullivan, S. Carlsson, Cnn features off-the-shelf: an astounding baseline for recognition, in: Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition Work-shops CVPRW, 2014.
[31] S. Jialin Pan, Q. Yang, A survey on transfer learning, IEEE Trans. Knowl. Data Eng. 22 (10) (2010).
[32] R. Arroyo, P.F. Alcantarilla, L.M. Bergasa, E. Romera, Fusion and binarization of CNN features for robust topological localization across seasons, IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2016.
[33] S. Pang, Z. Yu, M.A. Orgun, A novel end-to-end classifier using domain transferred deep convolutional neural networks for biomedical images, Comput. Methods Programs Biomed. 140 (2017) 283–293.
[34] X.W. Gao, R. Hui, Z. Tian, Classification of CT brain images based on deep learning networks, Comput. Methods Programs Biomed. 138 (2017) 49–56.
[35] T. Kooi, G. Litjens, B.V. Ginneken, AG. Mérida, C.I. Sánchez, R. Mann, A. Heeten, N. Karssemeijer, Large scale deep learning for computer aided detection of mammographic lesions, Med. Image Anal. 35 (2017) 303–312.
[36] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, Adv. Neural Inf. Process. Syst. (2012) 1097–1105.
[37] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going deeper with convolutions, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015.
[38] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, CoRR, 2014 abs/1409.1556.
[39] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, A.C. Berg, Imagenet large scale visual recognition challenge, Int. J. Comput. Vis. 115 (3) (2015) 211–252.
[40] Y. Sun, D. Liang, X. Wang, X. Tang, DeepID3: face recognition with very deep neural networks, Comput. Vis. Pattern Recognit. (2015) arXiv:1502.00873.
[41] K.M. Amin, A. Shahin, Y. Guo, A novel breast tumor classification algorithm using neutrosophic score features, Measurement 81 (2016) 210–220.
[42] C. Bishop, Pattern Recognition and Machine Learning (Information Science and Statistics), 1st edn. 2006. corr. 2nd Printing Ed. Springer, New York 2007.
[43] R.D. Labati, V. Piuri, F. Scotti, All-IDB: the acute lymphoblastic leukemia image database for image processing, 18th IEEE International Conference on Image Processing, 2011.
[44] "Mathwork.com,"; Mathwork, [Online]. Available: https://www.mathworks.com/.
[45] S.M. Smits, A. Leyte, "Clinical performance evaluation of the cellavision image capture system in the white blood cell differential on peripheral blood smears, J. Clin. Pathol. 67 (2) (2014) 168–172.