Engineering Applications of Artificial Intelligence 72 (2018) 415–422

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

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

Leukemia diagnosis in blood slides using transfer learning in CNNs and SVM for classification

Luis H.S. Vogado *, Rodrigo M.S. Veras, Flavio. H.D. Araujo, Romuere R.V. Silva, Kelson R.T. Aires

Federal University of Piauí, Teresina, Brazil

ARTICLE INFO

Keywords:
Medical images
Leukemia diagnosis
Convolutional neural networks
Transfer learning

ABSTRACT

Leukemia is a pathology that affects young people and adults, causing premature death and several other symptoms. Computer-aided systems can be used to reduce the possibility of prescribing inappropriate treatments and assist specialists in the diagnosis of this disease. There is a growing use of Convolutional Neural Networks (CNNs) in the classification and diagnosis of medical image problems. However, the training of CNNs requires a large set of images. To overcome this problem, we use transfer learning to extract images features for further classification. We tested three state-of-the-art CNN architectures and the features were selected according to their gain ratios and used as input to the Support Vector Machine classifier. The proposed methodology aims to correctly classify images with different characteristics derived from different image databases and does not require a segmentation process. We built a new database from the union of three distinct databases presented in the literature to validate the proposed methodology. The proposed methodology achieved hit rates above 99% and outperformed nine methods found in the literature.

1. Introduction

Over the years, multiple medical aid systems have been proposed. Among the diseases aided by computer systems, leukemia is the one that has the highest number of fatalities among adolescents and children, and the risk of developing it is higher in children up to five years of age. Leukemia is a type of cancer that originates in the bone marrow (Fig. 1a) and causes abnormal proliferation of white blood cells (Fig. 1b). To diagnose leukemia, specialists can carry out various tests and exams, including physical examinations, blood tests, blood counts, myelograms, lumbar punctures and bone marrow biopsies. Microscopic analysis is the most economical method of carrying out the initial screening of patients with leukemia. This type of test is done manually, which may generate fatigue in operators. Therefore, a low-cost system that is automatic and robust is necessary to avoid the operator's influence.

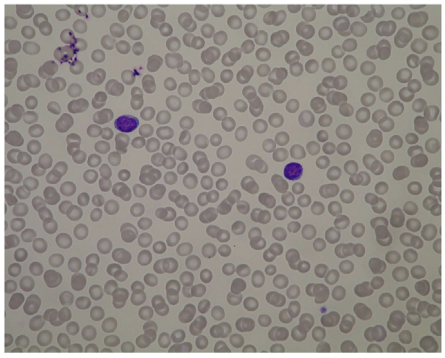
Many computer-aided diagnosis systems were developed with the use of image processing and computational intelligence techniques. These systems usually have some steps such as: preprocessing, segmentation, feature extraction, and classification. Feature extraction and classification are the steps that best define the diagnosis performed by computer-aided diagnosis systems. However, to achieve better results, a proper segmentation can provide an adequate feature extraction and consequently a reasonable classification.

In this work, we propose a leukemia diagnosis system that does not require the segmentation process (commonly used in state-of-the-art techniques). The methodology uses pre-trained Convolutional Neural Network (CNN) models (AlexNet Krizhevsky et al., 2012, Vgg-f Chatfield et al., 2014 and CaffeNet Jia et al., 2014) to extract features directly from the images without any previous preprocessing. Then, the obtained features will be used for the following classification with a Support Vector Machine (SVM) (Cortes and Vapnik, 1995). We used three hybrid datasets to evaluate the performance of the methodology, one with blood smears containing only one leukocyte per image, one with many leukocytes per image, and the last one with both types of images. To demonstrate the robustness of our approach, we compared the results obtained by our methodology with other state-of-the-art methods.

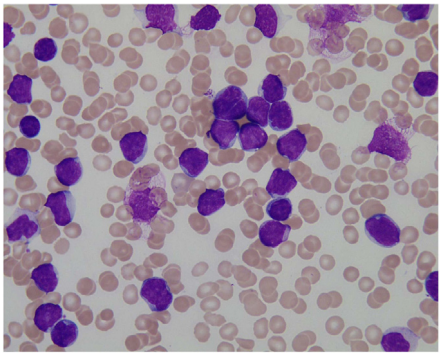
The remainder of the paper is organized as follows: Related works are presented in Section 2, and the proposed method is introduced in Section 3. In Section 4, we present the experiments and also describe the image datasets used in the tests and the evaluation of the method. We discuss the results in Section 5. Finally, the conclusions and perspectives on future works are given in Section 6.

* Correspondence to: Universitária Avenue, Postal code: 64049-550, Brazil.
E-mail addresses: lhvogado@gmail.com (L.H.S. Vogado), rveras@ufpi.edu.br (R.M.S. Veras), flavio86@ufpi.edu.br (F.H.D. Araujo), romuere@ufpi.edu.br (R.R.V. Silva), kelson@ufpi.edu.br (K.R.T. Aires).

https://doi.org/10.1016/j.engappai.2018.04.024
Received 18 October 2017; Received in revised form 14 April 2018; Accepted 20 April 2018
Available online 8 May 2018



(a)



(b)

Fig. 1. (a) ALL-IDB1 leukemic blood smear sample and (b) ALL-IDB1 non-leukemic blood smear sample.

2. Related works

Several methods for leukemia detection have been proposed over the years and some of these works presented solutions to the classification of the two most common types of leukemia: Acute Myeloid Leukemia (AML) and Acute Lymphoblastic Leukemia (ALL). Some works presenting state-of-the-art technologies only provide diagnosis by using images with one leukocyte per image (Mohaprata et al., 2011; Mohapatra et al., 2014; Neoh et al., 2015; Putzu et al., 2014; Singhal and Singh, 2016) and others that provide diagnosis by using images with multiple leukocytes (Madhukar et al., 2012; Vincent et al., 2015; Patel and Mishra, 2015; Agaian et al., 2018; Singh et al., 2017).

Table 1 present some works on leukemia detection and highlights the techniques of feature extraction and classification used. As can be seen in this table, most of the listed approaches combine texture (Haralick's features Haralick et al., 1973 and Local Binary Pattern (LBP) chen He and Wang, 1990); shape (area, perimeter, compactness, Fractal Dimension Pentland, 1984...) and color features (mean, standard deviation, uniformity, entropy...).

In the classification process, the most frequently classifiers used were: SVM, Multilayer Perceptron (MLP) (Popescu et al., 2009), Random Forest (RF) (Breiman, 2001), K-Nearest Neighbors (KNN) (Aha and Kibler, 1991), Radial Basis Function Network (RBFN) (Schwenker et al., 2001) and Naive Bayesian (NB) (Friedman et al., 1997).

The works presented in Table 1 achieves accuracy rates above 90%. However, they depend on the use of a proper segmentation process and do not show good performance when applied to databases with different characteristics. Recent works reported that the use of deep learning is producing better results than all classical techniques for different types of medical image applications (Castelluccio et al., 2015; Wang et al., 2016; Kumar et al., 2016). A large set of images is required to train a Convolutional Neural Network (CNN), but the available leukemia image databases are small and limited. Thus, to avoid the problem of having a limited number of images, we investigate the use of learning transfer to extract images features and subsequent classification by SVM. It is worth mentioning that this approach does not need the use of the segmentation process and is more robust to the characteristic variation of the images.

Castelluccio et al. (2015) present two methods of transfer learning in their work. The first one consists in fine-tuning the network, where the structure is modified, freezing high-level layers. The other method includes the extraction of the last network fully connected layer, obtained from the input image (Athiwaratkun and Kang, 2015). Then, it uses another classifier in the classification process. In Athiwaratkun and Kang (2015), the authors demonstrated that the layers with activation localized near the output layer could usually extract the best characteristics.

3. Proposed methodology

The method proposed in this work aims to diagnose leukemia using blood smear images. Following the flowchart shown in Fig. 2, it is possible to observe that the system uses an image without any preprocessing or segmentation as input. This is the main difference between our method and the state-of-the-art methods. The CNNs are used to describe the input image and the features are selected and reduced. In the classification step, the SVM is used to classify the images as pathological or not.

3.1. Convolutional neural networks

Currently, researchers working with artificial intelligence use deep learning techniques to create robust computational systems. Among these techniques, CNNs (Fig. 3) have been proposed to aid in diagnosis, and they have outperformed conventional methods of extracting features, obtaining better accuracy rates (Lecun et al., 2015; Litjens et al., 2017). Therefore, we use the power of the CNNs to carry out the construction of our approach.

In recent works, authors have presented two different ways of using the power of CNNs. The first is the usual method in which the training is performed with a large set of data. The second is through the transfer of learning using pre-trained networks (Tajbakhsz et al., 2016). In our work, we used pre-trained CNNs in a large natural image database. Thus, the neural network can assimilate generic features, facilitating its application in small databases. This technique can be used in several types of tasks, for example in the extraction of features from face images, objects, and diseases. The success of the results depends on the similarity of the images from the database used to extract the features and the images from the training set (Shin et al., 2016).

We use CNNs pre-trained on a large natural database of images with 1000 categories, called ImageNet (Russakovsky et al., 2015). Among these categories, the database also contains images of lymphocytes and lymphoblasts. During the development of our approach, three models were selected: AlexNet (Krizhevsky et al., 2012), CaffeNet (Jia et al., 2014), and Vgg-f (Chatfield et al., 2014). These models have similar architectures, and the main difference between them is the size of the filters used in the convolutional layers and the number of neurons in the fully connected layers.

Alexnet: Krizhevsky et al. (2012) developed the architecture for the ILSVRC-2010 competition to carry out training and classification of ImageNet database. It comprises eight layers that need to be trained, five convolutional layers with filters with sizes of 5×5 and 7×7, followed by three fully-connected layers, as well as max-pooling layers.

Caffenet: This architecture was developed by the Berkeley Vision and Learning Center (BVLC) and is considered one of the most popular CNNs in deep learning (Jia et al., 2014). It comprises five convolutional layers, each followed by a pooling layer, and three fully-connected layers. The

Table 1
Related works.

Work	Year	Descriptor	Classifier	Images	Image type	A (%)
Mohaprata et al. (2011)	2011	Shape, texture and color features	SVM	108	Leukocyte	93.00
Mohapatra et al. (2014)	2014	Shape and texture features	Ensemble (RBFN + SVM + MLP + KNN + NB)	104	Leukocyte	94.73
Neoh et al. (2015)	2015	Shape, texture and color features	Dempster–Shafer (SVM + MLP)	180	Leukocyte	96.72
Putzu et al. (2014)	2015	Shape, texture and color features	SVM	267	Leukocyte	93.63
Singhal and Singh (2016)	2016	Texture features	SVM	260	Leukocyte	93.80
Madhukar et al. (2012)	2012	Shape, texture and color features	SVM	98	Slide	93.50
Vincent et al. (2015)	2015	Texture features	MLP	100	Slide	97.70
Patel and Mishra (2015)	2015	Shape, texture, statistical and color features	SVM	27	Slide	93.75
Agaian et al. (2016)	2016	Shape, texture and color features	SVM	98	Slide	>94.00
Singh et al. (2017)	2016	Shape and texture features	MLP	108	Slide	97.20

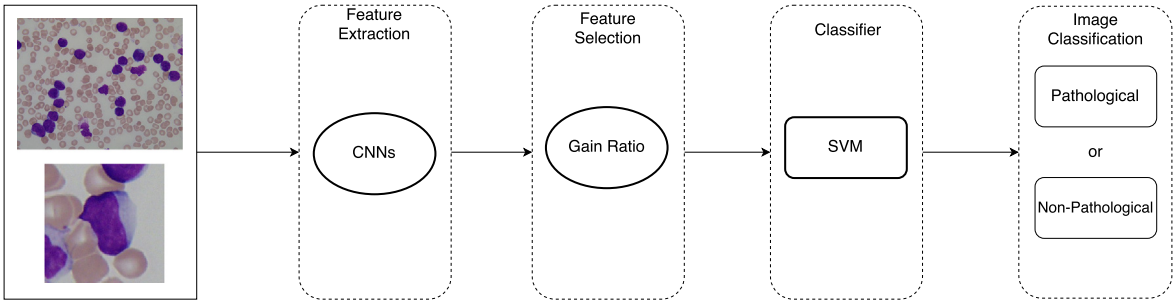


Fig. 2. Flowchart of the proposed leukemia diagnosis approach.

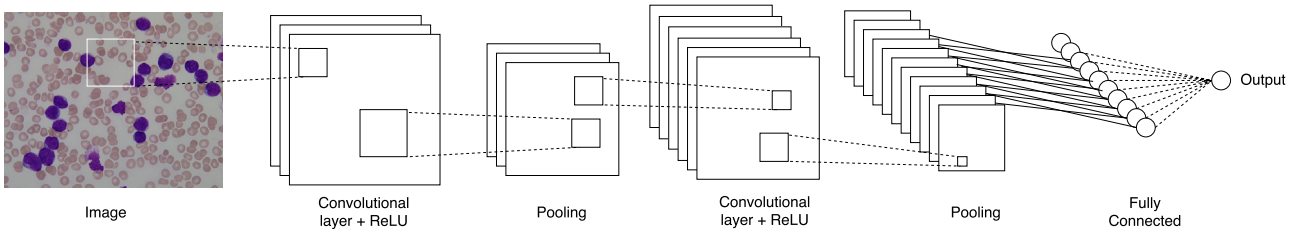


Fig. 3. A simplified illustration of the CNN architecture.

main difference between this architecture and AlexNet (Krizhevsky et al., 2012) is the order of the pooling layer and the normalization layer.

Vgg-f: This model was proposed in 2014 by Chatfield et al. (2014), and was based on AlexNet. This architecture was proposed with two others: Vgg-s and Vgg-m. The main difference between these three models is the number of layers and the size of the convolutional filters. The size of the filters influences the computing power demanded by the network. The smaller the filter, the less the network will require the computer. However, larger filters take advantage of more neighborhood information. The main difference between this architecture and AlexNet is that Vgg-f has fewer convolutional filters in the first, third, and fourth convolutional layers.

The number of features taken from the last fully connected layer of each architecture is 4096.

3.2. Feature selection

After generating the feature vector, we performed an attribute selection. This selection aims to eliminate unnecessary characteristics and consequently to simplify the prediction model, reduce the computational cost, and provide a better understanding of the results obtained. According to Guyon and Elisseeff (2006), attribute selection techniques are primarily employed to identify relevant attributes and essential information.

We choose the gain ratio algorithm to perform this selection. This algorithm calculates the information gain, which uses entropy as a measurement of the degree of impurity and decision trees to represent the attributes of the vector. Its performance is considered better than the information gain, so it provides better accuracy regarding the complexity of the generated trees. Let be S set consisting of s data samples with m distinct classes, Eq. (1) presents the information calculation used to classify a given sample.

$$Info(S) = -\sum_{i=1}^{m} p_i \log_2(p_i), \tag{1}$$

where p_i is the probability of that the sample belongs to a C_i class. The entropy calculation of a given attribute A that has v values is shown in Eq. (2):

$$Entropy(A) = -\sum_{i=1}^{m} Info(S)\frac{s_{1i} + s_{2i} + \cdots + s_{mi}}{s}, \tag{2}$$

and the variable s_{ij} represents the number of samples belonging the C_i class of the subset S_i. The attribute gain A is represented by Eq. (3).

$$Gain(A) = Info(S) - Entropy(A). \tag{3}$$

Eq. (4) presents the information value generated by dividing the data set of S into v partitions.

$$SInfo(S) = -\sum_{i=1}^{v} \left(\frac{|S_i|}{|S|}\right) \log_2 \left(\frac{|S_i|}{|S|}\right). \tag{4}$$

Finally, the gain ratio is defined as the result of dividing the solution to Eq. (3) by the solution to Eq. (4). The attributes are ranked according to the values of their gain ratios, and the n attributes with the highest values are selected.

To classify the attributes resulting from the feature selection, we chose the SVM. This classifier seeks to separate two classes of a dataset using a separation line or hyperplane by using the similarities between the features of each instance. Other classifiers were used in the tests, are: MLP, RF and KNN. In our tests, the SVM classifier obtained the best results. Therefore, we report the results only for this classifier.

4. Experiments

To evaluate the obtained results, we used two image databases, one containing only one leukocyte per image and the other with many leukocytes. Also, we used four classic metrics from the literature: precision, accuracy, recall, and the Kappa index. We implemented the feature extraction in MATLAB and the feature selection and classification using the WEKA tool (Hall et al., 2009).

4.1. Image database

Several authors have tested their methods only on homogeneous databases or private databases. However, one of the challenges in creating a medical assistance system is the ability to diagnose the disease in databases with distinct characteristics. Thus, in this work, we constructed three heterogeneous image databases based on traditional databases presented in the literature. Two of these databases are divided according to the number of leukocytes in the images. The Hybrid-Leukocyte database (Fig. 4) is formed from four databases that present only one leukocyte per image, namely ALL-IDB1 (Labati et al., 2011), ALL-IDB2 (Labati et al., 2011), Leukocytes (Sarrafzadeh and Dehnavi, 2015), and CellaVision (Rollins-Raval et al., 2012). The second database constructed is the Hybrid-Slide database (Fig. 5). This database contains many leukocytes per image and is also formed from four traditional databases, namely ALL-IDB1 (Labati et al., 2011), Atlas, and two databases proposed in Omid Sarrafzadeh and Rabbani (2014) and Sarrafzadeh et al. (2015). The third database is called Hybrid-Complete and is composed of the two hybrid databases presented previously. Tables 2 and 3 present the standard databases used to construct the Hybrid-Leukocyte and Hybrid-Slide databases as well as the number of pathological and non-pathological images and the total number of images in each database.

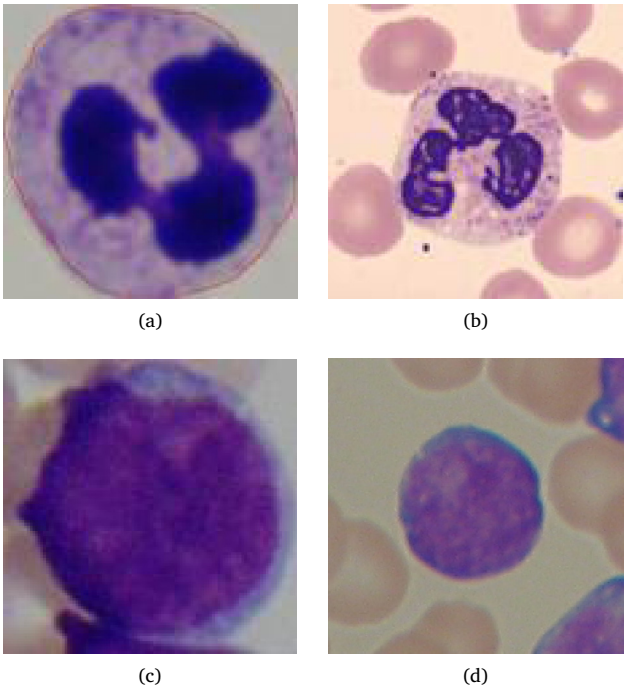


Fig. 4. Examples of leukocytes (a) and (b), and lymphoblasts (c) and (d) of the Hybrid-Leukocyte database.

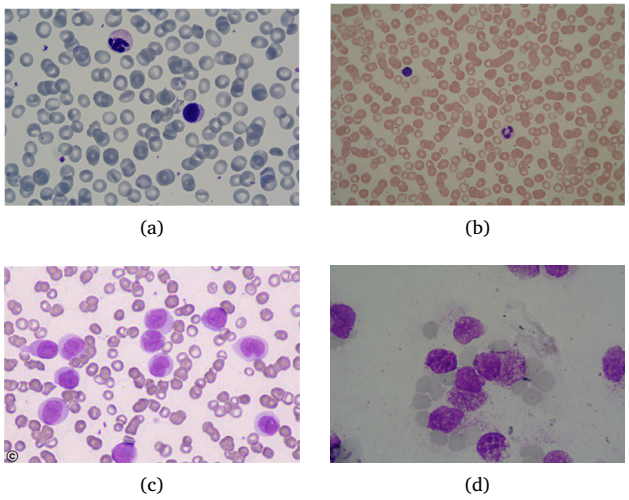


Fig. 5. Examples of leukocytes (a) and (b), and lymphoblasts (c) and (d) of the Hybrid-Slide database.

4.2. Metrics for performance evaluation

We used standard metrics based on a confusion matrix to evaluate the results. The confusion matrix is a table that shows the classification result. Here, it consists of four values: true positive (TP), false positive (FP), false negative (FN), and true negative (TN). From these values, we calculate the precision (P), recall (R), accuracy (A), and Kappa index (K) (Powers, 2011).

The Kappa index is used as an appropriate measure of accuracy because it represents the entire confusion matrix and takes all matrix elements into account, rather than just those that depict the number of correct classifications, which occur when calculating the overall classification accuracy. The quality levels indicated by the Kappa index were classified as bad ($K \leq 0.2$), reasonable ($0.21 \leq K \leq 0.4$), good ($0.41 \leq K \leq 0.6$), very good ($0.61 \leq K \leq 0.8$), and excellent ($K \geq 0.81$).

5. Results and discussions

We followed two approaches during the development of the proposed methodology: by using the outputs of each one of the three architecture separately and by using the concatenation of all vectors. We also evaluate their results individually, comparing them with existing state-of-the-art works. We made this evaluation because there are no studies that classify blood smears with multiples and with only one leukocyte in the same image, as the one proposed in this work.

We carried out empirical tests to define the dimensionality of the final vector, always searching for a compromise between accuracy and vector size. In these tests, we applied the gain ratio algorithm and assembled the vectors with the features ranked in descending order of gain ratio. Thus, to perform the experiments, we used the ranked features, starting from the one with the highest gain ratio and adding the others incrementally to complete 4096 features.

The experiments with vector concatenation were performed in the same way. However, due to the large number of features, we selected the first 4096 of the 12288. For verification of the results, we chose k-fold cross-validation, with the value of k being 5. We picked this amount because it presented a larger set of images to be tested, corresponding to 20% of the total for tests and 80% for training. In SVM we select the kernel radial basis function without optimization using the standard values obtained by WEKA.

Table 2
Databases used to build the Hybrid-Leukocyte database.

	ALL-IDB 1	ALL-IDB 2	Leukocytes	CellaVision	Total
Images	510	130	149	102	891
Non-pathological	0	130	149	102	381
Pathological	510	0	0	0	510

Table 3
Databases used to build the Hybrid-Slide database.

	ALL-IDB 1	Atlas	Omid Sarrafzadeh and Rabbani (2014)	Sarrafzadeh et al. (2015)	Total
Images	108	88	154	27	377
Non-pathological	59	0	154	0	213
Pathological	49	88	0	27	164

5.1. Hybrid-Leukocyte database

Fig. 6 illustrates the hit rates of the approach with selected vectors of different sizes for the Hybrid-Leukocyte database. We observed that the selected vectors with a dimensionality gain ratio of less than 500 already present high hit rates of above 99%. This demonstrates that this database does not require many characteristics to achieve a high accuracy rate. Consequently, as the size of the vector grows, these rates stabilize and stay around this percentage. Also, we can see that when using only 5% of the features, the proposed approach achieves excellent accuracy rates. This emphasizes the need for attribute selection, which provides more speed in the classification step. This speed helps provide a service to many people in a short period of time. This is also essential in medical aid systems.

As mentioned above, in this work, we look for the best balance between accuracy and attribute vector size. In other words, the smallest vector that could achieve reasonable success rates. Thus, Table 4 shows the hit rates and size vector chosen for each CNN.

We can observe, in Table 4, that Vgg-f and AlexNet presented the same success rates. Despite this, the confusion matrix of the two networks is different. The AlexNet obtained: TP = 509, FP = 1, TN = 379 and FN = 2 while the Vgg-f obtained: TP = 508, FP = 2, TN = 380 and FN = 1. Considering that AlexNet achieved the same rates with a shorter attribute vector than Vgg-f, we conclude that AlexNet is a better descriptor than Vgg-f. However, among all results, the concatenated and selected vector achieved an accuracy of 100.00% and was also the smallest vector among the four approaches. This is because the output of each network has an architecture that characterizes the image of the single leukocyte in different ways. Consequently, the concatenation of the vectors takes advantage of the characteristics extracted by the various filters existing in each CNN; in other words, through this concatenation more attributes of borders and texture are harnessed. This approach was chosen as the best among the four because it presented the best accuracy and greater robustness.

We performed the same tests for the MLP, RF, and KNN classifiers (with K ranging from 3 to 5). In Table 5, we present the obtained results; we found that SVM presented the best accuracy with the lowest number of features.

Table 6 shows a comparison of the accuracy of the proposed methodology with other state-of-the-art studies for the Hybrid-Leukocyte database. Analyzing this table, we can see the superiority of the results of the proposed methodology, demonstrating the robustness of the CNNs. Because it lacks the segmentation step, the proposed approach is not susceptible to errors during the extraction of features. In contrast, all the related works use different segmentation methods and are sensitive to errors when executed in heterogeneous databases. Another factor that demonstrates the robustness of our methodology is the presence of a heterogeneous and vast image database used in the experiments.

Table 5
Results of accuracy for each classifier in the Hybrid-Leukocyte database.

Classifiers	*A* (%) (81 features)
MLP	98.87
RF	98.87
KNN(3)	99.66
SVM	**100**

Table 6
Comparison of the accuracy obtained by the proposed methodology and other works presented in the literature for the Hybrid-Leukocyte database.

Work	Number of images	*A* (%)
Mohaprata et al. (2011)	108	93.00
Mohapatra et al. (2014)	270	94.73
Neoh et al. (2015)	180	96.72
Putzu et al. (2014)	267	93.63
Rawat et al. (2015)	260	93.80
Proposed methodology	891	**100.00**

5.2. Hybrid-slide database

Unlike the Hybrid-Leukocyte database, the Hybrid-Slide database has images with multiple leukocytes, representing the same pattern as the images used by doctors in hospitals. Fig. 7 shows the hit rates obtained by each of the attribute vectors selected with the gain ratio. AlexNet presented a low performance compared to other architectures. Meanwhile, CaffeNet reached the best rate of accuracy with the least attributes. Vgg-f was more consistent obtaining good results when compared to the architecture concatenation. Table 7 shows the results of the proposed methodology achieved for each of the different architectures. The Vgg-f and the concatenation of all the attributes of the CNNs presented better results; however, the Vgg-f size vector was three times smaller than that of the concatenated vector and was therefore the best choice. Based on the vector sizes presented in Tables 4 and 7, we observed that more features are required to classify the images with many leukocytes per image. In Table 8, we present the results of the proposed approach for each classifier. SVM and KNN achieve the

Table 4
Results of accuracy, precision, recall, and Kappa index and the size vector chosen for each CNN for the Hybrid-Leukocyte database.

	A (%)	*P* (%)	*R* (%)	*K*	Vector size
AlexNet	99.66	99.70	99.70	0.9931	290
CaffeNet	99.88	99.00	99.00	0.9977	84
Vgg-f	99.66	99.70	99.70	0.9931	773
AlexNet + CaffeNet + Vgg-f	**100**	**100**	**100**	**1**	**81**

Table 7
Results of accuracy, precision, recall, and Kappa index and the size vector chosen for each CNN for the Hybrid-Slide database.

	A (%)	*P* (%)	*R* (%)	*K*	Vector size
AlexNet	98.14	98.10	98.10	0.9623	1429
CaffeNet	98.93	99.00	98.90	0.9785	**113**
Vgg-f	**99.20**	**99.20**	**99.20**	**0.9838**	973
AlexNet + CaffeNet + Vgg-f	99.20	99.20	99.20	0.9838	3094

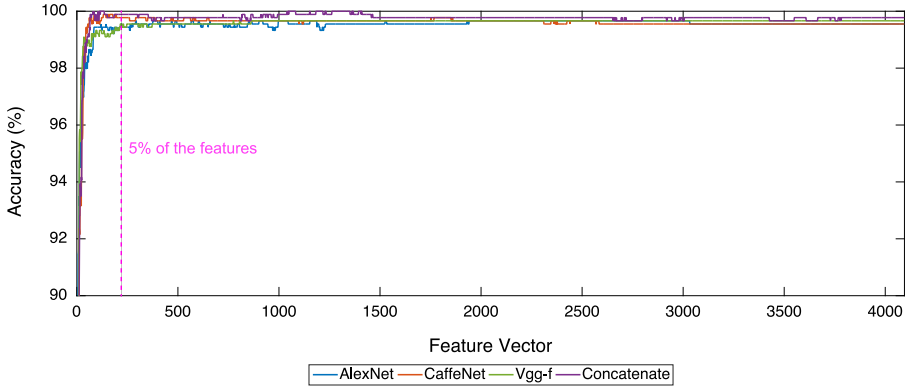


Fig. 6. Accuracy obtained by the proposed methodology, varying the size of the feature vector of each CNN for the Hybrid-Leukocyte database.

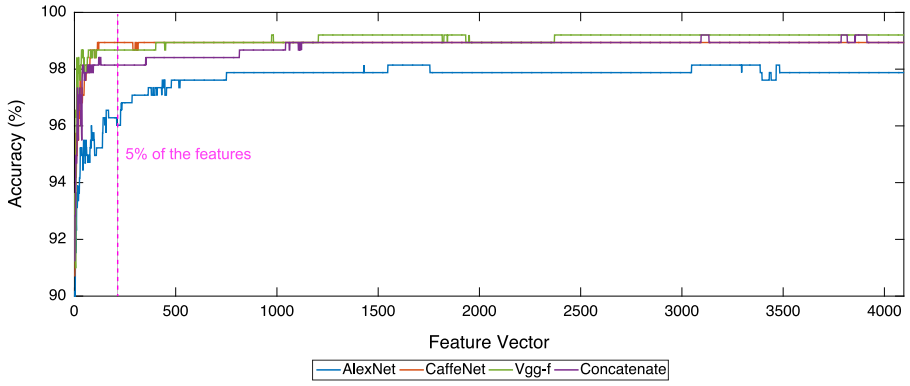


Fig. 7. Accuracy obtained by the proposed methodology, varying the size of the feature vector of each CNN for the Hybrid-Slide database.

Table 8
Results of accuracy for each classifier in the Hybrid-Slide database.

Classifiers	*A* (%) (973 features)
MLP	98.93
RF	98.40
KNN(3)	99.20
SVM	**99.20**

same accuracy rate, which demonstrates the excellent performance with different classifiers.

Table 9 shows the comparison of the accuracy of the proposed methodology with other state-of-the-art methods for the Hybrid-Slide database. It is worth mentioning that the works presented in Table 6 were developed for classifying images with only one leukocyte. Therefore, they do not work for the Hybrid-Slide database.

Analyzing Table 9, we can see that the proposed methodology performed better than other works presented in the literature. We believe that the results can still be improved by a fine-tuning operation (Tajbakhsz et al., 2016), since the learning transfer method used cannot wholly abstract the characteristics of the blood images since the networks were trained to extract generic and non-specific features of the disease. However, to carry out this operation of fine-tuning, a more substantial number of images are required.

Table 9
Comparison of the accuracy obtained by the proposed methodology and other works presented in the literature for the Hybrid-Slide database.

Work	Number of images	*A* (%)
Madhukar et al. (2012)	98	93.50
Vincent et al. (2015)	100	97.70
Patel and Mishra (2015)	27	93.75
Agaian et al. (2016)	98	>94.00
Singh et al. (2017)	108	97.20
Proposed methodology	377	**99.20**

5.3. Hybrid-complete database

Fig. 8 shows the hit rates of the approach with the selected vectors of different sizes for the Hybrid-Complete database. Analyzing this figure, we observed that the separate architectures achieved better results. While the accuracy of its concatenation grows with the number of attributes, but it does not present a superior result.

Table 10 presents the best results for each architecture and the size of the feature vector. The accuracy rates obtained by each CNN were almost the same. However, the Vgg-f feature vector was much smaller in size than the feature vectors of the other CNNs. Thus, this was the

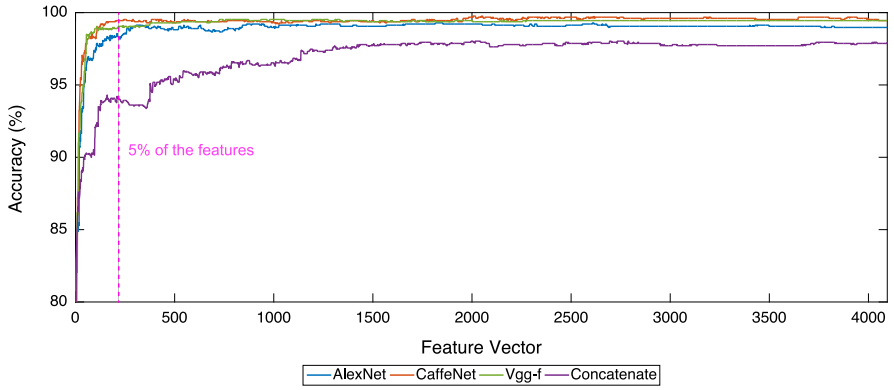


Fig. 8. Accuracy obtained by the proposed methodology, varying the size of the feature vector of each CNN for the Hybrid-Complete database.

chosen CNN of our methodology. The extracted attribute concatenation obtained the lowest accuracy among the approaches.

By ranking the attributes with the gain ratios in decreasing order, we observed that the difference in the gain of each attribute was small. Thus, more characteristics are required to achieve high accuracy rates. The results obtained in this database emphasize the use of CNNs as descriptors of leukemic images and also the robustness of the method proposed in this work. Table 11 presents the results for each classifier and the SVM presents the best accuracy. The KNN classifier with three clusters achieves good results; however it is not enough to validate the proposed approach.

Table 10
Results of accuracy, precision, recall, and Kappa index and the size vector chosen for each CNN for the Hybrid-Complete database.

	A (%)	Vector size
AlexNet	99.29	1813
CaffeNet	**99.76**	2003
Vgg-f	99.52	**724**
AlexNet + CaffeNet + Vgg-f	98.02	2015

Table 11
Results of accuracy for each classifier in the Hybrid-Complete database.

Classifiers	*A* (%) (2003 features)
MLP	99.13
RF	98.02
KNN(3)	99.21
SVM	**99.76**

6. Conclusion

The work presented in this paper describes a new methodology for the diagnosis of leukemia in blood images using Convolutional Neural Networks (CNNs). Based on the results obtained by the proposed approach, it was possible to validate the robustness of pre-trained CNNs for extracting features in relation to classical state-of-the-art methods. Through the selection of attributes, we observed that more characteristics are required to classify the images with many leukocytes, while fewer features are required for images with only one leukocyte.

Besides, the main advantage of the proposed methodology that allows it to perform better than other state-of-the-art methods is that it does not need a segmentation process. Furthermore, we proved the robustness of our methodology by using three heterogeneous databases.

As future work, the technique of fine-tuning in the CNN to improve the abstraction of leukemia information is proposed. Besides improving the CNNs, new databases of images with a larger amount of data will be used, allowing this system to be validated and used in daily life, helping physicians and patients to diagnose this disease.

Acknowledgments

The authors would like to thank the Brazilian National Counsel of Technological and Scientific Development (CNPq) (136240/2016-0) and the Federal University of Piauí (UFPI) for sponsoring our research.

References

Agaian, S., Madhukar, M., Chronopoulos, A.T., 2018. A new acute leukaemia-automated classification system. Comput. Meth. Biomech. Biomed. Eng.: Imaging Vis. 6 (3), 303–314.

Aha, D., Kibler, D., 1991. Instance-based learning algorithms. Mach. Learn. 6, 37–66.

Athiwaratkun, B., Kang, K., 2015. Feature representation in convolutional neural networks, CoRR abs/1507.02313. arXiv:1507.02313.

Breiman, L., 2001. Random forests. Mach. Learn. 45 (1), 5–32.

Castelluccio, M., Poggi, G., Sansone, C., Verdoliva, L., 2015. Land use classification in remote sensing images by convolutional neural networks, pp. 1–11, CoRR abs/1508.00092.

Chatfield, K., Simonyan, K., Vedaldi, A., Zisserman, A., 2014. Return of the devil in the details: delving deep into convolutional nets. In: British Machine Vision Conference, arXiv:1405.3531.

chen He, D., Wang, L., 1990. Texture unit, texture spectrum, and texture analysis. IEEE Trans. Geosci. Remote Sens. 28 (4), 509–512.

Cortes, C., Vapnik, V., 1995. Support-vector networks. Mach. Learn. 20 (3), 273–297.

Friedman, N., Geiger, D., Goldszmidt, M., 1997. Bayesian network classifiers. Mach. Learn. 29 (2–3), 131–163.

Guyon, I., Elisseeff, A., 2006. An introduction to feature extraction. In: Guyon, I., Gunn, S., Nikravesh, M., Zadeh, L. (Eds.), Studies in Fuzziness and Soft Computing. Springer, Berlin, Heidelberg, pp. 1–25.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H., 2009. The weka data mining software: An update. SIGKDD Explor. Newslett. 11 (1), 10–18.

Haralick, R.M., Shanmugam, K., Dinstein, I., 1973. Texture features for image classification. IEEE Trans. Syst. Man Cybern. 3 (6), 610–621.

Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., Darrell, T., 2014. Caffe: Convolutional architecture for fast feature embedding. In: Proceedings of the 22Nd ACM International Conference on Multimedia. MM '14, ACM, New York, NY, USA, pp. 675–678.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. In: Pereira, F., Burges, C.J.C., Bottou, L., Weinberger, K.Q. (Eds.), Advances in Neural Information Processing Systems 25. Curran Associates, Inc., pp. 1097–1105.

Kumar, A., Kim, J., Lyndon, D., Fulham, M., Feng, D., 2016. An ensemble of fine-tuned convolutional neural networks for medical image classification. IEEE J. Biomed. Health Inform. 21 (1), 1–9.

Labati, R.D., Piuri, V., Scotti, F., 2011. All-idb: The acute lymphoblastic leukemia image database for image processing. In: 18th IEEE International Conference on Image Processing, ICIP, pp. 2045–2048.

Lecun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521 (7553), 436–444.

Litjens, G.J.S., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., van der Laak, J.A.W.M., van Ginneken, B., Sánchez, C.I., 2017. A survey on deep learning in medical image analysis, CoRR abs/1702.05747.

Madhukar, M., Agaian, S., Chronopoulos, A.T., 2012. New decision support tool for acute lymphoblastic leukemia classification. In: Proc. SPIE 8295, Image Processing: Algorithms and Systems X; and Parallel Processing for Imaging Applications II, pp. 8295 –8295 –12.

Mohapatra, S., Patra, D., Satpathy, S., 2014. An ensemble classifier system for early diagnosis of acute lymphoblastic leukemia in blood microscopic images. Neural Comput. Appl. 24, 1887–1904.

Mohaprata, S., Samanta, S.S., Patra, D., Satpathi, S., 2011. Fuzzy based blood image segmentation for automated leukemia detection. In: 2011 International Conference on Devices and Communications, pp. 1–5.

Neoh, S.C., Srisukkham, W., Zhang, L., Todryk, S., Greystoke, B., Lim, C.P., Hossain, M.A., Aslam, N., 2015. An intelligent decision support system for leukaemia diagnosis using microscopic blood images. Sci. Rep. 5 (14938), 1–14.

Omid Sarrafzadeh, A.T.H.U.B., Rabbani, Hossein, 2014. Selection of the best features for leukocytes classification in blood smear microscopic images. Proc. SPIE 9041, 9041–9041–8.

Patel, N., Mishra, A., 2015. Automated leukaemia detection using microscopic images. Procedia Comput. Sci. 58, 635–642.

Pentland, A.P., 1984. Fractal-based description of natural scenes. IEEE Trans. Pattern Anal. Mach. Intell. PAMI-6 (6), 661–674.

Popescu, M.-C., Balas, V.E., Perescu-Popescu, L., Mastorakis, N., 2009. Multilayer perceptron and neural networks. WSEAS Trans. Circuits Syst. 8 (7), 579–588.

Powers, D.M.W., 2011. Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. Int. J. Mach. Learn. Technol. 2 (1), 37–63.

Putzu, L., Caocci, G., Ruberto, C.D., 2014. Leucocyte classification for leukaemia detection using image processing techniques. Artif. Intell. Med. 62, 179–191.

Rawat, J., Singh, A., Bhadauria, H.S., Virmani, J., 2015. Computer aided diagnostic system for detection of leukemia using microscopic images. In: 4th International Conference on Eco-Friendly Computing and Communication Systems, pp. 748–756.

Rollins-Raval, M., Raval, J., Contis, L., 2012. Experience with cellavision dm96 for peripheral blood differentials in a large multi-center academic hospital system. J. Pathol. Inform. 3 (29), 1–9.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A.C., Fei-Fei, L., 2015. Imagenet large scale visual recognition challenge. Int. J. Comput. Vis. (IJCV) 115 (3), 211–252.

Sarrafzadeh, O., Dehnavi, A.M., 2015. Nucleus and cytoplasm segmentation in microscopic images using K means clustering and region growing. Adv. Biomed. Res. 79–87.

Sarrafzadeh, O., Rabbani, H., Dehnavi, A.M., Talebi, A., 2015. Detecting different subtypes of acute myelogenous leukemia using dictionary learning and sparse representation. In: ICIP. IEEE, pp. 3339–3343.

Schwenker, F., Kestler, H.A., Palm, G., 2001. Three learning phases for radial-basis-function networks. Neural Netw. 14 (4–5), 439–458.

Shin, H., Roth, H.R., Gao, M., Lu, L., Xu, Z., Nogues, I., Yao, J., Mollura, D.J., Summers, R.M., 2016. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. IEEE Trans. Med. Imaging 35 (5), 1285–1298.

Singh, G., Bathla, G., Kaur, S., 2017. Design of new architecture to detect leukemia cancer from medical images. Int. J. Appl. Eng. Res. 11 (10), 7087–7094.

Singhal, V., Singh, P., 2016. Texture features for the detection of acute lymphoblastic leukemia. In: Satapathy, S.C., Joshi, A., Modi, N., Pathak, N. (Eds.), Proceedings of International Conference on ICT for Sustainable. Springer Singapore, Singapore, pp. 535–543.

Tajbakhsz, N., Shin, J.Y., Gurudu, S.R., Hurst, R.T., Kendall, C.B., Gotway, M.B., Liang, J., 2016. Convolutional neural networks for medical image analysis: Full training or fine tuning? IEEE Trans. Med. Imaging 35, 1299–1312.

Vincent, I., Kwon, K.-R., Lee, S.-H., Moon, K.-S., 2015. Acute lymphoid leukemia classification using two-step neural network classifier. In: Frontiers of Computer Vision. (FCV), pp. 1–4.

Wang, D., Khosla, A., Gargeya, R., Irshad, H., Beck, A.H., 2016. Deep learning for identifying metastatic breast cancer, ArXiv E-Prints.