Diagnosing Leukemia in Blood Smear Images Using an Ensemble of Classifiers and Pre-trained Convolutional Neural Networks

Luis H. S. Vogado, Rodrigo de M. S. Veras, Alan R. Andrade, Flavio H. D. de Araujo,
Romuere R. V. e Silva, Kelson R. T. Aires
Department of Computer Science, Federal University of Piauí - Teresina, Brazil
lhvogado@gmail.com, rveras@ufpi.edu.br, alanribeiroandrade@gmail.com, flavio86@ufpi.edu.br,
romuere@ufpi.edu.br, krtaires@gmail.com

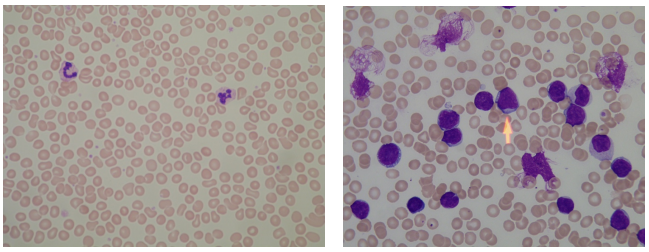
***Abstract*—Leukemia is a worldwide disease. In this paper we demonstrate that it is possible to build an automated, efficient and rapid leukemia diagnosis system. We demonstrate that it is possible to improve the precision of current techniques from the literature using the description power of well-known Convolutional Neural Networks (CNNs). We extract features from a blood smear image using pre-trained CNNs in order to obtain an unique image description. Many feature selection techniques were evaluated and we chose PCA to select the features that are in the final descriptor. To classify the images on healthy and pathological we created an ensemble of classifiers with three individual classification algorithms (Support Vector Machine, Multilayer Perceptron and Random Forest). In the tests we obtained an accuracy rate of 100%. Besides the high accuracy rate, the tests showed that our approach requires less processing time than the methods analyzed in this paper, considering the fact that our approach does not use segmentation to obtain specific cell regions from the blood smear image.**

I. Introduction

Diagnosis is an important process performed by physicians, which consists in determining the presence or absence of diseases based on a dataset. These data are essential for the identification of diseases and can be composed of signs, symptoms, images, exams, among others. An erroneous diagnosis, caused by an unsuccessful examination, can cause side effects to the patient, due to a possible prescription of medicines that are not appropriate for the treatment of a specific disease. To assist specialists at this crucial stage, there are low-cost computational systems that analyze and process the data, providing diagnostic assistance.

Over the years, multiple medical aid systems have been proposed. Diseases such as glaucoma [1], skin cancer [2], breast cancer [3] and leukemia [4] have been addressed in these systems. Early diagnosis of these diseases is critical to the success of their treatments, which are costly and complex. Even though in some cases the treatment does not cure the patient, it prolongs their survival.

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 5 years of age. Leukemia is a cancer that originates in the bone marrow (Figure 1a) and is characterized by the abnormal proliferation of white blood cells (Figure 1b). The diagnosis of leukemia can be done through various tests and exams, including physical examination, blood test, blood count, myelogram, lumbar puncture and bone marrow biopsy. The use of microscopic analysis is the most cost-effective approach for the initial screening of patients with leukemia. This type of test is done manually, which may generate fatigue in operators. Therefore, there is need for a low-cost system that is automatic and robust to avoid the influence of the operator.



(a) (b)

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

Many computer-aided diagnosis systems are developed with the use of image processing and computational intelligence techniques. These systems usually have steps such as: pre-processing, segmentation, feature extraction and classification. The steps that best define the diagnosis performed by these systems are the feature extraction and classification. However, to achieve more robustness, these steps end up depending on their predecessors, which means that a good segmentation can provide a good feature extraction and consequently a good classification.

In recent years, researchers working on artificial intelligence have been using deep learning through Convolutional Neural Networks (CNNs), enabling the creation of powerful computer systems for medical assistance. The power of these networks is applied in various areas, including the processing of signals, natural language, image and video. However, it requires machines with large processing capacity and a huge amount of

data to train and adapt these networks to perform the desired task.

The proposed approach in this work describes a leukemia diagnosis system that does not require the segmentation process (commonly present in state-of-the-art works). In this work we evaluated the use of multiple pre-trained CNNs, AlexNet [5], CaffeNet [6] and Vgg-f [7], as well as a combination of the three CNNs using a concatenation of their feature vectors. For the extraction step, we selected Vgg-f [7], considering the results obtained in our tests. Due to the high number of features obtained in the feature extraction step, we needed to use a technique for the feature selection. Many feature selection techniques were evaluated and we chose Principal Component Analysis (PCA) [8] to select the features that are in the final descriptor. In the ensemble's construction, several classifiers were empirically tested, and three were chosen considering the classifiers with the best results among all the classifiers that were tested. We propose an ensemble with three classifiers, the Support Vector Machine (SVM) [9], Multilayer Perceptron (MLP) [10] and Random Forest (RF) [11]. We sought to test the proposed system using a blood smear image database containing multiple nuclei per image. The results obtained by our system are compared to other systems present in the state-of-the-art.

The remainder of the text is divided as follows: related works are presented in Section II. The proposed system is presented in Section III. We describe the image databases used in the tests and present results and discussion in IV. Finally, we have conclusions and prospects for future works in Section V.

II. Related Work

Among the leukemia diagnosis systems developed over the years, some works present solutions using blood images for the classification of the two most common types of leukemia: Acute lymphoblastic leukemia (ALL) and acute myeloid leukemia (AML).

In the work proposed by Madhukar et al. [12], a system was developed to detect ALL using images from a single database, these images have multiple nuclei per image. The pre-processing step consists in the conversion of the image from RGB to the L*a*b color space. In the segmentation step, the unsupervised algorithm K-means is applied to components *a and *b from the converted image, with the number of groups equals to three. In the feature extraction step, they used shape features (e.g. area, perimeter, compactness, solidity, eccentricity, elongation and form-factor), Gray Level Co-occurrence Matrix (GLCM) [13] and Fractal dimension [14] as descriptors. In order to evaluate the system, they used 98 blood images from the ALL-IDB1 database [15]. The classification was performed using Support Vector Machine and three techniques for cross-validation: k-fold, Hould-Out and Leave-One-Out. After analyzing the results, the authors concluded that the technique that obtained the best accuracy was Leave-One-Out with 93.50%.

Vincent et al. [16] proposed the use of neural networks as classifiers. The method proposed in this work starts converting the image from RGB to the L*a*b color space. The resulted image is used in the clustering algorithm k-means, which separates the image into three different classes based on their color information. Constrast enhacement, auto-thresholding and morphological operations are applied in order to obtain the nucleus segmented image. The feature extraction and classification stages are subdivided into two steps. The first feature vector set obtained consists of five textural features, four Gray Level Co-occurrence Matrix (GLCM) features (e.g. energy, entropy, contrast, and correlation) and one fractal feature which is represented by Hausdorff Dimension. These feature vectors are analyzed by PCA algorithm which produces the input source for the first neural network classifier, the purpose of this classifier is to classify the cells in normal and abnormal. The same algorithm is applied to the second feature extraction process, though the extracted features are different. Since the second classifier needs a better differentiation, five geometrical features (e.g. cell area, nucleus area, cytoplasm area, nucleus-to-cytoplasm area ratio, and nucleus-to-cell area ratio) are extracted and analyzed by PCA algorithm in order to produce input for second neural network classifier that identifies AML and ALL. Both neural networks are trained using Levenberg-Marquardt (LM) algorithm [17]. The system achieved an accuracy of 97.70% using 100 blood images from ALL-IDB1 base.

In Patel and Mishra [18], the authors presented an automatic system for the detection of leukemia using microscopic blood images. This work can be divided into preprocessing, segmentation, feature extraction and classification stages. In the first stage, filters were used to remove possible noises in the image, to facilitate the segmentation of the image. The authors, unlike other state-of-the-art works, do not make changes in the color space, using the original RGB color space. In the segmentation step, the image is converted to grayscale and the clustering algorithms K-means and Zack [19] are applied. In the feature extraction stage, color, geometry, texture and statistics features were used. ALL-IDB1 is used to evaluate this system, however, only 27 images were used in the tests. The system achieved an accuracy of 93.57% using SVM.

The system proposed by Agaian et al. [20] presented an approach for the classification of blood images with multiple nuclei. The authors converted the images from RGB to the L*a*b color space and applied the clustering algorithm K-means. The chosen features were: shape, color, GLCM, Haar wavelet and Fractal dimension. SVM was used as the classifier. The system obtained an accuracy above 94.00% using 98 images from ALL-IDB1.

After analyzing the works presented in this section, it is possible to observe that many state-of-the-art methods use similar descriptors to classify blood images. However, in recent researches, authors have been applying deep leaning techniques as descriptors [21] and classifiers [3], [22], which can be used in the development of powerful computational systems. We noted that these techniques are being used in

computer-aided diagnosis systems, but there are no works involving leukemia. The number of images available for our tests (108) is not high enough to retrain a CNN in an efficient manner. Therefore, we decided to use the CNN for feature extraction.

III. Proposed System

The method proposed in this work aims to diagnose leukemia using blood smear images. Following the flowchart shown in Figure 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 methods present in the state-of-the-art. From the input image, the CNN is used to describe it, and the feature vector obtained is then reduced using the PCA. For the classification stage, a ensemble of classifiers is proposed in this work, bringing more reliability to the results and thus classifying the images in healthy or not. These steps are best described throughout the text, justifying the use of each component in the overall flowchart of the approach.

A. Convolutional Neural Network

The feature extraction process consists of the representation of a redundant set of data through unique features that make it different from other sets. There are numerous features that can be extracted from an image, for example: color, shape and texture. Among the several descriptors proposed in the state-of-the-art, the use of CNNs in this extraction process has been gaining prominence.

A CNN is a network formed by several layers that can be used in object recognition and image classification. Among these layers, we have the convolutional layers that can alter the representation of the data through filters. Usually after a convolutional layer, an activation function is used, these functions perform non-linear transformations in the data, in order to generate linearly separable outputs. One of the most common functions used with this purpose is the Rectified Linear units (ReLu), presented in Equation 1, where x is the input to a neuron.

$$f(x) = max(0, x) \tag{1}$$

The pooling layer succeeds the previous layers and reduces the amount of features of the resulting data. The fully connected layers are responsible for gathering all the features of the descriptors so that they can be classified by the final layer. Figure 3 presents the organization of the aforementioned layers.

In recent works, authors have presented two different ways of using the power of a CNN. The first is the usual way, by performing the training with a large set of data. The second way is the transfer of learning using pre-trained networks.

In this study, we used the learning transfer technique [23] where the CNN is trained using a large natural base of images. This training allows the CNN to assimilate generic features, which facilitates its applicability 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 depend on the similarity of the images from the base used to extract the features and the images from the training set.

Two options of learning transfer are presented in the work of Castellucio et al. [21]. The first consists in fine-tuning the network, where the structure is modified, freezing high-level layers. The second is the extraction of the last fully connected layer of the network, obtained from the input image [24]. Then, it uses another classifier in the classification process. In this work, we chose the second option, presented in the flow chart in Figure 2.

The CNNs we used were pre-trained using the natural image database ImageNet. As mentioned before, we used Vgg-f [7] in the feature extraction step of our system.

a) Vgg-f:: This model was proposed in 2014 by Chatfield et al. [7], and it 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 convulational filters. The size of the filters influence the computing power demanded by the network, the smaller the filter, the less the network will require of the computer. Larger filters take advantage of more neighborhood information. The main differences between this architecture and AlexNet is in the smaller amount of convolutional filters and the dense connectivity between the convolution layers present in Vgg-f, containing eight learning layers, five convolutional layers and three fully connected layers.

We also tested other two CNNs, AlexNet [5] and CaffeNet [6], for comparison. These architectures are described below.

b) AlexNet: This architecture was developed by Krizhevsky et al. [5] for the ILSVRC-2010 competition in order to carry out training and classification of ImageNet database. It comprises eight layers that need to be trained, five convolutional layers with filters of size 5x5 and 7x7, followed by three fully-connected layers, as well as max-pooling layers.

c) 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 [6]. It comprises five convolutional layers, each followed by a pooling layer, and three fully-connected layers. The main difference between this architecture and AlexNet [5] is in the order of the pooling layer and the normalization layer. CaffeNet is also easier to manipulate and change properties in comparison to the architectures analyzed in this paper. Besides that, its API is available in C++, Matlab and Python.

The number of features taken from the last fully connected layer of each architecture is 4096. In the concatenated network, the resulting vector of the extraction is the fusion of the attributes from each model, resulting in a total of 12288 characteristics.

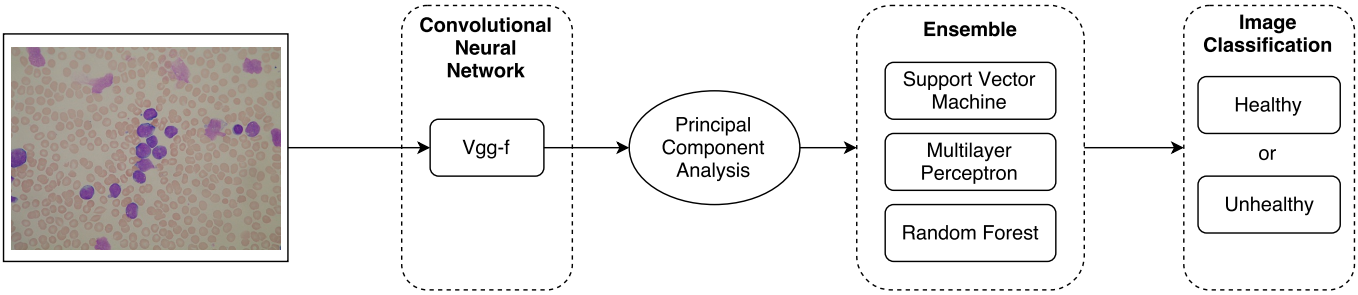


Figure 2. Flowchart of the proposed approach.

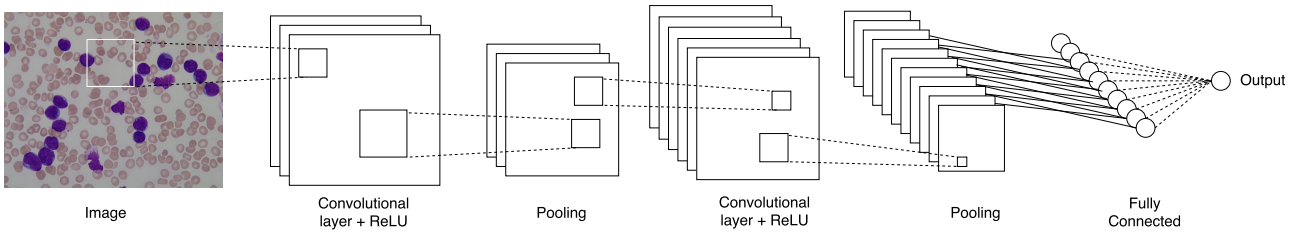


Figure 3. A simplified illustration of the CNN architecture.

B. Feature Selection

After generating the feature vector, we performed a selection of attributes. This selection aims to eliminate unnecessary attributes, and consequently simplify the prediction model, reduce the computational cost, as well as provide a better understanding of the results found. According to [25], attribute selection techniques are primarily employed to identify relevant attributes and essential information.

An optimal selection of attributes for classification problems requires an exhaustive search of all possible subsets of attributes [26], making it impractical when the number of attributes is too high. For this reason, several researchers have developed different attribute selection techniques. Each of these techniques uses distinct selection criteria and search algorithms to evaluate, and to find heuristically the most appropriate subset of attributes.

After performing many tests, we selected attributes using the PCA algorithm [8]. The application of PCA is made in several problems, among them, we can mention facial recognition, passport verification, medical records, etc.

C. Ensemble of Classifiers

The use of ensembles in machine learning has been increased in the last years. Composed by several classifiers, this classification technique is considered more efficient than the use of only one classifier, since it provides a greater reliability in the results. Works from the literature([22], [27]) disseminate the use of this technique in computer-aided systems.

The choice for multiple classifiers, with different properties and their combination, corroborates with the ensemble's efficiency. In our work, we selected three classifiers based on tests and works from the literature, they are: Support Vector Machine (SVM), Multilayer Perceptron (MLP) and Random Forest (RF).

There are many ways of combining classifiers, usually the individual output of each classifier is used in this process, which minimizes the occurrence of erroneous decisions. Each image is classified as ill or not by each classifier, the results obtained by these classifiers are then combined in order to obtain the final result. One of the most used means in this combination is the majority vote. This aims to evaluate the output of each classifier and if a majority agrees in a certain class, then this will be its final classification. In addition to the majority vote, there are other rules such as the weighted majority vote, edge count, median, average and probability product. [28].

In our work, we used the rule of majority voting to combine the outcomes obtained by each classifier. In order to perform the training and testing of the data set, we decided to use the k-fold cross validation, with the value of k being 5. To evaluate the results obtained, we used a image database with multiple nuclei per image. The implementation of the feature extraction step of our method was made in MATLAB, while the features reduction and the classification of images were made with the WEKA tool [29]

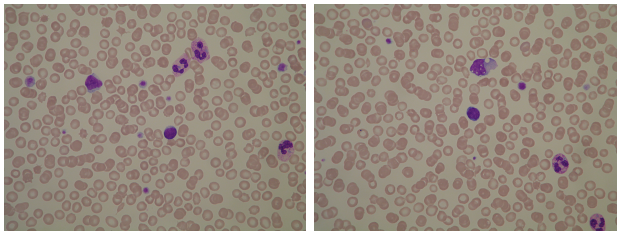
IV. Experiments

A. Image Database

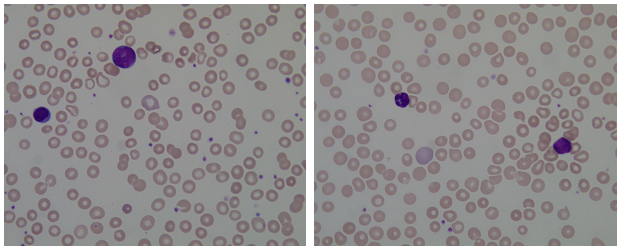
Proposed in [15], the ALL-IDB is a database which is divided into two distinct versions, called ALL-IDB1 and ALL-IDB2. All images in the datasets have a native resolution equal to 2592 $\times$ 1994, captured with a PowerShot G5 camera.

ALL-IDB1 has 108 images (59 healthy images and 49 images with cancer) of blood smears, containing multiple nuclei per image. The lymphoblasts were labeled by specialists,

resulting in 510 elements. ALL-IDB2 has 260 images, each of them contains one lymphoblast or lymphocyte per image. In this work we used ALL-IDB1, considering the fact that this database only presents images containing multiple nuclei and the focus of this work is to classify images with that property. In Figure 4 we can see examples of healthy (Figure 4a) and unhealthy (Figure 4b) images of the ALL-IDB1 base.



(a) Images without leukemia.



(b) Images with leukemia.

Figure 4. ALL-IDB1 images sample.

B. Experimental Results and Discussion

In our experiments, we used the previously presented architectures, AlexNet, CaffeNet, Vgg-f and a concatenation of the three to perform the feature extraction step. For the verification of the results, we chose k-fold cross-validation, with the value of k being 5. This value was chosen because it presented a larger set of images to be tested, which correspond to 20% of the total for tests and 80% for training.

In Table I, we present empirical tests performed with the purpose of defining the rate of variation of the data applied to the PCA, we used intervals of 0.1 between each variation. These experiments seek the best compromise between accuracy and attribute vector size. We observed that the architectures achieved excellent results, achieving better results than those presented by state-of-the-art methods. However, we aimed to achieve the best possible results with the least amount of attributes. This occurred with the variation of 0.4 in the vector extracted from Vgg-f. This being the smallest vector that presents 100% hit using the database, with only 6 attributes.

In the same table, we can observe the results of Vgg-f compared to the outputs of the other two architectures, as well as the combination the the three of them. The three architecture models used in this work reached relatively close results. This indicates that shallower networks grasp general features that are applicable to a wider variety of images. The features extracted by these architectures are less semantically optimized for natural images. However they are more generalizable and adaptable when transferred to the medical imaging domain. We carried out an analysis on the relevance of the features from the concatenated vector, we observed that the features coming from Vgg-f are the ones that present the greatest gain ratio.

It is also possible to observe that the concatenation of the networks also obtained excellent hit rates. This is due to the fact that the output of each network has a set of data that describes it in a unique way. The reason behind the concatenation of these features was to obtain a set of features capable of representing the data present in the images in different ways. However, by concatenating these features, the size of the features vector is tripled. Thus, the reduction of the features vector was carried out, which consequently discarded the less relevant features. After analysis of the results, we noted that Vgg-f presented results that are equivalent to the concatenation of networks, which highlights its effectiveness, considering the fact that it does not need to perform a concatenation of other architectures to obtain satisfactory results, and for that reason, it has a smaller computational time in comparison to the combination of architectures.

Much of the work in the literature presents systems of classification of leukemia with only one nucleus per image, whereas the classification with several nuclei does not have as many approaches as the former. The classification of images with multiple nuclei is relevant to the diagnosis of the disease, because it presents the same pattern of images used by physicians in hospitals. A segmentation step and cutting of images containing multiple nuclei into single nucleus images is then unnecessary to perform the desired task. In Table II we compare the results obtained by the proposed system with other systems that seek to perform the diagnosis of leukemia in blood smear images containing multiple nuclei per image.

In the work proposed by Agaian et al. [20], the authors did not present the exact accuracy of their system, only emphasized that it was greater than 94%. In another work published by the same authors ([12]), the result was 93.50%, using the same image database. The work presented by Patel and Mishra in [18] reached an accuracy of 93.75%, which is higher than the result obtained by Madhukar et al. in [12]. However, the smaller amount of images used in [18] makes the system not so reliable when compared to other works that were tested with a larger set of images. The system proposed in this work obtained the highest accuracy among all systems described in this paper, with an accuracy of 100%. Besides that, our system was tested using a higher number of images than the others, providing a greater reliability in the results and thus validating the approach using the ALL-IDB1 base.

A factor that can contribute to the instability of state-of-the-art methods is the use of a segmentation step that depends on the specific characteristics of a base. As presented in [30], the authors compared their segmentation method with methods from the literature using different bases. As a result, it was noted that many state-of-the-art methods presented problems with bases that have distinct characteristics. Therefore, there is

Table I
ANALYSIS OF THE COMPROMISE BETWEEN THE ACCURACY AND THE FEATURE VECTOR SIZE.

	Accuracy(%) | Number of features																	
CNN | PCA	0.1		0.2		0.3		0.4		0.5		0.6		0.7		0.8		0.9	
AlexNet	93.51	1	92.59	2	98.14	4	99.07	7	99.07	11	98.14	19	98.14	28	98.14	42	97.22	64
CaffeNet	95.37	1	99.07	3	99.07	4	100	8	99.07	13	100	20	98.14	31	97.22	44	93.59	65
Vgg-f	95.37	1	93.51	2	99.07	3	**100**	**6**	99.07	8	100	14	100	22	100	36	100	58
AlexNet + CaffeNet + Vgg-f	92.59	1	94.44	2	99.07	4	100	7	100	12	100	20	100	31	98.14	46	95.37	68

Table II
ACCURACY OBTAINED BY THE METHODS IN THE CLASSIFICATION STAGE USING IMAGES FROM ALL-IDB1.

Methods	**Number of images**	**A**(%)
Madhukar et al. [12]	98	93.50
Vincent et al. [16]	100	97.70
Patel and Mishra [18]	27	93.75
Agaian et al. [20]	98	>94.00
Proposed	**108**	**100**

no guarantee that a method will present a good segmentation using images from multiple bases. With that information, we can say that the absence of prior segmentation is a point in favor of our approach. Because problems with possible segmentation errors do not influence the final result of our method.

V. CONCLUSION

The feature extraction and classification stages are considered the most important steps in computer-aided diagnosis systems. We observed that the methods from the literature presented promising results. However, these methods need to be able to perform the task on larger image sets to confirm the reliability of their results. The work presented in this paper describes a new system for the diagnosis of leukemia in blood images using a Convolutional Neural Network (Vgg-f), PCA and an ensemble with three classifiers. Based on the results obtained by the methodology, it is possible to validate its robustness in comparison to other works with an accuracy of 100%.

For future work, we propose the use of an fine tuning in the architecture, in order to improve the information abstraction of leukemia. The current approach does not require a great amount of processing, but it still does not represent the maximum power of CNNs. In addition to the improvements in the system, we intend to use new image databases, with a greater amount of data, so the system can be validated and used in daily life, helping physicians and patients in the diagnosis of this disease.

ACKNOWLEDGMENT

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

REFERENCES

[1] X. Chen, Y. Xu, D. W. K. Wong, T. Y. Wong, and J. Liu, "Glaucoma detection based on deep convolutional neural network," in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, vol. 37, August 2015, pp. 715–718.

[2] J. Kawahara and G. Hamarneh, *Multi-resolution-Tract CNN with Hybrid Pretrained and Skin-Lesion Trained Layers*. Springer International Publishing, 2016.

[3] D. Wang, A. Khosla, R. Gargeya, H. Irshad, and A. H. Beck, "Deep Learning for Identifying Metastatic Breast Cancer," *ArXiv e-prints*, june 2016.

[4] M. Madhukar, S. Agaian, and A. T. Chronopoulos, "Automated screening system for acute myelogenous leukemia detection in blood microscopic images," *IEEE Systems Journal*, vol. 8, no. 3, pp. 995–1004, 2014.

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

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

[7] K. Chatfield, K. Simonyan, A. Vedaldi, and A. Zisserman, "Return of the devil in the details: Delving deep into convolutional nets," in *British Machine Vision Conference*, 2014.

[8] K. Pearson, "On lines and planes of closest fit to systems of points in space," *Philosophical Magazine*, vol. 2, pp. 559–572, 1901.

[9] C. Cortes and V. Vapnik, "Support-vector networks," in *Machine Learning*, 1995, pp. 273–297.

[10] M.-C. Popescu, V. E. Balas, L. Perescu-Popescu, and N. Mastorakis, "Multilayer perceptron and neural networks," *WSEAS Trans. Cir. and Sys.*, vol. 8, no. 7, pp. 579–588, Jul. 2009.

[11] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[12] M. Madhukar, S. Agaian, and A. T. Chronopoulos, "New decision support tool for acute lymphoblastic leukemia classification," pp. 829 518–829 518–12, 2012.

[13] R. Haralick, K. Shanmugam, and I. Dinstein, "Texture features for image classification," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 3, no. 6, 1973.

[14] Poggio, Koch, K. Arakawa, and E. Krotkov, "A. pentland. fractal-based description of natural scenes." Tech. Rep., 1992.

[15] R. D. Labati, V. Piuri, and F. Scotti, "All-idb: The acute lymphoblastic leukemia image database for image processing." *Image Processing (ICIP), 2011 18th IEEE International Conference on*, pp. 2045–2048, 2011.

[16] I. Vincent, K.-R. Kwon, S.-H. Lee, and K.-S. Moon, "Acute lymphoid leukemia classification using two-step neural network classifier," *Frontiers of Computer Vision (FCV)*, pp. 1–4, 2015.
[17] M. I. A. Lourakis, *A brief description of the Levenberg-Marquardt algorithm implemented by levmar*, Foundation for Research and Technology - Hellas, Vassilika Vouton, P.O. Box 1385, GR 711 10 Heraklion, Crete, GREECE, 2005.
[18] N. Patel and A. Mishra, "Automated leukaemia detection using microscopic images," vol. 58, pp. 635–642, 2015.
[19] G. W. Zack, W. E. Rogers, and S. A. Latt, "Automatic measurement of sister chromatid exchange frequency," *Journal of Histochemistry & Cytochemistry*, pp. 741–753, 1977.
[20] S. Agaian, M. Madhukar, and A. T. Chronopoulos, "A new acute leukaemia-automated classification system," *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, pp. 1–12, 2016.
[21] M. Castelluccio, G. Poggi, C. Sansone, and L. Verdoliva, "Land use classification in remote sensing images by convolutional neural networks," *CoRR*, vol. abs/1508.00092, 2015.
[22] A. Kumar, J. Kim, D. Lyndon, M. Fulham, and D. Feng, "An ensemble of fine-tuned convolutional neural networks for medical image classification," *IEEE Journal of Biomedical and Health Informatics*, vol. PP, pp. 1–9, 2016.
[23] H. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. J. Mollura, and R. M. Summers, "Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning," *CoRR*, vol. abs/1602.03409, 2016.
[24] B. Athiwaratkun and K. Kang, "Feature representation in convolutional neural networks," *CoRR*, vol. abs/1507.02313, 2015.
[25] I. Guyon and A. Elisseeff, "An introduction to feature extraction," in *Feature extraction*. Springer, 2006, pp. 1–25.
[26] J. Reunanen, "Overfitting in making comparisons between variable selection methods," *Journal of Machine Learning Research*, vol. 3, no. Mar, pp. 1371–1382, 2003.
[27] S. Mohapatra, D. Patra, and S. Satpathy, "An ensemble classifier system for early diagnosis of acute lymphoblastic leukemia in blood microscopic images," *Neural Computing and Applications*, vol. 24, pp. 1887–1904, 2014.
[28] M. P.Ponti, "Combining classifiers: From the creation of ensembles to the decision fusion." in *SIBGRAPI Tutorials*. IEEE Computer Society, 2011, pp. 1–10.
[29] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The weka data mining software: An update," *SIGKDD Explor. Newsl.*, vol. 11, no. 1, pp. 10–18, Nov. 2009.
[30] L. H. S. Vogado, R. de M S Veras, A. R. Andrade, R. R. V. e Silva, F. H. D. de Araujo, and F. N. S. de Medeiros, "Unsupervised leukemia cells segmentation based on multi-space color channels," in *International Symposium on Multimedia*. IEEE, 2016, pp. 451–456.