

# Symbolic Learning using Brain Programming for the Recognition of Leukemia Images

Rocio Ochoa-Montiel<sup>1,2</sup>, Humberto Sossa<sup>1,3</sup>, Gustavo Olague<sup>4</sup>, Mariana Chan-Ley<sup>4</sup>, José Menendez<sup>4</sup>

<sup>1</sup> Instituto Politécnico Nacional,  
Centro de Investigación en Computación,  
Mexico

<sup>2</sup> Universidad Autónoma de Tlaxcala,  
Facultad de Ciencias Básicas Ingeniería y Tecnología,  
Mexico

<sup>3</sup> Tecnológico de Monterrey,  
Mexico

<sup>4</sup> Centro de Investigación Científica y de Educación Superior de Ensenada,  
Laboratorio EvoVision,  
Mexico

{ma.rocio.ochoa, humbertosossa, gustavo.olague}@gmail.com,  
{mchan, jmenendez}@cicese.edu.mx

**Abstract.** In this work, we propose an approach of symbolic learning for the recognition of leukemia images. Image recognition for cancer detection is often a subjective problem due to different interpretations by experts of the medical area. Feature extraction is a critical step in image recognition, and current automatic approaches are unintelligible since they need to be adapted to different image domains. We propose the paradigm of brain programming as a symbolic learning approach to address aspects involved in the derivation of knowledge that allows us to recognize subtypes of leukemia in color images. Experimental results provide evidence that the multi-class recognition task is achieved through the solutions discovered from multiple runs of the bioinspired model.

**Keywords.** Leukemia recognition, symbolic learning, brain programming, evolutionary computer vision.

## 1 Introduction

Visual analysis of biomedical images is an essential task for the diagnosis of illnesses. Techniques of artificial vision allow identification,

recognition, and count in biological smears for diagnostic purposes, treatment, or classification of new pathologies [1, 8, 21, 10].

Artificial visual models in the medical area offer an excellent alternative for major problems that affect both the national and international community. One of these problems is opportune cancer detection, whose cost of diagnosis is significantly elevated. According to the World Health Organization (WHO), in 2018, at least 9.6 million people worldwide died from cancer, being nearly 1 in 6 of all global deaths. Leukemia is a type of cancer that is critical because it is a leading cause of death for children and adolescents worldwide.

Furthermore, in many low- and middle-income countries, only 20% are cured due to numerous factors like the inability to obtain an accurate diagnosis, the inaccessible therapy due to lack of access to essential medicines, and others [24]. There exists clear evidence of the efforts to understand leukemia using artificial vision

techniques. In this regard, until the year 2010, Kampen reports more than 226,267 publications about this topic, and this number continues to grow [7].

Visual properties of blood cells suggested from literature focused on hematological diseases and are studied in the biological image recognition by handcraft approaches [19]. In particular, traits as shape, color, and the distribution of some elements into the cell are meaningful for recognition.

Since feature extraction used in handcraft approaches is driven by human reasoning, these can provide practical issues for developing models with more robust and explainable learning techniques [13, 14]. In this work, we introduce brain programming as a symbolic learning method to address the problem of leukemia image recognition.

In the next Section, we recall the related works. Section 3 presents the theoretical concepts of Brain programming and the proposed methodology. In Section 4, Experiments and Results are shown. Conclusions are included in Section 6.

## 2 Background

Although leukemia recognition has been addressed for a long time, there are important drawbacks regarding computer diagnosis such as the specificity of the methods and their ad-hoc design focusing on specific datasets. The generation of meta-data is often carried out during experimentation or statistical analysis, which is time-consuming and expensive. On the other hand, automatic approaches categorize images through information extracted directly from the images. These approaches are competitive, generating opaque model predictions; also, there are limitations regarding image size and hardware resources to prove these models. In Table 1, we present some previous works on leukemia cell recognition.

In this work, we address the problem of leukemia image recognition using symbolic learning through a paradigm named brain programming. Symbolic learning explores the implications of artificial intelligence research through methods based

on high-level symbolic representations (human-readable) of problems, logic, and search [5]. In this regard, Brain programming (BP) is a paradigm of evolutionary computer vision that aims to emulate the behavior of the brain for vision problems according to neuroscience knowledge. In [15, 3], the authors introduce BP, tackling diverse problems of computer vision.

Genetic programming (GP) is the method used by brain programming to discover a set of evolutionary visual operators (EVOs) embedded within a hierarchical structure called the artificial visual cortex (AVC) [15]. These EVOs are functions for the description of the image classes.

## 3 Brain Programming for the Recognition of Leukemia Images

The leukemia image recognition problem is introduced from the standpoint of data modeling. Since, a minimization problem requires to find a solution  $\mathbf{L}_{min} \in S$  such that  $f(\mathbf{L}_{min})$  is a global minimum on  $S$ , then:

$$\exists \mathbf{L}_{min} \in S : f(\mathbf{L}_{min}) \leq f(\mathbf{L}). \quad (1)$$

In contrast to conventional methods, in which the aim is finding best-fit parameters; in GP and the recognition problem, the purpose is to find a function that satisfies the task of data modeling. Thus, image recognition is defined as:

$$\mathbf{y} = \min(f(\mathbf{x}, \mathbf{F}, \mathbf{T}, \mathbf{a})), \quad (2)$$

where the dataset is given by  $(\mathbf{y}, \mathbf{z})$ ,  $\mathbf{F}$  denotes the set of functions,  $\mathbf{T}$  represents the terminal set, and  $\mathbf{a}$  describes the parameters tuning the algorithm. To solve the problem, we require a method of feature extraction and a suitable criterion  $\mathbf{SC}$  for the minimization. The methodology requires the definition of two parts: 1) the AVC is the algorithm in charge of feature extraction, and 2) BP is the algorithm used to tuning  $(\mathbf{F}, \mathbf{T}, \mathbf{a})$  for each visual operator embedded into the AVC.

Regarding the algorithm to minimize the criterion  $\mathbf{SC}$ , we propose a Multi-layer Perceptron (MLP) as classifier that is used to learn a mapping  $f(\mathbf{x})$  where the descriptors  $\mathbf{x}_i$  are associated to

Table 1. Related Works

Reference	Origen dataset	Cell type	Images	Image resolution	Color space	Descriptor	Classifier
[6]	Public	Leukemia & healthy	300	1712X1368, 257x257	CIELab	shape,texture, color, derived	PSO <sup>a</sup> , other
[20]	Public	Leukemia & healthy	260	257x257	CMYK	shape,texture, color, wavelet	SSOA <sup>b</sup>
[12]	Public,private	Leukemia & healthy	768	variable	RGB, CMYK	shape,texture, color	SVM <sup>c</sup> , KNN <sup>d</sup> , NB <sup>e</sup> , DT <sup>f</sup>
[4]	Private	Leukemia: L1, L2, L3, M1, M2, M3, M5, M6	120	—	CIELab	shape,texture, color	Fuzzy DT <sup>f</sup>
[11]	Private	CLL Leukemia & healthy	1010	360x360	RGB	shape, derived	SVM <sup>c</sup> , ANN, KNN <sup>d</sup> , DT <sup>f</sup> , AdaBoost
[16]	Public	Leukemia & healthy	108	—	CMYK	shape,texture, color	SVM <sup>c</sup>
[18]	Private	Leukemia: L1, L2, M2, M3, M5	500	800x600	RGB	shape,texture, derived	PSO <sup>a</sup>
[23]	Public,private	Leukemia & healthy	891	variable	RGB	derived	CNN <sup>g</sup>
[17]	Public	Leukemia: L1, L2, M2, M3, M5 & healthy	420	1000x1000	RGB	shape,texture, color	SVM <sup>c</sup> , GA <sup>h</sup>
[9]	Private	Leukemia & healthy	295	2582x1948	RGB	shape,texture, color	SVM <sup>c</sup> , K-means

<sup>a</sup> PSO—Particle Swarm Optimization, <sup>b</sup> SSOA—Social Spider Optimization Algorithm, <sup>c</sup> SVM—Support Vector Machine, <sup>d</sup> KNN—K-Near Neighbor, <sup>e</sup> NB—Naive Bayes, <sup>f</sup> DT—Decision Tree, <sup>g</sup> CNN—Convolutional Neural Network, <sup>h</sup> GA—Genetic Algorithm

labels  $y_i$ . In this work, we address the problem as a multiclass classification task. Hence, it is assumed that in the minimization problem the variables  $((\mathbf{x}, \mathbf{y}), \mathbf{F}, \mathbf{T}, \mathbf{a}, \mathbf{SC})$  are related in such a way that the objective is to associate the descriptors (domain) and the labels (codomain).

Since BP is a paradigm consisting of two main stages, in the first stage, the purpose is to discover functions to optimize complex models by adjusting operations within them. In the second stage, the parts (programs) are applied to a hierarchical model for the feature extraction. It is noteworthy that the second stage uses the concept of composition of functions to extract features from images. Thus, an outstanding characteristic of BP is the possibility of changing the model to solve either focus of attention to produce saliency [3] and classification problems [2]. Fig. 1 presents the general scheme of the proposal.

### 3.1 Stages of Brain Programming

#### 3.1.1 Initialization

The evolutionary process of BP begins with a randomized initial generation. In this way, a set of initialization variables are defined such as population size, size of solutions or individuals, and crossing-mutation probabilities.

An individual represents a computer program written with a set of syntactic trees included in hierarchical structures. These individuals contain four kinds of functions, one for each visual operator (VO). Expert knowledge is used to define the procedures to create trees whose nodes are selected from a pool of functions and terminals, which are shown in Table 2. A more extensive description of these functions and terminals can be found in [2].

An individual consists of a set of functions taken from Table 2 and encoded in a multi-tree representation. A variable number of syntactic trees, ranging from four to 10, compose each individual. These trees regard each type of EVO (orientation, color, shape, and mental maps). Crossover and mutation operations were designed considering this representation. In this way, BP creates symbolic solutions (individuals) to recognition problems. We use the AVC model to deal with the leukemia recognition problem, so after completing each generation, the individuals in the population are evaluated to score the fitness.

#### 3.1.2 Feature Extraction and Classification with the Artificial Visual Cortex

In contrast to conventional evolutionary algorithms that commonly apply a fitness function to evaluate

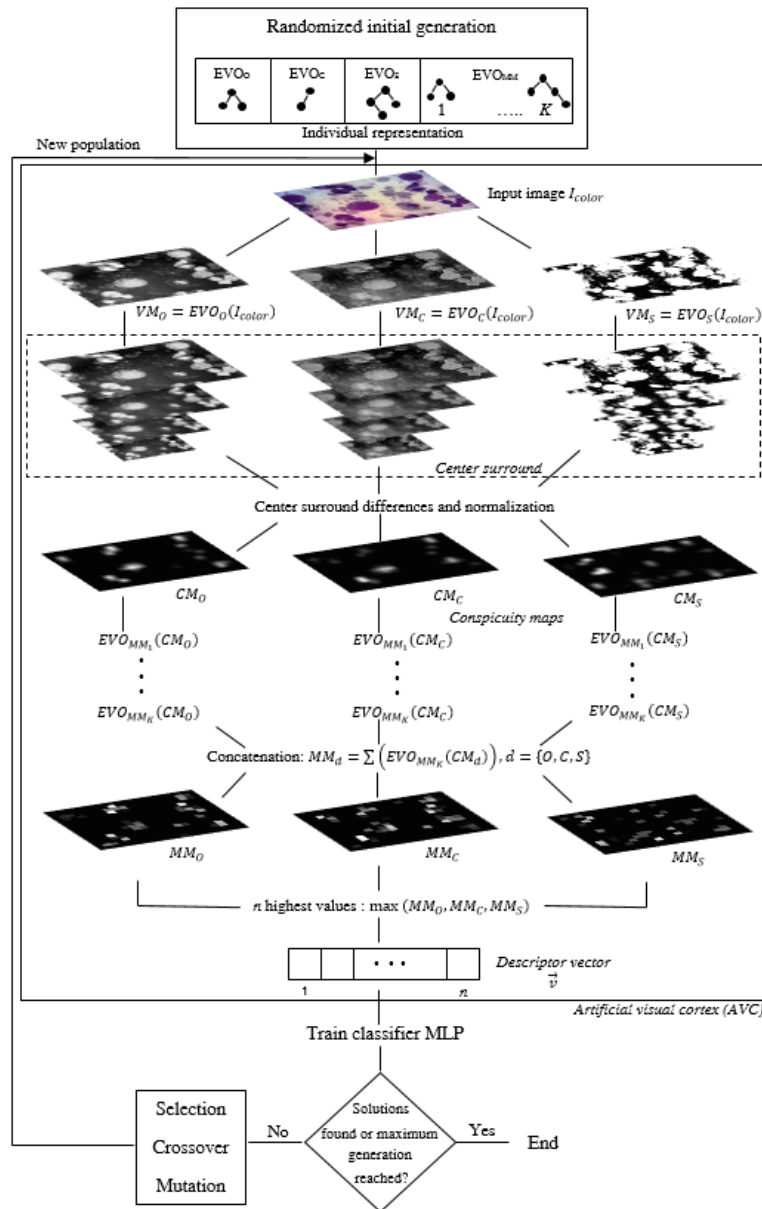


Fig. 1. General flowchart of the methodology

individuals' quality, in BP like GP, the evaluation consists of a set of EVOs designed to extract features from input images.

Since the AVC models some aspects of the human visual cortex, each layer of the artificial visual cortex computes mathematical operations that represent a visual function. The image's visual

features are selected to construct an abstract representation of the object of interest. In this way, the model finds salient points in the image to generate an image descriptor used for the classification.

The AVC is composed of two phases. In the first, the features that describe the object are acquired

and transformed, whereas, in the second phase, the descriptor obtained in the previous stage is used to classify the object.

The first phase is based on the psychological model of visual attention proposed by [22], in which basic features such as orientation, color, and shape are computed in parallel. Thus, the input to the model is an RGB image  $I$  defined as follows [15].

**Image as the graph of a function.** Let  $f$  be a function  $f : U \subset \mathbb{R}^2 \rightarrow \mathbb{R}$ . The graph or image  $I$  of  $f$  is the subset of  $\mathbb{R}^3$  that consists of the points  $(x, y, f(x, y))$ , in which the ordered pair  $(x, y)$  is the value at that point. This is, the image  $I = \{(x, y, f(x, y)) \in \mathbb{R}^3 | (x, y) \in U\}$ .

Note from this definition that images are variations in the intensity of light along the two-dimensional plane of camera sensors. In this way, multiple color channels are considered to create the set  $I_{color} = \{I_r, I_g, I_b, I_c, I_m, I_y, I_k, I_h, I_s, I_v\}$ , whose elements refer to the color components of RGB, HSV, and CMYK color spaces.

The following step is the decomposition of the image into relevant characteristics. The orientation, color, and shape dimensions transform independently the input images  $I_{color}$  to emphasize specific aspects of the object. Hence, individuals represent possible configurations for feature extraction that describe input images. These are optimized through the evolutionary process. After applying each EVO, a visual map (VM) generated for each dimension  $d$  (orientation, color, and shape) represents a partial output within the whole process. These are topographic maps that refer to the characteristics of the image.

From the obtained VMs, the next step is to compute a center-surround process. First, scale invariant features are extracted and stored in a conspicuity map (CM). The CM is calculated as the difference between different scales that are obtained through a pyramid of 9 levels  $P_d^\sigma = \{P_d^{\sigma=0}, P_d^{\sigma=1}, P_d^{\sigma=2}, \dots, P_d^{\sigma=8}\}$ . A Gaussian smoothing filter on each VM is used to calculate each pyramid. This produces an image half the size of the input map. The process is repeated 8 times to obtain the pyramid of 9 levels.

In the next step, the differences between each pyramid level  $P_d^\sigma$  are calculated using Eq. (3) as follows:

$$Q_d^j = P_d^{\sigma=\lfloor \frac{j+9}{2} \rfloor + 1} - P_d^{\sigma=\lfloor \frac{j+2}{2} \rfloor + 1}, \quad (3)$$

where  $j = 1, 2, \dots, 6$ . Each level of  $P_d^\sigma$  is normalized and scaled to the dimension of the VM using polynomial interpolation. Finally, the six levels are combined into a single map with a summation, and a CM is obtained for each dimension.

The second phase of the AVC begins with the description and classification. This phase aims to synthesize the whole information obtained into a vector descriptor of the image input to a MLP classifier. To begin, a mental map MM is built from the CMs using Eq. 4, where  $d$  is the dimension, and  $k$  is the cardinality of the set  $EVO_{MM}$ . This MM discriminates the unwanted information, highlighting the most salient features of the object. The EVOs are defined through syntactic trees, and the MMs occupies the fourth position of the tree onward:

$$MM_d = \sum_{i=1}^k EVO_{MM_i}(CM_d). \quad (4)$$

From the MMs obtained and concatenated with the remainder of syntactic trees, the generated program is applied to each image. The  $n$  highest values are used to define the descriptor vector  $\vec{v}$  for the image in turn. In this way, the next step is to train a classifier using the feature vectors from the dataset. In this work, a MLP is trained to create a model  $f(\mathbf{x})$  that maps a set of descriptors vectors  $\mathbf{x}_i$  to their corresponding labels  $\mathbf{y}_i$ , satisfying Eq. 2.

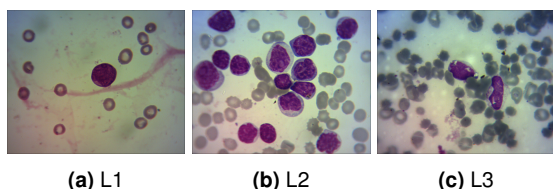
Selection, crossover, and mutation processes are performed as suggested in [15]. Finally, the stop conditions are: (1) the algorithm reaches a predefined number of generations, or (2) the algorithm fitness reaches an optimal value; in this case that all images are correctly classified.

## 4 Experiments and Results

Experiments were executed on a computer with Intel Core i9-7900X CPU 3.31Ghz, 64GB RAM, 222Gb hard drive, 64-bit Windows10 Enterprise

**Table 2.** Functions and Terminals for the evolutionary visual operators (EVOs)

Functions	Terminals
<b>Orientation (<math>EVO_o</math>)</b>	
$A + B, A - B, A * B, A/B,  A + B ,  A - B , \inf(A, B), \sup(A, B), \sqrt{A}, A^2, \log(A), \text{thr}(A), \text{round}(A), \lfloor A \rfloor, \lceil A \rceil, G_{\sigma=1}(A), G_{\sigma=2}(A),  A , D_x(A), D_y(A), k + A, k - A, k * A, k/A$	$I_r, I_g, I_b, I_c, I_m, I_y, I_k, I_h, I_s, I_v, G_{\sigma=1}(I_x), D_x(I_x), D_y(I_x), D_{yy}(I_x), D_{xx}(I_x), D_{xy}(I_x)$
<b>Color (<math>EVO_c</math>)</b>	
$A + B, A - B, A * B, A/B, k + A, k - A, k * A, k/A, \text{thr}(A), \text{round}(A), \lfloor A \rfloor, \lceil A \rceil, \sqrt{A}, A^2, \log(A), (A)^c, \exp(A)$	$I_r, I_g, I_b, I_c, I_m, I_y, I_k, I_h, I_s, I_v, \text{Oppr-g}(I_{rgb}), \text{Oppb-y}(I_{rgb})$
<b>Shape (<math>EVO_s</math>)</b>	
$A + B, A - B, A * B, A/B, k + A, k - A, k * A, k/A, \text{thr}(A), \text{round}(A), \lfloor A \rfloor, \lceil A \rceil, A \oplus SE_{dm}, A \oplus SE_s, A \oplus SE_d, A \ominus SE_{dm}, A \ominus SE_s, A \ominus SE_d, Sk(A), \text{Perim}(A), A \otimes SE_{dm}, A \otimes SE_s, A \otimes SE_d, \text{That}(A), \text{Bhat}(A), A \odot SE_s, A \odot SE_s$	$I_r, I_g, I_b, I_c, I_m, I_y, I_k, I_h, I_s, I_v$
<b>Mental Maps (<math>EVO_{MM}</math>)</b>	
$A + B, A - B, A * B, A/B,  A + B ,  A - B , \sqrt{A}, A^2, \log(A), D_x(A), D_y(A),  A , k * A, G_{\sigma=1}(A), G_{\sigma=2}(A)$	$CM_d, D_x(CM_d), D_y(CM_d), D_{xx}(CM_d), D_{yy}(CM_d), D_{xy}(CM_d)$

**Fig. 2.** Types of lymphocytic leukemia cells

Edition operating system, graphics processing unit (GPU) GeForceGTX 1080, and MATLAB R2018a.

From parameters values in Table 3, the evolutionary loop starts by computing the fitness of each AVC using an MLP to calculate the classification rate using the training and validation sets. The MLP has one intermediate layer with 50 neurons.

In the next step, a set of AVCs is selected from the population with a probability based on fitness using a roulette-wheel selection, and the best AVC is retained for further processing. The new individual is created from the selected AVC by applying a crossover or mutation at chromosome or gene levels as in [15]. Although we do not use

**Table 3.** Initialization values for the algorithm

Parameter	Description
Generation	30
Population size	30 individuals
Initialization	Ramped Half and Half
Crossover rate	0.4
Mutation	0.1
Tree depth	Dynamic depth selection
Dynamic max depth	50 levels
Maximum length of genes	10
Selection	Roulette-wheel
Elitism	Keep the best individual

a strategy for bloat handling, we limit to 10 the maximum gene length.

The evolutionary loop ends until a classification rate is equal to 100%, or the algorithm reaches the number of generations  $N = 30$ .

#### 4.1 Dataset

The dataset used is composed of bone marrow smear images from three subtypes of Acute Lymphoblastic Leukemia (ALL): L1, L2, and L3, like in previous work [13]. Images RGB are

in BMP format with a resolution of  $1280 \times 1024$  pixels. Image acquisition was made employing an optical microscopic with a magnification of 1250 times and a camera coupled to the microscope with a resolution of 1.3 megapixels. The images were resized to  $256 \times 320$  pixels using bicubic interpolation due to the high computational cost. We use 217 images per class. Typically, the images contain one or more interest cells that appear like irregular purple regions, as you can see in Fig. 2.

We divide the dataset into three parts; the learning set, the validation set, and the testing set. Fig. 3 shows the details of the data division. To obtain a reliable fitness, each new individual is estimated by the average classification error rate with the MLP using five-fold cross-validation. The learning set is randomly divided into five equal parts and perform five training cases with the MLP on 4 out of 5 and the result is computed with the remaining validation 1 set.

To select the best-performed solution, we test the classification error for every fold on the validation set 2. Hence, we select one solution with the best validation error as the (near-)optimal feature descriptor for the final testing.

Finally, the test set is divided as a five-fold with the aim of computing statistical results of the best solution discovered in the previous stage. We apply the same process for the learning set, and the overall classification result is calculated as the average of the 5 MLP test-fold accuracies.

## 4.2 Results

The following is a description of the results from the evolutionary process to recognize leukemia images. To evaluate the proposed method, we repeat the above process seven times. From these, the best solution is shown in Table 4; while Fig. 4 illustrates the range of descriptor values of the best solution found.

Since we use a balanced dataset and 261 images for the test, in Fig. 4 the image index 1...87 corresponds to class L1, the following 87 to class L2 images, and the remainder to group L3 images. It is worth noting that the descriptors' values clearly show the difference between categories,

which in some cases can be evaluated with simple techniques instead of the MLP classifier; thus simplifying the overall process.

The depth and number of nodes quantify the complexity of the best individual, see Fig. 5 (a)-(b). It depicts the complexity of the evolutionary run for each generation, and we recognize that both variables—the number of nodes and depth of trees—decrease as the generations progress. This means that the final solution is of lower complexity. The genetic diversity found in the population at each generation along the run that produces the fittest individual in the experiment is presented in Fig. 6.

Diversity is defined as the percentage of operators' uniqueness within the population. It should be noted that the best individual is a structure of EVOs and same as complexity, diversity decrease as the generations progress.

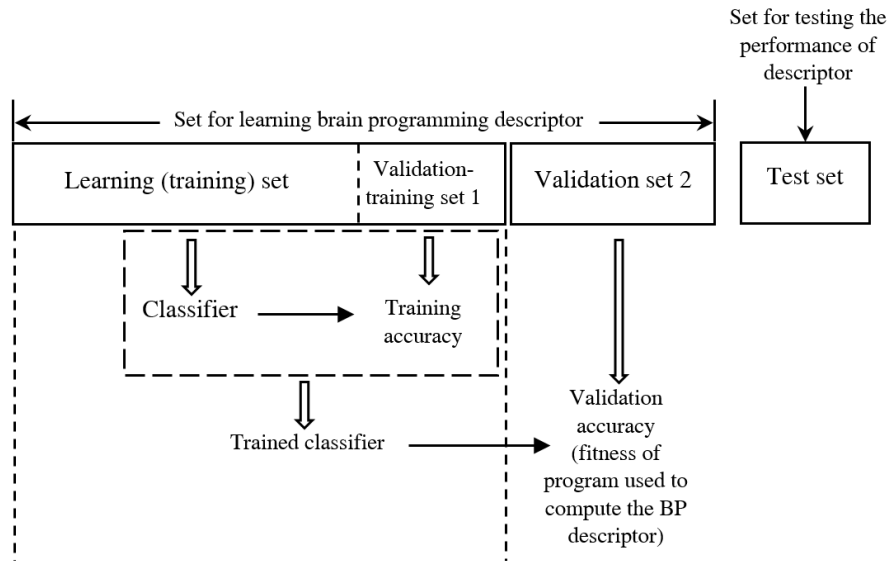
Fitness behavior along the evolution is shown in Fig. 7. It depicts average, standard deviation, and best so far run corresponding to the best solution.

Additionally, we evaluated the proposal with a balanced dataset composed of 261 images of 3 types of myelocytic leukemia: M3, M4, and M5. This dataset was acquired under the same conditions mentioned in Section 4.1. Fig. 8 presents typical images of these cells.

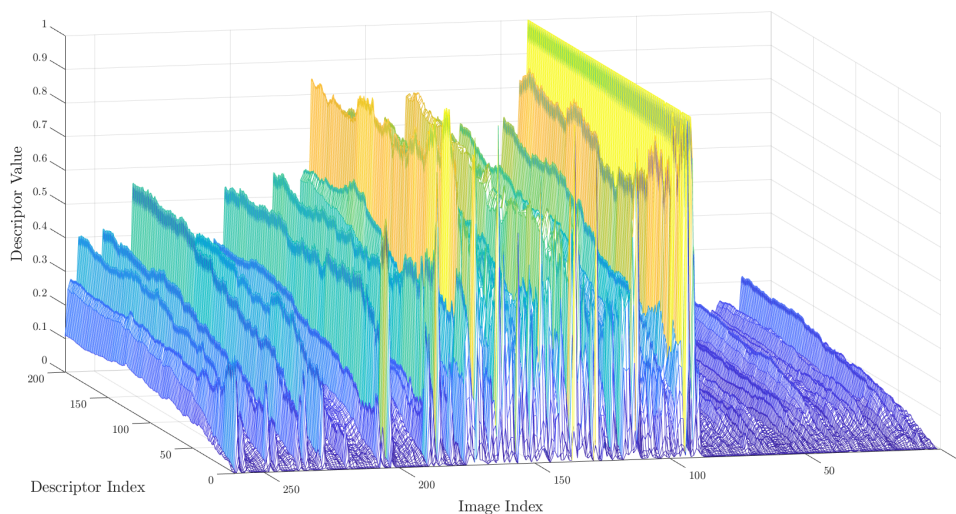
In contrast to the above experiment, in this evaluation, we executed only two experiments. However, the results are similar for both datasets. The validation accuracy for the best solution was 91.02%, which is a competitive result considering that the content of images is visually more different than the previous problem.

Although we do not find a similar symbolic learning approach for the recognition of these subtypes of leukemia, we provide next another proposal for comparison with our model. Thus, experiments with a convolutional neural network have been performed using the same datasets.

The network structure consists of four convolutional and max-pooling layers. The feature maps are flattened and reduced to an output of size three. Data augmentation was used on the training set. Augmentation operations contained horizontal and vertical reflexion (flipping).



**Fig. 3.** Division of dataset



**Fig. 4.** Descriptors of the best solution

Accordingly, the result is the addition of a random number of augmented images to the training set in each epoch.

The number of filters in the four convolutional layers are 4, 8, 16, and 32. Five-fold cross-validation was used to assess training

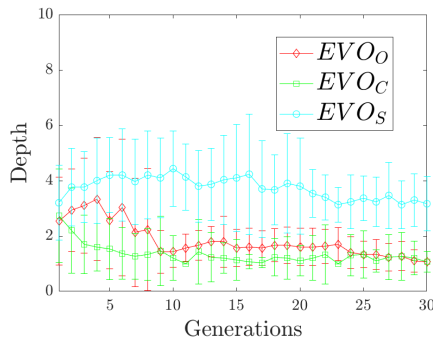
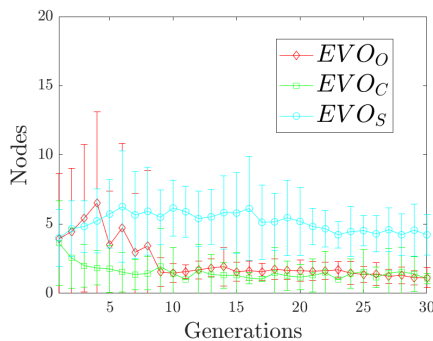
performance using the sets of learning and validation shown in Fig. 3. The training was stopped when the validation loss did not decrease for 20 epochs.

After this, the best net model from the five-fold was selected and used with the test set to assess



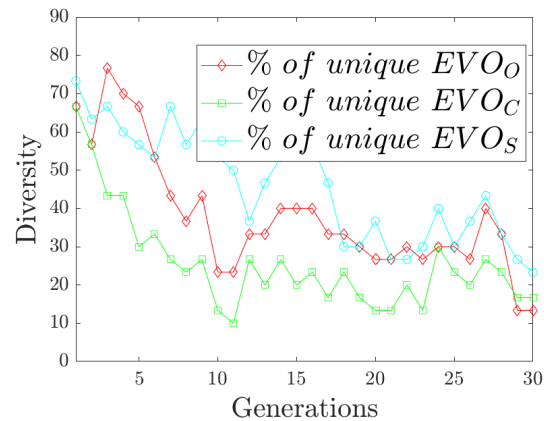
**Table 4.** Structure of best solution after seven experiments

Validation accuracy=92.82%
$EVO_o = G_{\sigma=1}(S)$
$EVO_c = M$
$EVO_s = \text{round}(0.36 - K)$
$EVO_{MM_1} = D_y(D_y(CM_d))$
$EVO_{MM_2} = G_{\sigma=1}( D_x(D_y(CM_d)) /D_y(CM_d))$
$EVO_{MM_3} = D_y(D_y(    D_x(D_x(CM_d)) + CM_d    ))$
$EVO_{MM_4} = G_{\sigma=1}(D_x(CM_d)/D_x(D_x(CM_d)))$

**(a) Depth****(b) Nodes****Fig. 5.** Complexity of the best individual

the net performance.

To evaluate the global classification performance, we repeated the experiment for ten times, from which an accuracy of  $97.05\% \pm 2.56$  (mean $\pm$ s.d., runs=10) for the classes L1, L2 and L3; and an accuracy of  $97.51\% \pm 1.37$  (mean $\pm$ s.d., runs=10) for the classes M3, M4 and M5. For one of the folds, the training and validation accuracy are shown in Fig. 9.

**Fig. 6.** Genetic diversity

From the results, it is worth noting that although a CNN shows higher performance for the classification task, a critical problem is the lack of information about the learning process that led to the solution. This fact is undesirable since experts need to identify the illness from the features of the image and to know the way to learn to recognize these cell types.

## 5 Conclusions

This work proposes the use of evolutionary vision for the recognition of leukemia cell images. Since the biological visual cortex inspires the methodology, then visual extraction and description are addressed through a hierarchical structure and function composition using a set of mathematical operations. Hence, the results show that all functions embedded within this structure can be discovered by the evolutionary cycle.

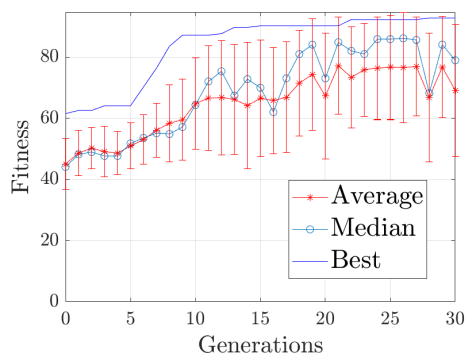


Fig. 7. Fitness behavior of the best solution

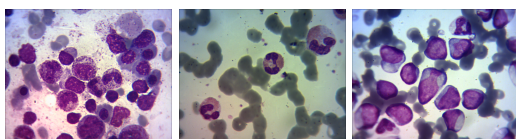


Fig. 8. Types of myelocytic leukemia cells

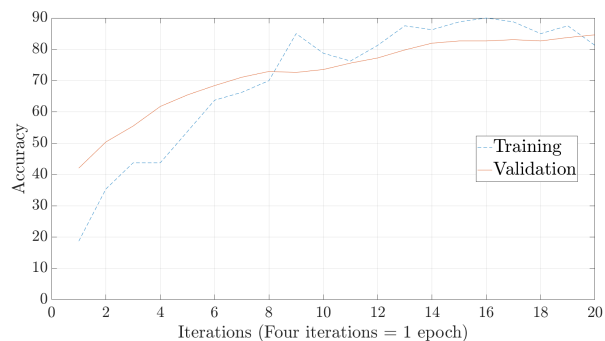
Furthermore, the characteristics of the proposed methodology apply new structures for the machine vision task. As it has been shown, the approach also can be framed as an optimization problem due to its structural characteristics, and as a consequence, is susceptible to being improved.

In conclusion, since in the problem of leukemia cells recognition, it is of utmost importance to know how the features are derived in a natural way, as well as their meaning and their significance for the recognition task, in this work was proposed the use of symbolic learning as a white box methodology to study the problem of leukemia cells recognition.

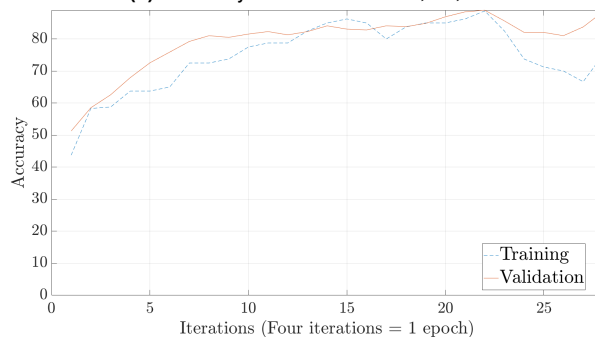
Thus, the model can be applied in diverse fields. The task of object recognition requires a clear explanation to understand better the subject of the research.

## Acknowledgments

Authors would like to acknowledge the support provided by the Instituto Politécnico Nacional under projects: 20200630 and 20210788, CONACYT under projects: 65 (Fronteras de la Ciencia) and



(a) Accuracy for the classes L1, L2, L3



(b) Accuracy for the classes M3, M4, M5

Fig. 9. Performance of the convolutional neural network

6005 (FORDECYT-PRONACES), and CICESE through the project 634-135 to carry out this research. First author thanks the Autonomous University of Tlaxcala, Mexico for the support. Authors also express their gratitude to the Applied Computational Intelligence Network (RedICA).

## References

1. **Bhattacharjee, R., Saini, L.-M. (2015).** Robust technique for the detection of acute lymphoblastic leukemia. 2015 IEEE Power, Communication and Information Technology Conference (PCITC), pp. 657–662. DOI: 10.1109/PCITC.2015.7438079.
2. **Chan-Ley, M., Olague, G. (2020).** Categorization of digitized artworks by media with brain programming. *Appl. Opt.*, Vol. 59, No. 14, pp. 4437–4447. DOI: 10.1364/AO.385552.
3. **Dozal, L., Olague, G., Clemente, E., Hernández, D. E. (2014).** Brain programming for the evolution of an artificial dorsal stream. *Cognitive Computation*,

- Vol. 6, No. 3, pp. 528–557. DOI: 10.1007/s12559-014-9251-6.
4. **Fatichah, C., Tangel, M.-L., Yan, F., Betancourt, J.-P., Widyanto, R., Dong, F., Hirota, K. (2015).** Fuzzy feature representation for white blood cell differential counting in acute leukemia diagnosis. *International Journal of Control, Automation and Systems*, Vol. 13, No. 3, pp. 742 – 752. DOI: <https://doi.org/10.1007/s12555-012-0393-6>.
  5. **Haugeland, J. (1985).** *Artificial Intelligence: The Very Idea*. Cambridge, Mass: MIT Press.
  6. **Jothi, G., Hannah, I., Ahmad, T.-A., K-Renuga, D. (2018).** Rough set theory with jaya optimization for acute lymphoblastic leukemia classification. *Neural Computing and Applications*, Vol. 31, No. 9, pp. 5175 – 5194. DOI: <https://doi.org/10.1007/s00521-018-3359-7>.
  7. **Kampen, K.-R. (2012).** The discovery and early understanding of leukemia. *Leukemia research*, Vol. 36, No. 1, pp. 6–13. DOI: 10.1016/j.leukres.2011.09.028.
  8. **Khobragade, S., Mor, D.-D., Patil, C.-Y. (2015).** Detection of leukemia in microscopic white blood cell images. 2015 International Conference on Information Processing (ICIP), pp. 435–440. DOI: 10.1109/INFOP.2015.7489422.
  9. **Laosai, J., Chamnongthai, K. (2018).** Classification of acute leukemia using medical-knowledge-based morphology and cd marker. *Biomedical Signal Processing and Control*, Vol. 44, pp. 127 – 137. DOI: <https://doi.org/10.1016/j.bspc.2018.01.020>.
  10. **Mishra, S., Majhi, B., Sa, P.-K. (2016).** A survey on automated diagnosis on the detection of leukemia: A hematological disorder. 2016 3rd International Conference on Recent Advances in Information Technology (RAIT), pp. 460–466. DOI: 10.1109/RAIT.2016.7507945.
  11. **Mohammed, E.-A., Mohamed, M.-M.-A., Naugler, C., Far, B.-H. (2017).** Toward leveraging big value from data: chronic lymphocytic leukemia cell classification. *Network Modeling Analysis in Health Informatics and Bioinformatics*, Vol. 6, pp. 6 – 23. DOI: <https://doi.org/10.1007/s13721-017-0146-9>.
  12. **Moshavash, Z., Danyali, H., Helfroush, M.-S. (2018).** An automatic and robust decision support system for accurate acute leukemia diagnosis from blood microscopic images. *Journal of Digital Imaging*, Vol. 31, No. 5, pp. 702 – 717. DOI: <https://doi.org/10.1007/s10278-018-0074-y>.
  13. **Ochoa-Montiel, R., Martínez, L., Sossa, H., Olague, G. (2020).** Handcraft and automatic approaches for the recognition of leukemia images. *Research in Computing Science*, Vol. 149, No. 11, pp. 271 – 280.
  14. **Ochoa-Montiel, R., Olague, G., Sossa, H. (2020).** Expert knowledge for the recognition of leukemic cells. *Appl. Opt.*, Vol. 59, No. 14, pp. 4448–4460. DOI: 10.1364/AO.385208.
  15. **Olague, G., Clemente, E., Dozal, L., Hernández, D. (2014).** Evolving an artificial visual cortex for object recognition with brain programming. In **Schuetze, O., Coello, C. A. C., Tantar, A.-A., Tantar, E., Bouvry, P., Moral, P. D., , Legrand, P.**, editors, *EVOLVE - A Bridge between Probability, Set Oriented Numerics, and Evolutionary Computation III*, volume 500 of *Studies in Computational Intelligence*. Springer Heidelberg, pp. 97–119. DOI: 10.1007/978-3-319-01460-9-5.
  16. **Putzu, L., Caocci, G., Di-Ruberto, C. (2014).** Leucocyte classification for leukaemia detection using image processing techniques. *Artificial Intelligence in Medicine*, Vol. 62, No. 3, pp. 179 – 191. DOI: <https://doi.org/10.1016/j.artmed.2014.09.002>.
  17. **Rawat, J., Singh, A., HS, B., Virmani, J., Devgun, J.-S. (2017).** Computer assisted classification framework for prediction of acute lymphoblastic and acute myeloblastic leukemia. *Biocybernetics and Biomedical Engineering*, Vol. 37, No. 4, pp. 637 – 654. DOI: <https://doi.org/10.1016/j.bbe.2017.07.003>.
  18. **Reta, C., Altamirano, L., Gonzalez and Raquel Díaz-Hernández, J., Peregrina, H., Olmos, I., Alonso, J., Lobato, R. (2015).** Correction: Segmentation and classification of bone marrow cells images using contextual information for medical diagnosis of acute leukemias. *PLoS ONE*, Vol. 10, No. 07. DOI: 10.1371/journal.pone.0134066.
  19. **Rodak, B. F., Carr, J. H. (2016).** *Clinical Hematology Atlas*. W. B. Saunders Co.
  20. **Sahlol, A.-T., Abdeldaim, A.-M., Hassanien, A.-E. (2018).** Automatic acute lymphoblastic leukemia classification model using social spider optimization algorithm. *Soft Computing*, Vol. 23, No. 15, pp. 6345 – 6360. DOI: <https://doi.org/10.1007/s00500-018-3288-5>.
  21. **Savkare, S.-S., Narote, S.-P. (2015).** Blood cell segmentation from microscopic blood images. 2015 International Conference on Information Processing

(ICIP), pp. 502–505. DOI: 10.1109/INFOP.2015.7489435.

22. **Treisman, A.-M., Gelade, G. (1980).** A feature-integration theory of attention. *Cognitive Psychology*, Vol. 12, No. 1, pp. 97 – 136. DOI: [https://doi.org/10.1016/0010-0285\(80\)90005-5](https://doi.org/10.1016/0010-0285(80)90005-5).
23. **Vogado, L. H., Veras, R. M., Araujo, F. H., Silva, R. R., Aires, K. R. (2018).** Leukemia diagnosis in blood slides using transfer learning in cnns and

svm for classification. *Engineering Applications of Artificial Intelligence*, Vol. 72, pp. 415 – 422. DOI: <https://doi.org/10.1016/j.engappai.2018.04.024>.

24. **World Health Organization (2021).** Cancer.

*Article received on 11/08/2020; accepted on 21/12/2020.  
Corresponding author is Humberto Sossa.*