

Evaluation of Heat Map Methods Using Cell Morphology for Classifying Acute Lymphoblastic Leukemia Cells

José de J. Velázquez-Arreola, Nohemí Sánchez-Medel,
Oliver A. Zarraga-Vargas, Raquel Díaz-Hernández*,
Leopoldo Altamirano-Robles

National Institute of Astrophysics, Optics and Electronics, Puebla,
Mexico

raqueld@inaoep.mx

Abstract. Explainable artificial intelligence (XAI) is a field of research that has attracted the interest of researchers in recent years. These algorithms seek to provide transparency to artificial intelligence (AI) models. One application of these algorithms is in the medical area, created as an auxiliary tool for corroborating predictions obtained by an AI when classifying pathologies, for example, Acute Lymphoblastic Leukemia (ALL). The present work evaluates the amount of information heat maps provide and how they relate to the blood components' morphological characteristics. For the assessment, four Convolutional Neural Network (CNN) models were retrained and fine-tuned to classify unsegmented images (ALL_IDB2 database). Subsequently, their respective heat maps were generated with the LRP (Layer-wise Relevance Propagation), Deep Taylor, Input*Gradient, and Grad-Cam methods. The best results were obtained with the GoogleNet model and the Grad-Cam heat map generation method, having a percentage of 43.61% of relevant pixels within at least one cell morphological feature present. Moreover, the most significant pixels are within the nucleus, with 73.97% of important pixels inside. According to the results, the Grad-Cam method best relates the relevant pixels generated in the heat map to the morphology of the cell of interest to classify a healthy or diseased cell.

Keywords. Explainable artificial intelligence (XAI), heatmaps, acute leukemia lymphoblastic (ALL), grad-cam, cell morphology.

1 Introduction

Deep learning applications, surpassing human capabilities in tasks like image and speech

recognition and recommendation systems, have received substantial attention.

Despite their achievements, these applications cope with a critical shortfall in both explainability and reliability.

Deep learning models are commonly perceived as complex black boxes, presenting challenges in understanding their intricate underlying mechanisms. Their inability to justify decisions and predictions undermines human trust, heightening concerns, especially considering the potential life-threatening errors that artificial intelligence algorithms may make depending on the application.

For example, a flaw in the computer vision system of an autonomous car could result in a catastrophic crash, while in healthcare, where decisions directly impact human lives, the stakes are considerably high.

In response to these challenges, many methods have emerged to address the need for transparency and reliability in deep learning applications. Notably, explainable Artificial Intelligence has emerged as a focal point in machine learning research.

These methods aim to explain machine and deep learning models in a manner easily understandable by humans. The categorization of interpretability methods is based on how they provide explanation information, encompassing visual, textual, and mathematical or numerical approaches. This paper evaluates different visual interpretability methods for classifying Acute Lymphoblastic Leukemia cells.

1.1 The Problem

In the medical field, there has been an increase in the activity of digitizing pathological studies for medical diagnosis. The digitalization opens the door to life-saving artificial intelligence (AI) applications. One of the branches of application of these explanatory methods is to use them as an auxiliary tool in validating the predictions made by a neural network.

However, to diagnose some diseases through digital microscopic images, as in the case of Acute Lymphoblastic Leukemia (ALL), it is necessary to pay attention to the morphological characteristics present in the cells of interest.

From this arises the need to evaluate whether heat maps, as a visual explanatory method, are appropriate to highlight the morphological characteristics present. Thus, the expert can consider them as an aid for the corroboration of the classification, giving rise to the diagnosis of the disease.

1.2 Cell Morphology of Acute Lymphoblastic Leukemia

White blood cells are essential to the human body's immune system. They have a specific morphology depending on the type of blood component. Leukemia is an alteration in the production and malformation of these cells.

The French-American-British (FAB) classifies Leukemia into Acute and Chronic Lymphoblastic Leukemia. It also categorizes them into ALL-L1: small uniform cells; ALL-L2: large, varied cells; and ALL-L3: large, mixed cells with vacuoles (bubble like features).

The nuclear and cytoplasmic structures can differentiate between healthy and diseased cells. Acute Myeloid Leukemia (AML) and Chronic Myeloid Leukemia (CML) are also caused by abnormal myelocytes.

The authors in [1, 2, 3], argue that benign and malignant cells can be discriminated by their nuclear structure, nucleus-to-cytoplasm ratio, color, and texture. Fig. 1 illustrates an example of the difference between the structure of healthy and cancerous cells.

It shows that cancer cells have an irregular structure, and the shape of the nucleus is

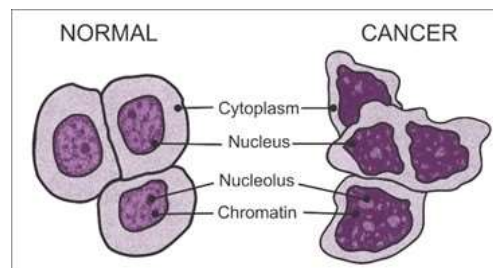


Fig. 1. Example of the difference between the structure of healthy and cancerous cells collapsed; this is how hematologists describe the ALL.

This work evaluates whether the heat maps produced by selected methods are related to at least one of these morphological features. This work evaluates whether the heat maps produced by selected methods are related to at least one of these morphological features.

1.3 Explanatory Methods

In this work, we use methods that generate heat maps; these are considered a way to explain the functioning of a neural network. In [4] define three concepts that are usually misused and interchanged. As a result of their research, they conclude the following definitions:

- Interpretability is the ability to explain or provide meaning in terms understandable to a human being.
- Explainability is associated with the notion of explanation as an interface between humans and decision-makers. At the same time, it accurately represents the decision-maker and is understandable to humans.
- Transparency: A model is considered transparent if it is understandable. Since a model can have different degrees of comprehensibility, transparent models are divided into three categories: simulatable, decomposable, and algorithmically transparent.

Explainability is critical for the safety, approval, and acceptance of AI systems for clinical use. At work [5] is a comprehensive overview of techniques that apply XAI to improve various properties of ML models and systematically

classifies these approaches, comparing their respective strengths and weaknesses.

In recent years, different heat map generation algorithms have been proposed to understand neural networks better. These methods include Deep Taylor, Input*Gradient, and LRP, among others. However, comparing results between these methods is somewhat complex because it is necessary to replicate each of these methods separately.

The author in [6] made available the iNNvestigate library; this tool solves the problem of method comparison, providing a standard interface and implementing several published methods for heat map generation, facilitating the analysis of neural network predictions by generating heat maps. This work uses this library to create heat maps of the strategies implemented therein, specifically the Deep Taylor, Input*Gradient, and LRP methods.

Grad-Cam [7], uses the gradient of the classification score related to the convolutional features determined by the network to understand which parts of the image are most important for classification. For this work, the algorithm implemented in MATLAB software is used.

1.4 Retrained Models

Four different models of CNNs were used in this work:

- VGG-19 is a convolutional neural network with 19 depth layers [8].
- ResNet18 is a neural network with 18 depth layers [9].
- ResNet50 is a neural network with 50 depth layers [9].
- GoogleNet is a neural network with 22 depth layers [10].

These neural networks were pre-trained with more than one million images from the ImageNet database. The pre-trained network can classify images into 1000 object categories (e.g., keyboard, mouse, pencil, and many animals). As a result, the network has learned feature-rich representations for a wide range of objects. The size of the network's image input is 224 by 224 pixels.

The idea behind selecting these architectures was to experiment with small and large models. In addition to belonging to the best-known models, these models usually perform better when transfer learning is done using other datasets. Therefore, these models were retrained with the database images described in section 3.1.

2 Related Work

Table 1 compares recent works that present different techniques to solve the problem. All authors focus on classifying images containing ALL cell types, using techniques to relate them to the morphology of the cell. Most authors using CNN models highlight the ResNet50, VGG's, and Inceptions models. On the other hand, few authors use a method of visual explanation.

The authors in [11], evaluate different algorithms to calculate heat maps using a hematologist specialized in ALL diagnosis. Generated heat maps were assessed with the help of five hematologists and experts in morphological cell classification. The evaluation focused on the amount of information provided by the heat maps and how they relate to morphological characteristics present in the classified cells.

Results of the best heatmaps and hematologist evaluations are presented in this work. The central outcome is that the heatmaps must include morphological information to be a valuable tool for medical diagnosis systems.

Following the same line of research expressed in [11], the present work represents an extension in the sense that a reference map with the morphology of each cell in the analyzed images was produced, which allows quantification of what percentages of the pixels marked as significant by the heat map, fall into morphologically coherent entities. This way, evaluating which algorithm is more relevant concerning cell morphology is possible.

In [22] it is proposed a method of classification and explanation. The proposed method contemplates segmentation of the morphological characteristics of the cells. Subsequently, the method uses a ResNet50 network that performs the classification, obtains the respective heat map, and generates an explanation of spatial features.

Table 1. Comparative table of related works

Author	Type Blood Cell	Model	XAI Method
N. Jiwani et al. in [12]	ALL	No	No
Jiang et al. [13]	ALL	Wavelet	No
Abir et al. [14]	ALL	Resnet50, DenseNet121 and VGG16	LIME
Nayeon Kim [15]	ALL Pro-B	InceptionV3, Res-Net101V2, InceptionResNetV2, and VGG19	LIME and DCN
Ochoa-Montiel et al. [16]	ALL	Random Forest, LeNet, AlexNet	No
Maaliw R.R. et al. [17]	ALL	Transfer learning InceptionV3, Xception InceptionResNetV2	No
Velázquez-Arreola et al. [19]	ALL	VGG16, VGG19, ResNet50, and MobileNet V1	LRP, Deep Taylor, and Input*Gradient methods
Diaz R. J. et al. [22]	ALL	Modified ResNet-50	Grad CAM

Both steps are considered visual explanatory methods. In addition, they perform heat map generation experiments, with the cell segmented and unsegmented. They conclude that better results are obtained when the heat map is generated with the segmented cell.

The main difference between the work presented here and the work discussed above is that we evaluate the number of most relevant pixels according to the segmentation categories. Experiments are also performed by combining different CNN models and heat mapping methods.

The experiments aim to identify which combination is the most efficient in relating relevant pixels against morphological features using the unsegmented image. Consequently, our work differs substantially from the work explained above.

3 Methodology

This section describes the database used for retraining and evaluation. It also presents the semi-manual segmentation procedure to create the reference map. This map will be used to compare the heat maps and thus evaluate the position of the most relevant pixels. Finally, the general methodology of this article is described.

3.1 Image Database

The Acute Lymphoblastic Leukemia Imaging Database for Image Processing (IDB-ALL) [17] is publicly accessible, and the categories are balanced. It features microscopic images of blood samples. It is a database intended to evaluate and compare algorithms for image segmentation and classification in Acute Lymphoblastic Leukemia (ALL).

Each image in the database was identified and classified by a group of oncologists with expertise in identifying ALL diseased cells. The photos are divided into diseased and healthy, with 130 images for each category. The images have a resolution of 257 x 257 pixels in RGB.

This work divided the images into 100 images for each category (healthy and diseased). With the remaining subset of images, 30 per category, a section of never-before-seen images was created. The latter will be used exclusively to evaluate the models after the entire retraining process and generate heat maps.

As described in [16], the typical datasets used for leukemia cell recognition have drawbacks related to category imbalance problems. In other cases, they were constructed from different sources or acquisition conditions. Leukemia cells

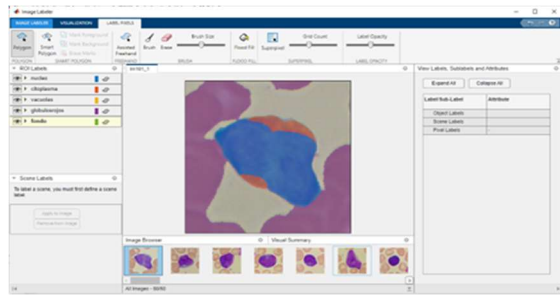


Fig. 2. Segmentation of an image from the IDB-ALL2 database using Image Labeler software

contain two or more categories of Leukemia, including its subtypes (Lymphocytic Leukemia and Myeloid Leukemia, Chronic or Acute). For our purpose, this work is focused on the ALL type.

The ALL-IDB2 dataset [17] is small. However, it is one of the most widely used datasets. For example, in [3, 16, 18], is publicly accessible, and the categories are balanced. For this reason, an ALL-IDB2 data set was selected for this work.

3.2 Data Segmentation (Semi-Manual)

As mentioned in section 1.2, cell morphology is used to identify ALL diseased cells. The main characteristics focus on the nucleus and cytoplasm. For this reason, a semi-manual segmentation is performed, highlighting five classes: nucleus, cytoplasm, vacuoles, red blood cells, and background.

The segmentation results will be considered the base reference (ground truth), which will later be used to evaluate the heat maps. The evaluation compares the most relevant pixels in each heat map and the ground truth corresponding to the original image.

The segmentation was performed using the MATLAB Image Labeler [19]. This application allows labeling reference images from a collection of pictures, defining rectangular region of interest (ROI) labels with aligned or rotated axes, line ROI labels, pixel ROI labels, polygon ROI labels, point ROI labels, projected cuboid ROI labels, and scene labels.

Fig. 2 illustrates how image segmentation from the IDB-ALL2 database was performed with the Image Labeler application. The segmentation process was carried out by two doctoral students

who worked on this research and supervised by two hematologists with experience in cellular morphology from the Mexican public health system. The image was segmented into five categories: nucleus, cytoplasm, vacuole, red blood cells, and background. These categories were chosen at the suggestion of the hematologists.

3.3 General Methodological Process

The general methodology is composed of the following steps:

1. Obtaining the image database (ALL-IDB2). Then, separate the images into two folders: 200 images for training and 60 images that will be used as never-seen-before. Photographs of healthy and diseased cells are included.
2. Segmentation of each image using the Image Labeler application [19]. A matrix of the same size as the segmented image will be generated. These matrices will be the ground truth reference used for the evaluation.
3. Using the 200 images for retraining, we applied data augmentation by the traditional method (rotation and reflection) [20] to have 1000 images at the end for each type of cell (Healthy or ALL). Using these images, we finally retrained the neural networks GoogleNet, ResNet18, ResNet50, and VGG19.
4. Generate the heat maps using the retrained models. Heat maps are generated with the Deep Taylor, Input*Gradient, and LRP methods using the iNNvestigate library. Grad-Cam type heat maps are generated with Matlab software.
5. Evaluation of the heat maps. For this process, only the most essential pixels for the neural network during classification are considered. Due to the color map used, it is evident that the pixels with the highest relevance are located in the red channel of the heat map images. Therefore, this channel is used to evaluate the heat maps. The evaluation process is described as follows:
 - a. We used the pixels of the red channel. Calculate the mean color depth of all pixels in this channel. The obtained value will be the reference value to

narrow down the pixels with the highest relevance.

- b. Identify the pixels above the reference value, then locate each pixel in the image of heat maps and the ground truth matrix. Then, count all pixels for each class according to the segmentation.
 - c. Save the number of pixels in each category in a table for later analysis.
6. We produce some graphs of the results and analyze them. The results and their analysis are described in the following section.

4 Experiments and Results

This section describes the experiments and their results, such as images generated, results tables, and graphs to analyze them.

4.1 Heatmaps Generated with iNNvestigate

As mentioned at the end of section 2, this work is an extension of [11]. That work explains the reasons why the iNNvestigate library is used and why only the Deep Taylor, Input*Gradient, and LRP methods were used. The same procedures are analyzed in the present work by continuing that research. Fig. 3 shows some heat maps obtained for a cell classified as healthy and a cell classified as unhealthy.

The heat map generated with the Deep Taylor method shows over the entire image different degrees of relevance, using a single color to show the significance of the pixels for the neural network. However, most of the time, it offers higher levels of relevance at the edges of the cells of interest.

The heat maps obtained by the Input*Gradient method generate pixels according to the color scale ranging from blue to red, with blue being the pixels with the lowest relevance and red pixels with the highest relevance.

With this method, identifying the shapes or regions of greater importance to the neural network in classification is a little more complex. The complexity arises because it generates relevant and non-relevant pixels very close to each other and with poorly defined regions compared to other methods.

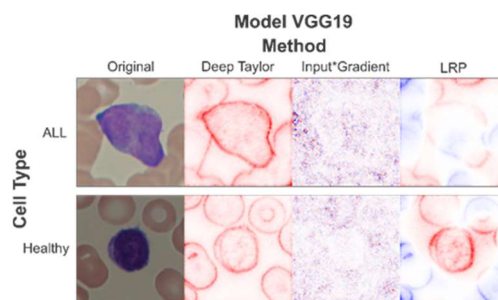


Fig. 3. Examples of heat maps obtained with the iNNvestigate library

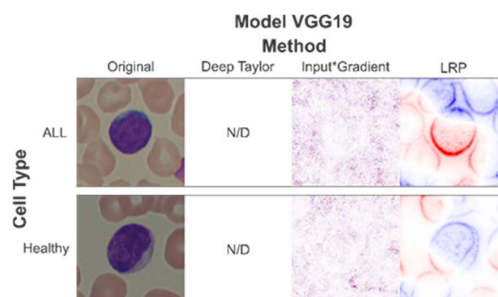


Fig. 4. Examples of heat maps where the Deep Taylor method is not defined

Finally, with the LRP method, the heat maps obtained, like the previous method, handle the color scale of blue and red. Unlike the Input*Gradient method, these heat maps show more clearly the regions of greater relevance to the neural network than those unimportant.

Fig. 4 shows two cells where the heat map is not defined for the Deep Taylor method. These cases, according to [6], are inherent to the technique since a value that works as a threshold is calculated. If this threshold is not exceeded, the heat map is not generated.

4.2 Grad-Cam Heat Maps

The authors in [7] propose a method of generating heat maps to provide "visual explanations" for decisions of a large class of convolutional neural network (CNN) based models, making them more transparent.

Their approach, Gradient Weighted Class Activation Mapping (Grad-CAM), uses the gradients of any target concept, flowing into the final convolutional layer to produce an approximate

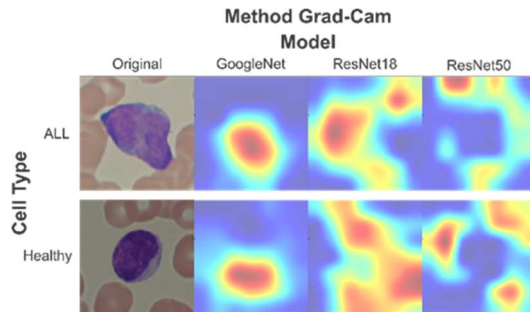


Fig. 5. The Figure shows results obtained with the Grad-Cam method and the GoogleNet, ResNet18, and ResNet50 models

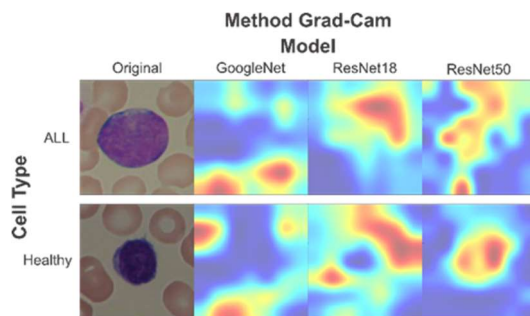


Fig. 6. The Figure shows results obtained with the Grad-Cam method and the GoogleNet, ResNet18, and ResNet50 models

Table 2. Table of heat map evaluation results. Retrieved from [19]

General Evaluation of the three models and cells types			
Method	Average	% Morphological Information	Num. Heat maps with the highest score
LRP	1.19	24%	4
Deep Taylor	1.69	34%	8
Input*Gradient	2.30	46%	25

location map that highlights the most essential pixels for the CNN network in class prediction.

Grad-CAM applies to a wide variety of CNN model families. These include CNNs with fully connected layers (e.g., VGG), CNNs used for structured outputs (e.g., subtitles), CNNs used in tasks with multimodal inputs (e.g., visual response to questions), or reinforcement learning without architectural changes or retraining.

In the context of image classification models, heat maps generated with this method provide

insight into the failure modes of these models (showing that seemingly unreasonable predictions have reasonable explanations), outperforming the approaches described in the previous section.

Finally, the authors designed and conducted human studies to measure whether Grad-CAM explanations help users establish adequate confidence in deep network predictions and showed that Grad-CAM enables untrained users to successfully discern a "stronger" deep network from a "weaker" one, even when both make identical predictions.

Figure 5 visualizes some heat maps obtained with the Grad-Cam method for the GoogleNet, ResNet18, and ResNet50 models, for a diseased and healthy cell image. This figure shows that the most relevant pixels with the GoogleNet model are mainly located within the same location as the cell of interest.

In the heat maps obtained with the ResNet18 model, some regions of interest are positioned within the concerned cell. However, most of these pixels are located outside the target cell.

Finally, in the heat maps that correspond to the ResNet50 model, the regions of interest usually are not related to the cell represented in the image. At first glance, it could be said that the Grad-Cam method, in combination with the GoogleNet model, is the best combination to generate heat maps that relate to the morphology of the target cell.

However, in Figure 6, we can see the results obtained with cells different from the previous figure. In this image, it is visualized that the heat map obtained is not always the best.

4.3 Evaluation by Experts

The authors in [11] present the results of evaluating the heat maps generated with the Deep Taylor, Input*Gradient, and LRP methods. The evaluation was performed by five pathologists with expertise in ALL cell classification.

From this work, it was obtained that the Input*Gradient method was the one that best visually related to the morphology of the target cell. The results are shown in Table 2. However, a database of segmented cells from the background was used in that work.

The paper presents feedback from one of the experts. It indicates that heat maps have low

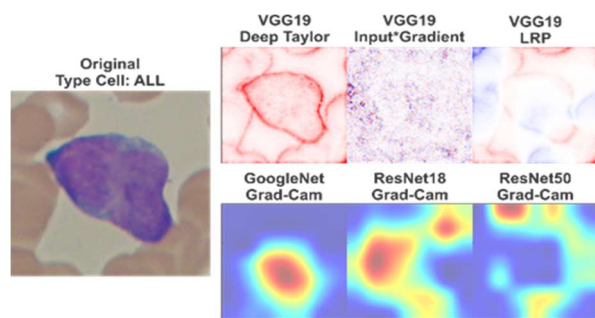


Fig. 7. Comparative image of the results obtained with the models and heat map generation methods evaluated

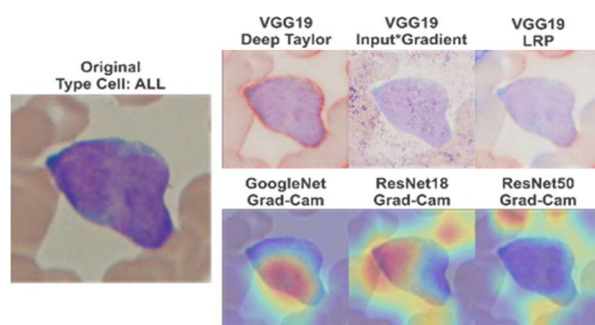


Fig. 8. Comparative image of the heat maps obtained superimposed on the ALL cell with which they were generated

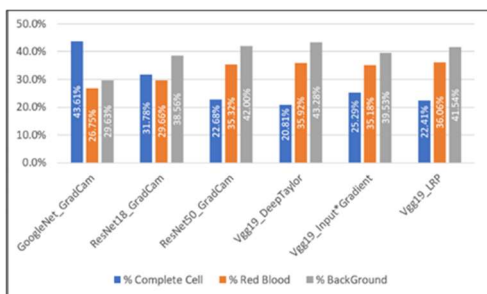


Fig. 9. Comparison between model and heat map generation method and their relationship with the central cell, red blood cells, and background, regarding percentages of relevant pixels

correlated information with the morphology of interest.

The present work is derived from the results presented in that paper. Here, we perform a computational evaluation of the heat maps using a semi-manual segmentation of the most significant morphological features, adding to the assessment of the Grad-Cam method.

We decided to use unsegmented images, i.e., the photos contain the cell of interest, the background, and other blood elements.

4.4 Heatmap Evaluation

Section 3.3 describes the methodological process used to evaluate the heat maps. The results obtained are shown here. In Fig. 7 shows the heat maps generated by the approaches and Fig. 8 shows an overlay of the heat map and the diseased cell with which they were produced as comparative images of the heat maps generated with the models and methods described above are displayed for an ALL-diseased cell.

If we know the number of relevant pixels placed within each segmented region, each category's percentage of pixels can be calculated based on a total of appropriate pixels.

The results of the evaluation performed in this work are shown in the graph in Fig. 9. From these results, it can be seen that the GoogleNet model and the Grad-Cam method is the combination of model and approach that best relates to the morphological characteristics of the cell of interest since 43.61% of the pixels marked as significant are located on the cell, 26.75% of the pixels are positioned in the red blood cells and the remaining 29.63% in the background.

In the second place, the ResNet18 model and the Grad-Cam method were set with 31.78% of the relevant pixels inside the interest cell, 29.66% in the red blood cells, and 38.56% in the background. The combination that obtained the worst result was the VGG19 model with the Deep Taylor method, with percentages of relevant pixels within the target cell of 20.81%, 35.92% in red blood cells, and the highest rate in the background at 43.28%.

We calculate the mean number of relevant pixels for each category. It is observed that the combination of the GoogleNet model and the GradCam architecture has a mean of 8071 pixels within the cell. In contrast, the red blood cell and background categories have the values of 4951 and 5484, respectively.

It is the only combination where the difference is more significant to a considerable extent between the complete cell versus the red blood cells and background categories. The results are shown in the graph of Fig. 10. The evaluation

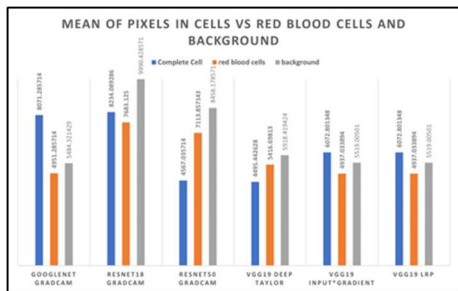


Fig. 10 Mean number of relevant pixels for each category using a combination of heatmap generation versus model

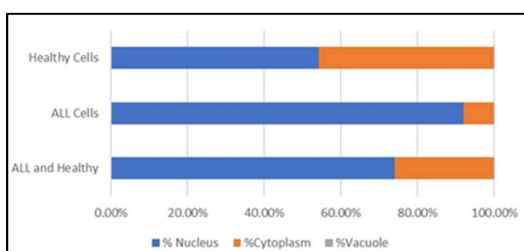


Fig. 11. Percentage of pixels with higher relevance located within the cell morphology of interest. Heat maps were created with the Google Net model and the Grad-Cam method

results in percentage terms of the best model-method combination are shown in Fig. 11.

The results show that the most relevant pixels for healthy cells are 54.26% within the nucleus and 45.74% in the cytoplasm. In the case of ALL cells, 91.90% are located within the nucleus, and 8.10% are located in the cytoplasm. The latter is related to the morphological characteristics that describe the disease because, in ALL cells, the cell's nucleus tends to have a larger area than the nucleus of a healthy cell.

In the database with which the heat maps were evaluated, there are three images where vacuoles were visualized. For this combination of the CNN model and heat map generation method, no pixels are of great relevance in their locations. Therefore, the vacuole category has 0% relevance.

5 Discussion

Performing a computational evaluation of the generated heat maps based on a map of

morphological features (nucleus, cytoplasm, vacuoles) for the classification of ALL cells, as well as red blood cells and background, allows to evaluate whether the CNN focuses on cell features of interest or other elements present in the image. In addition, it will enable the comparison of heat map generation methods to define which correlates with such morphological features.

According to the results obtained in this research, the GoogleNet model and the Grad-Cam method are the ones that best relate the natural morphological characteristics of the cell with the heat maps.

According to the results obtained in the present work, the evaluation made by expert pathologists in [11] can be corroborated.

6 Conclusion

Implementing heat maps in a neural network aims to identify the most critical regions or pixels for the neural network classification process. In this work, we generated heat maps with four different methods (Deep Taylor, Input*Gradient, LRP, and Grad-Cam) implemented on four different architectures (GoogleNet, ResNet18, ResNet50, VGG19).

The ALL-IDB2 database containing unsegmented images of the cell of interest with background and other blood elements present was used.

A ground truth map was generated and divided into three morphological features (nucleus, cytoplasm, and vacuoles), red blood cells, and background. Using the reference map, we evaluated the generated heat maps.

This evaluation concludes that the GoogleNet model focuses primarily on features present in the cell of interest. The Grad-Cam method is the heat map generation method that best expresses the relevance of CNNs. Combined with the GoogleNet model, it yields results that focus exclusively on the target cell.

7 Future Work

The generation of heat maps as a tool to explain the result of a prediction in an image is promising.

However, research is still required because the results of the heat maps should focus on showing the outcomes that hematologists expect. Most importantly, the construction of heat maps must include morphological features to be useful for medical specialists, so we will continue to explore the line of generating visual explanatory methods that focus exclusively on morphological features present in the cell of interest.

Acknowledgments

José de Jesús Velázquez-Arreola acknowledges CONAHCyT for the scholarship granted towards pursuing his PhD studies, CVU 778129.

References

1. **Hariprasath, S., Dharani, T., Shaikh-Mohammad, B. N. (2019).** Automated detection of acute lymphocytic leukemia using blast cell morphological features. 2nd International Conference on Advances in Science & Technology (ICAST).
2. **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>.
3. **Lamberti, W. F. (2022).** Classification of white blood cell leukemia with low number of interpretable and explainable features. DOI: 10.48550/arXiv.2201.11864.
4. **Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Herrera, F. (2020).** Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information fusion*, Vol. 58, pp. 82–115. DOI: 10.1016/j.inffus.2019.12.012.
5. **Weber, L., Lapuschkin, S., Binder, A., Samek, W. (2023).** Beyond explaining: opportunities and challenges of XAI-based model improvement. *Information Fusion*, Vol. 92, pp. 154–176, DOI: 10.1016/j.inffus.2022.11.013.
6. **Alber, M., Lapuschkin, S., Seegerer, P., Hägele, M., Schütt, K. T., Montavon, G., Kindermans, P. J. (2019).** iNNvestigate neural networks! *Journal of Machine Learning Research*, Vol. 20, pp. 1–8.
7. **Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D. (2017).** Grad-cam: Visual explanations from deep networks via gradient-based localization. *Proceedings of the IEEE international conference on computer vision*, pp. 618–626.
8. **Simonyan, K. (2015).** Very deep convolutional networks for large-scale image recognition. 3rd International Conference on Learning Representations (ICLR 2015), Computational and Biological Learning Society, pp. 1–14.
9. **He, K., Zhang, X., Ren, S., Sun, J. (2016).** Deep residual learning for image recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778.
10. **Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Rabinovich, A. (2015).** Going deeper with convolutions. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–9.
11. **Velázquez-Arreola, J., Zarraga-Vargas, O. A., Díaz-Hernández, R., Altamirano-Robles, L. (2023).** Evaluation of heatmaps as an explicative method for classifying acute lymphoblastic leukemia cells. *Proceedings 15th Mexican Conference, MCPR 2023*, pp. 252–260. DOI: 10.1007/978-3-031-33783-3_24.
12. **Jiwani, N., Gupta, K., Pau, G., Alibakhshikenari, M. (2023).** Pattern recognition of acute lymphoblastic leukemia (ALL) using computational deep learning. *IEEE Access*, Vol. 11, pp. 29541–29553. DOI: 10.1109/ACCESS.2023.3260065.
13. **Jiang, Z., Dong, Z., Wang, L., Jiang, W. (2021).** Method for diagnosis of acute lymphoblastic leukemia based on ViT-CNN ensemble model. *Computational Intelligence and Neuroscience*, Vol. 2021. DOI: 10.1155/2021/7529893.
14. **Abir, W. H., Uddin, M. F., Khanam, F. R., Tazin, T., Khan, M. M., Masud, M., Aljahdali,**

- S. (2022).** Explainable AI in diagnosing and anticipating leukemia using transfer learning method. *Computational Intelligence and Neuroscience*, Vol. 2022. DOI: 10.1155/2022/5140148.
- 15. Nayeon, Kim. (2021).** Deep learning technology-based model to identify benign and Pro-B acute lymphoblastic leukemia (ALL): Xception + LIME. *American Journal of Biomedical and Life Sciences*, Vol. 9, No. 5, pp. 279–285.
- 16. Ochoa-Montiel, R., Olague, G., Sossa, H. (2020).** Expert knowledge for the recognition of leukemic cells. *Applied Optics*, Vol. 59, No. 14, pp. 4448–4460. DOI: 10.1364/AO.385208.
- 17. Labati, R. D., Piuri, V., Scotti, F. (2011).** All-IDB: The acute lymphoblastic leukemia image database for image processing. 2011 18th IEEE International Conference on Image Processing, IEEE, pp. 2045–2048. DOI: 10.1109/ICIP.2011.6115881.
- 18. Maaliw, R. R., Alon, A. S., Lagman, A. C., Garcia, M. B., Susa, J. A. B., Reyes, R. C., Hernandez, A. A. (2022).** A multistage transfer learning approach for acute lymphoblastic leukemia classification. 2022 IEEE 13th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON) pp. 0488–0495. DOI: 10.1109/UEMCON54665.2022.9965679.
- 19. Mathworks.** Matlab Images Labeler application. <https://mathworks.com/help/vision/ref/imagelabeler-app.html>.
- 20. Mathworks.** Matlab augmented Images Datastore. <https://la.mathworks.com/help/deeplearning/ref/augmentedimagedatastore.html>.
- 21. Mamalakis, A., Barnes, E. A., Ebert-Uphoff, I. (2022).** Investigating the fidelity of explainable artificial intelligence methods for applications of convolutional neural networks in geoscience. *Artificial Intelligence for the Earth Systems*, Vol. 1, No. 4. DOI: 10.1175/AIES-D-22-0012.1.
- 22. Diaz-Resendiz, J. L., Ponomaryov, V., Reyes-Reyes, R., Sadovnychiy, S. (2023).** Explainable CAD system for classification of acute lymphoblastic leukemia based on a robust white blood cell segmentation. *Cancers*, Vol. 15, No. 13, pp. 3376. DOI: 10.3390/cancers15133376.

*Article received on 04/07/2023; accepted on 13/10/2023.
Corresponding author is Raquel Díaz-Hernández.