

# Deep Learning Technology-based Model to Identify Benign and Pro-B Acute Lymphoblastic Leukemia (ALL): Xception + LIME

**Nayeon Kim**

The Suwon University Lab, Suwon, Korea

**Email address:**

[k.jessica223@gmail.com](mailto:k.jessica223@gmail.com)

**To cite this article:**

Nayeon Kim. 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, 2021, pp. 279-285. doi: 10.11648/j.ajbls.20210905.19

**Received:** October 8, 2021; **Accepted:** October 27, 2021; **Published:** October 30, 2021

---

**Abstract:** Leukemia is a type of cancer that occurs when abnormal blood cells take place in the bone marrow. Leukemia can either be acute (fastly growing) or chronic (slowly growing) and it is considered as one of the most commonly diagnosed cancers for children younger than the age of 15 or adults older than the age of 55. Leukemia can be diagnosed through various types of tests and depending on the aggressiveness of the disease, the treatment may differ. To provide a low-cost, time-efficient solution, this study employs the deep learning technique to train the Xception, VGG16, VGG19, and MobileNet models to optimize the accuracy of medical image detection. Through medical imaging, the trained model is able to detect anomalies in the dataset and identify whether the given data is a benign acute lymphoblastic leukemia (ALL) or a Pro-B ALL. Overall, this VGG16 showed the most optimal performance in terms of accuracy and precision, producing a 98.5% accuracy in detecting abnormal regions from the dataset. This study also further used XAI technique and a deep convolutional neural network to visualize the results of anomalies. As a result, this paper concluded that both deep learning and machine learning techniques are yet to replace human resources and intelligence as the heatmap and the LIME portrayal identified different regions as abnormal parts, therefore proving the inconsistency of deep learning technology.

**Keywords:** Deep Learning, Machine Learning, Leukemia, XAI, Xception

---

## 1. Introduction

Leukemia is a cancer of blood cells, generally of the white blood cells (WBCs). White blood cells play a crucial role in the human body by protecting it from invasions by abnormal cells, viruses, fungi, bacteria, and any other foreign substances. A majority of the white blood cells are produced in the bone marrow and likewise, leukemia patients have abnormal cells growing in the bone marrow of bones. Common symptoms of leukemia range from simple fever and bone pain to seizures or loss of muscle control. Just like any other type of cancer, leukemia's effect on one's body heavily varies based on its aggressiveness. Leukemia can be either acute (rapid spread) or chronic (slow growth). Because of its various causes and risk factors such as genetic disorders or a family history of leukemia, leukemia is seen as one of the commonly diagnosed childhood cancers. With newly advanced discoveries of treatment, 90% of children

diagnosed with leukemia are likely to survive.

Although leukemia cannot be fully diagnosed through physical exams, patients most commonly discover and confirm their disease through imaging tests, blood tests, bone marrow biopsies, and aspiration. Once confirmed, patients go through one or more types of treatment: chemotherapy, radiation therapy, targeted therapy, biological or immunotherapy, or stem cell transplantation [1].

With the rapidly advancing technology, the medical industry is focused on experimenting with various technological tools and programs to produce a cost-, time-efficient method to diagnose and treat diseases. At the heart of this progression, deep learning and machine learning play a vital role. Deep learning technology has been used in the medical field for numerous purposes as chatbots and, most commonly, as medical imaging solutions to identify patterns of symptoms and specific types of diseases. Based on computational and mathematical algorithms, deep learning

technology can perform image segmentation and registration on the brain, lungs, tumor, and even on biological cells or membranes.

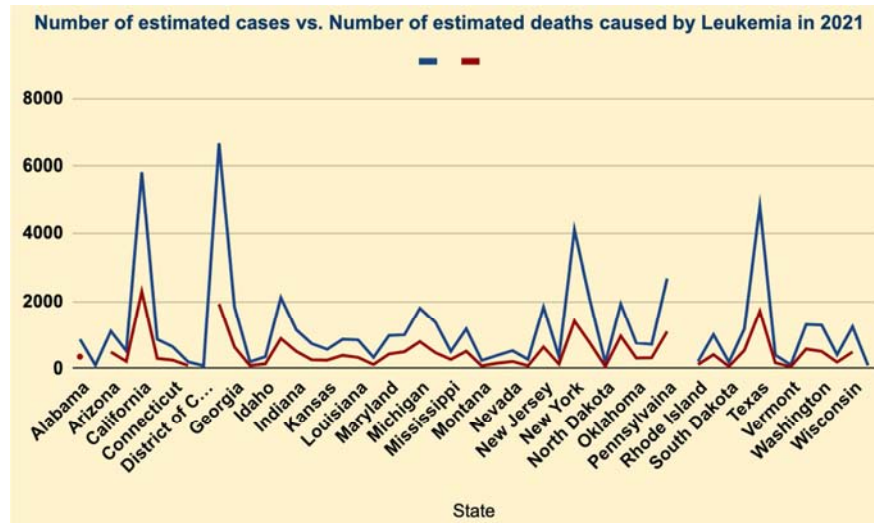


Figure 1. A comparison chart of estimated cases and estimated deaths of Leukemia in the United States in 2021 [2].

## 2. Objective

This paper aims to explain the use of deep learning technology in training the medical imaging technique to identify whether the given patient's image shows benign acute lymphoblastic leukemia (ALL) or Pro-B ALL. Just like any other type of cancer, the severity of the symptoms and treatments of leukemia may vary based on its aggressiveness, whether it's acute or chronic. A benign form of chronic leukemia does not require specific therapy while Pro-B ALL cannot be solely treated by intensive chemotherapy, and may require additional hematopoietic stem cell transplantation or targeted therapy [3]. As such, the acuteness of the disease may result in a completely different direction of treatment methods and outcomes; therefore, an accurate, efficient way of diagnosis is important. Through medical imaging techniques, this study trains the model to learn the distribution and patterns of normal (without disease) images to detect anomalies and identify whether the patient has benign or Pro-B ALL.

Our model was able to accurately distinguish the anomalies graphically, displaying them in red, while a majority of the studies we reviewed for training machine learning for leukemia lacked such results and clear reasoning behind what criteria or standards they used to detect anomalies. For instance, the hybrid inception v3 XGBoost model that S. Ramaneswaran and its team proposed included microscopic white blood cell images and an attention map for fine-tuned Inception v3 but failed to provide a clearly divided image of the cells that showed the readers how and why the model was able to detect leukemia from the medical imaging technique.

## 3. Literature Review

Amjad Rehman et al. propose a new segmentation

technique to accurately classify bone marrow images into four different categories: L1, L2, L3, or normal. The team initially processed through data collection obtained by the Amreek Clinical Laboratory Saidur Sharif Swat KP Pakistan, image acquisition, and segmentation to subdivide the region of interest before the classification phase. Using the Alexnet model with CNN, the team distinguished the data into its subtypes and normal condition through a deep learning technique known as transfer learning. As a result, the research work achieved an accuracy of 97.78%, successfully presenting a new method to gain higher classification accuracy compared to the existing automated methods [4].

S. Ramaneswaran et al. team proposes a hybrid Inception v3 XGBoost model that classifies acute lymphoblastic leukemia through identifying microscopic white blood cell images. The research team used a dataset called ISBI C-NMC 2019 dataset prepared at Laboratory Oncology, AIIMS, New Delhi, which consisted of a total of 10661 cell images. In this research, the team used XGBoost classifier as a classification head instead of traditionally utilizing the softmax classifier considering the effectiveness of using an XGBoost classification head with a CNN model. Through a total of two stages of training, the team experimented with various CNN backbone feature extractors and extracted image features from the model trained in the first stage. Using the PyTorch library to develop the deep learning models and the Adam optimizer, the team successfully achieved a weighted F1 score of 0.986 to accurately detect acute lymphoblastic leukemia cells [5].

Mustafa Ghaderzadeh and his team present a comprehensive review of published machine-learning-based leukemia detection and classification models that process peripheral blood smear (PBS) image analysis. The research team performed a systematic search in four databases—PubMed, Scopus, Web of Science, and ScienceDirect—and initially collected a total of 116 articles which they narrowed

down to 16 articles after reviewing the articles based on their inclusion and exclusion criteria. Through a thorough analysis of published studies, the team discovered that machine learning and deep learning algorithms were frequently used as machine vision techniques. Moreover, the team realized that the more prominent techniques of blood smear image segmentation are thresholding methods and object detection. And among these segmentation algorithms, the team presented three major computational core types which are pixel-based image segmentation, region-based segmentation, and shape-based segmentation. Overall, this study announced that the use of machine learning in PBS image analysis had an average accuracy of over 97%, proving machine-learning as a more accurate, more efficient, cheaper, and safer diagnostic method [6].

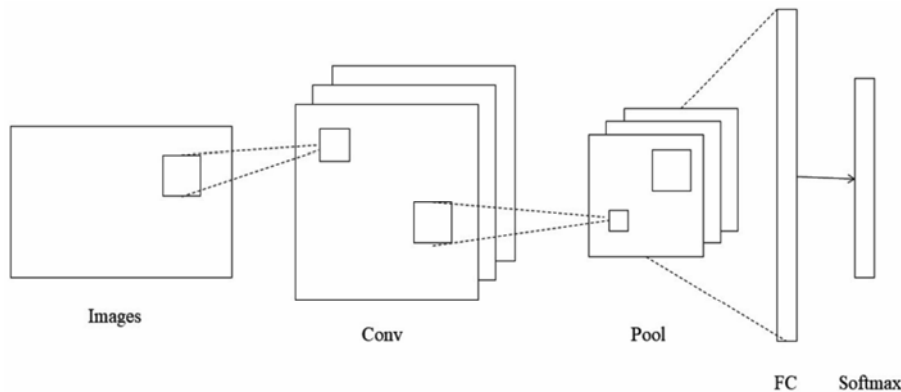
Mohamed Loey, Mukdad Naman, and Hala Zayed propose two automated classification models to detect leukemia by differentiating blood microscopic images. The research team used a dataset of 564 blood microscopic images consisting of half leukemia-free samples and half leukemia-affected samples. The study employed two classification models that both adopted transfer learning; the first model employed a pre-trained CNN called AlexNet to extract distinct features and classifiers in which the team discovered that the SVM classifier showed the highest accuracy. In this model, the team went through image pre-processing, feature extraction, and classification by starting with blood image conversion into a red-green-blue (RGB model) and resizing images to a fixed size. For the second model, the team used AlexNet for both feature extraction and classification which proved to be a superior method compared to the first model based on

different performance metrics [7].

Maneela Shaheen and its team propose a new AlexNet-based classification model to identify Acute Myeloid Leukemia (AML) in microscopic blood images. Under the four criteria—precision, recall, accuracy, and quadratic loss—the research team also compared the quality of performance between AlexNet and LeNet-5 based models. The study used a pre-classified dataset of 4000 images obtained by a tertiary care hospital in Peshawar, Pakistan along with other microscopic peripheral blood images found in public. Through training and comparison rounds, the team discovered that the AlexNet-based model showed 98.58% accuracy with 87.4% precision while the LeNet-5 based model showed 96.25% accuracy with 83.6% precision. Overall, the team concluded that the AlexNet-based model can analyze and detect vital features from medical images to accurately identify diseases [8].

## 4. Data Description

This study used a dataset of 3256 peripheral blood smear (PBS) images from 89 patients suspected of acute lymphoblastic leukemia (ALL). This dataset was provided by the bone marrow laboratory of Taleqni Hospital in Tehran, Iran. The given samples were either benign and malignant, where the latter had three subtypes of malignant lymphoblasts: Early Pre-B, Pre-B, and Pro-B acute lymphoblastic leukemia. In order to optimize the accuracy of this study, we only used the dataset of 504 images of benign lymphoblastic leukemia and 804 images of Pro-B acute lymphoblastic leukemia [9].



*Figure 2. Overall architecture of the convolutional neural network.*

## 5. CNN

Convolutional neural network (CNN) belongs to the deep learning models and is commonly used for image detection and classification. Unlike deep neural networks, it basically consists of different layers, including convolutional layer, pooling layer, and fully connected layer. As the goal of this algorithm is to manipulate the image-based data, the model fundamentally uses the jpeg or png format of the input data. When the model gets input data, the convolution layer

utilizes the filter to extract the features through the convolutional operation. The filter rotates the image, extracts the feature through convolutional operation, and convolves a filter map. After the convolution layer, the pooling layer is used to downsize the feature map. Max pooling and average pooling are the representative ones, and they extract the maximum and average values per each from the filter map. The fully connected layer is also known as the DNN, which is applied to classify the dataset into the given label [10]. The flatten layer must be preceded in order to insert into the fully connected layer because of a dimensional size [11].

## 6. XAI

Machine learning and deep learning models are “black box” models, which can not provide the reason for the result. Even though they yield high accuracy compared to other algorithms, this “black box” makes them difficult to be applied in various fields. The explainable artificial intelligence (XAI) aims to explain and provide a conclusive reason from the given result. Local interpretable Model-Agnostic Explanation (LIME) yields the explanation for a single data point and can be applied to any machine learning and deep learning algorithms.

The main idea of the LIME is that if the output of the model changes significantly when the input value is slightly changed, then it can be defined that the variable is an important one [12]. For the image dataset, LIME firstly splits the given dataset into the interpretable elements called superpixels. Then, a slight change is applied to the superpixels, by covering them in gray. Lastly, the gray image

is used as input data for the model and gets the result [13].

## 7. Xception

As the name suggests, the Xception model is based on Inception. A typical convolution operation simultaneously obtains correlation between channels and spatial correlation in one kernel, but Inception separates it to some extent. The Xception model further performs a depthwise separable convolution operation to completely separate correlations between channels and spatial correlations. Depthwise separable convolution operation, which is an extreme version of inception consists of pointwise convolution and depthwise convolution. Correlation between channels is calculated through a  $1 \times 1$  convolution operation in pointwise convolution and a  $3 \times 3$  convolution operation in depthwise convolution, while depthwise separable convolution operation is calculated in a reverse way [14].

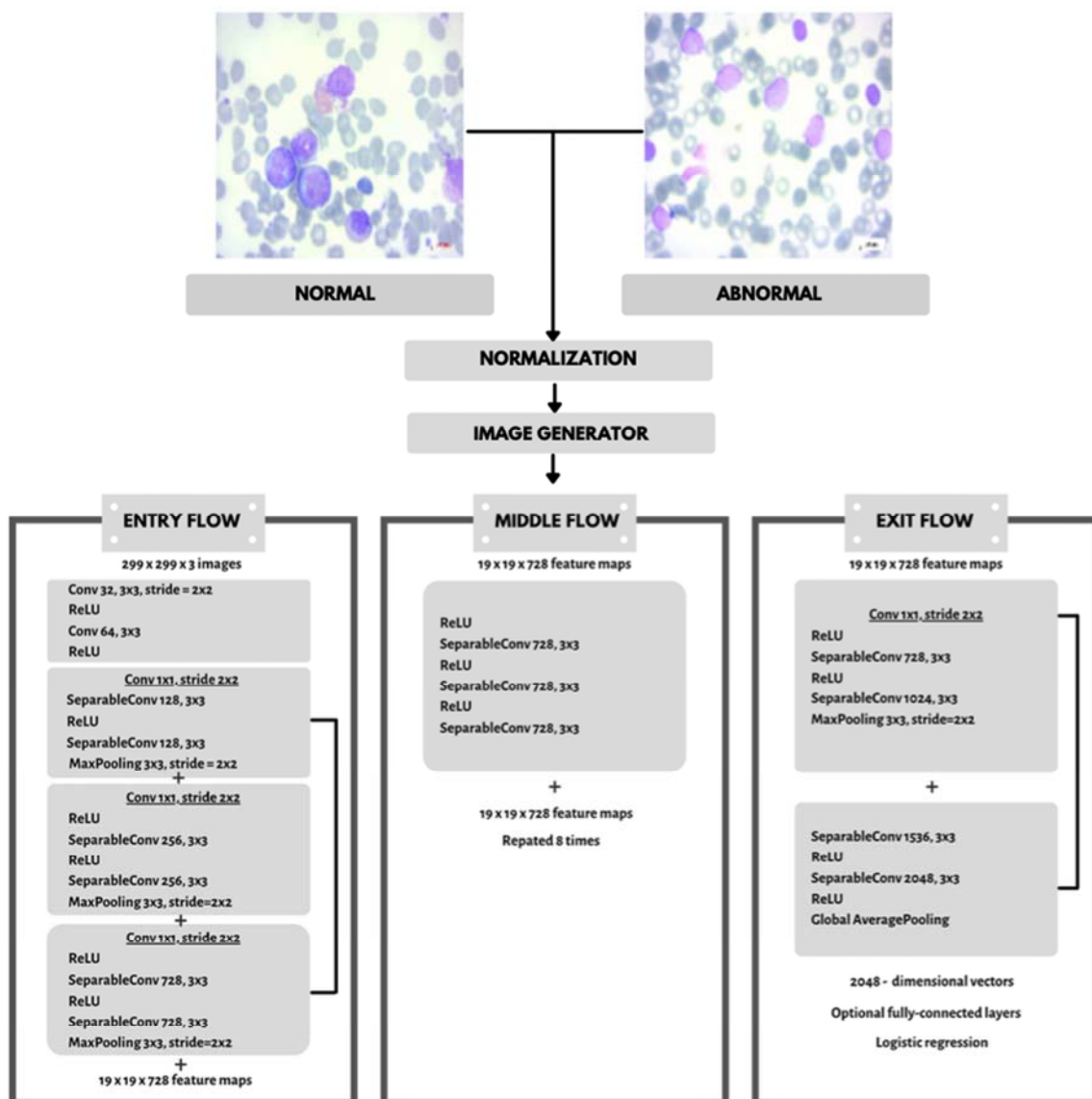


Figure 3. Proposed model flowchart.

## 8. Proposed Model

We initially found a dataset that consisted of both normal and abnormal images for our model input. For image normalization, we divided all of the dataset images by 255 pixels and, in order to enlarge the number of images, we used the image generator function. During this step, we adjusted each image by making small alterations such as rotating an image by 20 degrees or fixing its brightness. Then, we applied the Xception model and used binary classification – benign and Pro-B ALL. To visualize the results we included a heatmap that graphically showed the anomaly parts and used

“Lime” as a final step to distinguish the normal and abnormal regions by color.

## 9. Results

For this study, we applied deep learning models called Keras Applications which are commonly used for prediction, fine-tuning, and feature extraction. Among the various Keras Application models, we used Xception, VGG16, VGG19, and MobileNet.

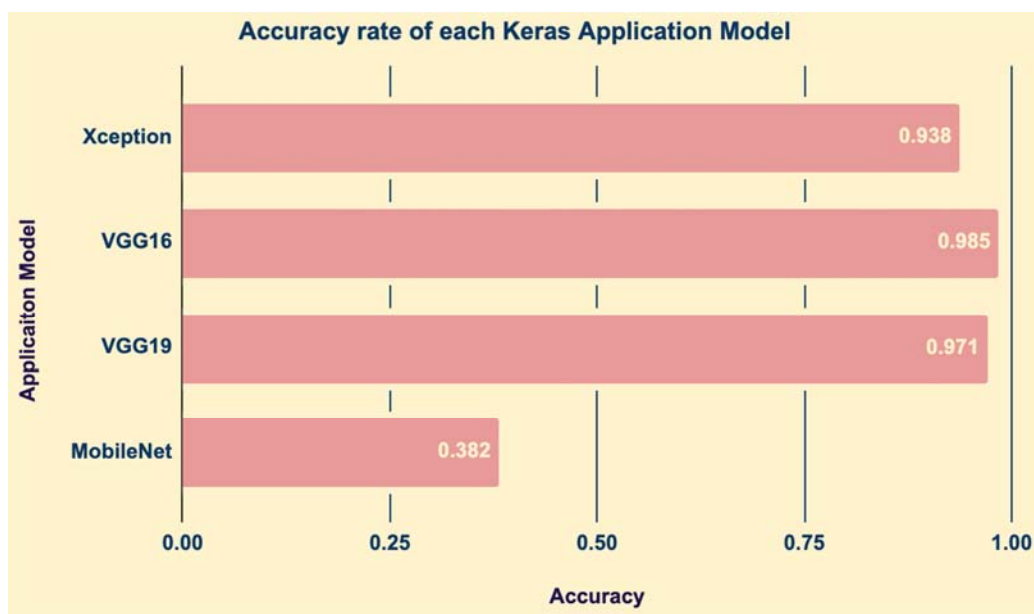


Figure 4. A comparison chart of accuracy rate of four Keras Application Models.

As shown above, the VGG16 model showed the highest accuracy rate of 98.5%, then VGG19 of 97.1%, Xception of 93.8%, and MobileNet of 38.2% respectively. Compared to

other convolutional neural networks (CNN), MobileNet is known to be a class of small, low power models, which explains the accuracy degradation [15].

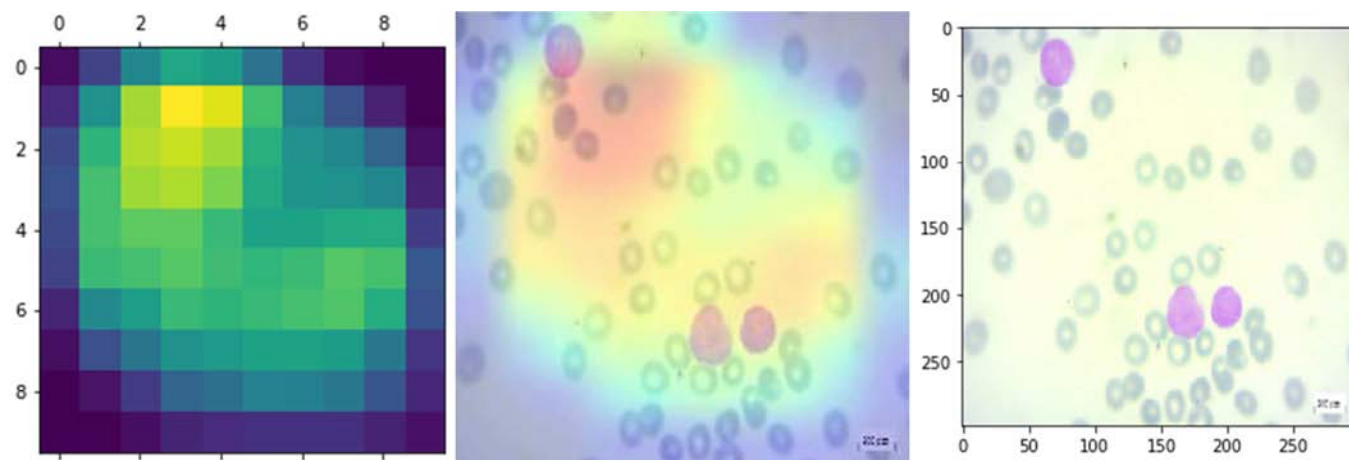
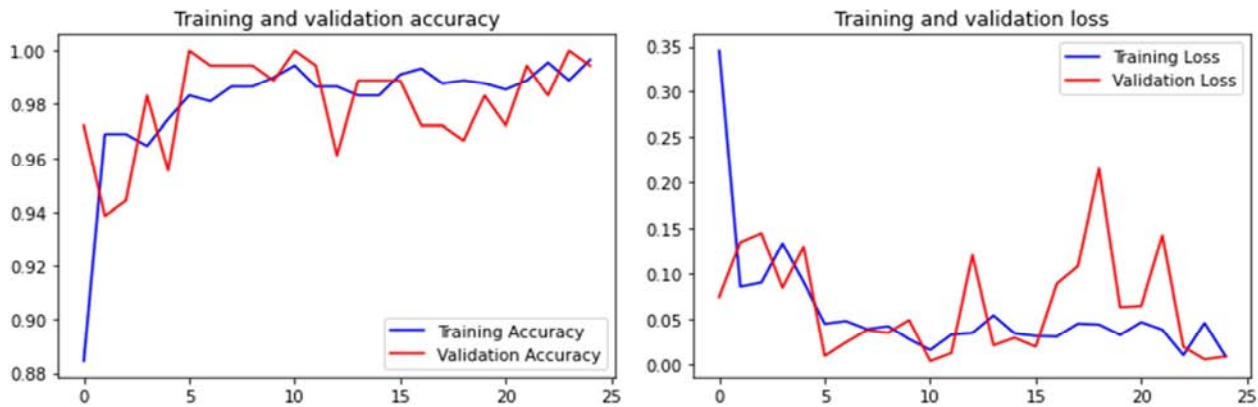


Figure 5. Heatmap of anomalies.

Moreover, as shown in figure 5, we used a deep convolutional neural network to develop a tumor region

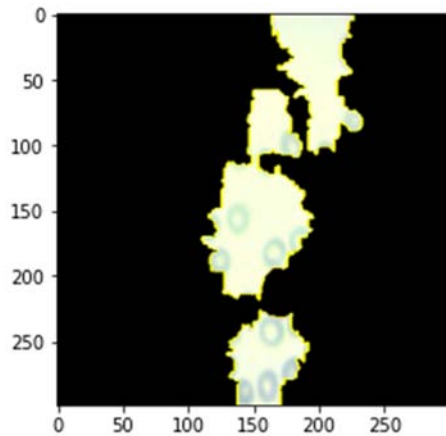
recognition system for the datasets. Through the heatmap, we visualized the anomaly parts of the data images, portraying

normal regions in purple and abnormal parts in red [16].



**Figure 6.** Training and validation accuracy and loss graphs.

For this study, we set the epoch value to 25. The training and validation graphs in figure 6 display the process of how this model was trained during the 25 cycles.



**Figure 7.** Lime portrayal.

Lastly, we applied one of the explainable AI (XAI) techniques called “Lime” to portray anomaly parts of the image. “Lime” stands for Local Interpretable Model-agnostic Explanations and it supports image classifiers in this study. As seen in figure 7, the black parts display the normal region while the colored region represents the anomalies.

## 10. Principal Finding

As shown in figures 5 and 7, the model does not identify the same region as an anomaly, presenting the inconsistencies of deep learning technologies. Although machine learning as a whole has been trained and targeted to diagnose and identify diseases, it is still too early to decisively say that AI can safely replace human diagnosis in terms of accuracy. Deep learning techniques are certainly cost and time-efficient solutions but the inconsistent accuracy and results of this study prove that machine learning still has abundant space for improvement to be settled as a primary medical replacement.

## 11. Limitation

The original dataset consisted of four types of leukemia: benign, early Pre-B, Pre-B, and Pro-B acute lymphoblastic leukemia. However, when we trained the model with four different Keras Applications, the maximum accuracy only roughly achieved 83%. Therefore, we had to perform the binary classification with only benign and Pro-B acute lymphoblastic datasets, and as a result, the overall accuracy of this study increased. Moreover, although we set Xception as our main Keras Application model, VGG16 and VGG19 showed higher accuracy rates.

## 12. Conclusion

Leukemia is a cancer of white blood cells that may either be chronic or acute. Because the severity of the disease may vary drastically, this study proposed to use deep learning technology to accurately detect whether the image contained a sign of benign acute lymphoblastic leukemia (ALL) or a Pro-B ALL. This paper took a step-by-step procedure from collecting data input, going through an image normalization, using an image generator function, to applying four Keras Application models – Xception, VGG16, VGG19, and MobileNet– to optimize the accuracy of medical imaging techniques. With a 98.5% accuracy, the VGG16 model showed a superior performance to our predicted model, Xception. One of the challenges we encountered was that when we tried to use all four of the classifications for our dataset – benign, early, pre, and pro – the maximum accuracy only reached about 83%, a comparably low rate to the 98.5% accuracy we achieved from a binary classification. To visualize the anomaly parts, this study further included a heatmap using a deep convolutional neural network and an XAI technique called “Lime” to graphically display the abnormal regions. Comparing the heatmap to the “Lime” graphics, we recognized that the detected anomaly parts were not identical, therefore proving the inconsistencies of deep learning technology. This weakness showed that machine

learning and AI are yet to replace human resources and intelligence. Therefore, for the future research, more sophisticated XAI technologies should be applied to the medical datasets so that doctors or researchers could utilize the AI techniques in the diagnosis of diverse diseases.

## References

- [1] Watson, S. (2019, May 11). Symptoms of Leukemia in Pictures: Rashes and Bruises. Healthline. <https://www.healthline.com/health/pictures-leukemia-rashes-bruises#symptoms>.
- [2] American Cancer Society: Cancer Facts & Statistics. (2021, October 8). Leukemia. <https://cancerstatisticscenter.cancer.org/#!/cancer-site/Leukemia>.
- [3] Gao, C. (2019, November 14). Clinical-biological characteristics and treatment outcomes of pediatric pro-B ALL patients enrolled in BCH-2003 and CCLG-2008 protocol: a study of 121 Chinese children. Cancer Cell International. <https://cancerbiomedcentral.com/articles/10.1186/s12935-019-1013-9>.
- [4] Rehman, A., Abbas, N., Saba, T., Rahman, S. I. U., Mehmood, Z., & Kolivand, H. (2018). Classification of acute lymphoblastic leukemia using deep learning. *Microscopy Research and Technique*, 81 (11), 1310–1317. <https://doi.org/10.1002/jemt.23139>.
- [5] Ramaneswaran, S., Srinivasan, K., Vincent, P. M. D. R., & Chang, C. Y. (2021). Hybrid Inception v3 XGBoost Model for Acute Lymphoblastic Leukemia Classification. *Computational and Mathematical Methods in Medicine*, 2021, 1–10. <https://doi.org/10.1155/2021/2577375>.
- [6] Ghaderzadeh, M., Asadi, F., Hosseini, A., Bashash, D., Abolghasemi, H., & Roshanpour, A. (2021). Machine Learning in Detection and Classification of Leukemia Using Smear Blood Images: A Systematic Review. *Scientific Programming*, 2021, 1–14. <https://doi.org/10.1155/2021/9933481>.
- [7] Loey, M., Naman, M., & Zayed, H. (2020). Deep Transfer Learning in Diagnosing Leukemia in Blood Cells. *Computers*, 9 (2), 29. <https://doi.org/10.3390/computers9020029>.
- [8] Shaheen, M., Khan, R., Biswal, R. R., Ullah, M., Khan, A., Uddin, M. I., Zareei, M., & Waheed, A. (2021). Acute Myeloid Leukemia (AML) Detection Using AlexNet Model. *Complexity*, 2021, 1–8. <https://doi.org/10.1155/2021/6658192>.
- [9] *Acute Lymphoblastic Leukemia (ALL) image dataset*. (2021, April 30). Kaggle. <https://www.kaggle.com/mehradaria/leukemia>.
- [10] Sainath, T. N., Vinyals, O., Senior, A., & Sak, H. (2015). Convolutional, Long Short-Term Memory, fully connected Deep Neural Networks. *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Published. <https://doi.org/10.1109/icassp.2015.7178838>.
- [11] Albawi, S., Mohammed, T. A., & Al-Zawi, S. (2017). Understanding of a convolutional neural network. *2017 International Conference on Engineering and Technology (ICET)*. Published. <https://doi.org/10.1109/icengtechnol.2017.8308186>.
- [12] Sahay, S., Omare, N., & Shukla, K. K. (2021). An Approach to identify Captioning Keywords in an Image using LIME. *2021 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS)*. Published. <https://doi.org/10.1109/icccis51004.2021.9397159>.
- [13] Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). “Why Should I Trust You?” *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Published. <https://doi.org/10.1145/2939672.2939778>.
- [14] Chollet, F. (2017). Xception: Deep Learning with Depthwise Separable Convolutions. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Published. <https://doi.org/10.1109/cvpr.2017.195>.
- [15] Yun, S., & Wong, A. (2021). Do All MobileNets Quantize Poorly? Gaining Insights into the Effect of Quantization on Depthwise Separable Convolutional Networks Through the Eyes of Multi-scale Distributional Dynamics. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 2447–2456).
- [16] He, T., Guo, J., Chen, N., Xu, X., Wang, Z., Fu, K., Liu, L., & Yi, Z. (2020). MediMLP: Using Grad-CAM to Extract Crucial Variables for Lung Cancer Postoperative Complication Prediction. *IEEE Journal of Biomedical and Health Informatics*, 24 (6), 1762–1771. <https://doi.org/10.1109/jbhi.2019.2949601>.