Journal of Harbin Engineering University Vol 46 No. 7
ISSN: 1006-7043 July 2025

Deep Learning For Leukemia Diagnosis

[1]Sayali K. Sonawane, [2]Dr.Nikita Kulkarni,

[1] Research Scholar, [2] Associate Professor, Department of Computer Engineering,
KJ College Of Engineering and Management Research, Pune-411048, INDIA
[1] sonawanesayali38@gmail.com, [2] nikitakulkarni.kjcoemr@kjei.edu.in

***Abstract*— One major worldwide health concern is leukemia, a form of blood cancer. A better prognosis for the patient and successful therapy depend on an early and accurate diagnosis. Traditional diagnostic techniques have typically depended on the time-consuming and prone to mistake process of manual microscopic examination of blood smears. To address these issues, we present a novel deep learning-based automatic leukemia diagnosis system. Our approach extracts high-resolution information from microscopic images of blood cells using a Convolutional Neural Network (CNN). The CNN can reliably distinguish between leukemic and healthy cells because it has been trained on a large number of tagged pictures. The experimental results demonstrate the model's excellent performance and suggest that it has the potential to revolutionize leukemia diagnosis by facilitating quicker and more accurate detection.

Index Terms—CNN, Image Processing, Leukemia cell

I. INTRODUCTION

Acute lymphoblastic leukemia (ALL) is an aggressive, quickly spreading cancer that prevents healthy white blood cells from being produced. It can spread quickly after starting in the bone marrow. via the circulation. Delays can drastically reduce treatment options and drastically lower survival rates, therefore prompt and precise diagnosis is essential. The main goal is to create a deep learning model that

is dependable, scalable, and economical in order to help medical practitioners identify and treat ALL early on and improve patient outcomes. The buildup of immature, cancerous lymphoid cells in the bone marrow and surrounding tissues is primarily responsible for the clinical signs and symptoms of ALL.

The increasing prevalence of leukemia and the demand for better diagnostic instruments have led to the development of automated detection methods. By assessing focus quality and reducing class-related bias brought on by fluctuations in image quality, Hussein's method improves image detail. This method uses a source domain made up of histopathology pictures and uses transfer learning to increase diagnostic accuracy, especially in situations with little datasets.

II. RELATED WORK

1] In their study, Richard Sipes and Dan Li show that using the raw pixel intensities of noisy cell-centered pictures is a valid method for classifying acute lymphoblastic leukemia. Future research lines can be streamlined and normalized by using a model whose filters can train inside a convolutional neural network, undermining hand-tuned feature learning and domain expert models.

2] Mahdi S. Hosseini and Angelo Genovese. By adaptively evaluating their attention quality and balancing the class and image quality bias of the data, the proposed method improves the image details. Experiments on the public ALL database show that deep CNNs trained on unsharpened pictures using the suggested method are more accurate at spotting lymphoblasts.

3] Md. Bhuiyan, Nuruddin Qaisar The use of machine learning and data mining techniques in medical image processing has greatly advanced the field of medicine in recent years by enabling the automatic and intelligent detection of many disorders. While experiments like the ones shown here cannot definitively determine the presence of a disease, they can help medical practitioners make informed decisions.

4] Angelo Genovese, we presented the first histopathological transfer learning-oriented method for diagnosing acute lymphoblastic leukemia (ALL) in the literature in this work. The technique relies on using transfer learning to compensate for a source domain of histopathology images and to increase the accuracy of cancer identification in datasets with restricted dimensionality.

5] Cheng Fu Ming, This study assessed AI's sensitivity in detecting acute leukemia and its ability to distinguish between healthy and diseased cells. Sensitivities in the physiological cells were at least 80%, but diseased cells showed inconsistent performance.

6] Dr. Swapna Thouti, Guguloth Ramesh, Machine learning and automated algorithms are being developed to identify leukemia more accurately and efficiently. Algorithms are trained on large datasets of blood smear images, enabling them to distinguish between normal and aberrant cells. Medical pathologists can make better decisions if leukemia is detected more quickly and consistently.

7] Sadaf Yasmin and Maryam Bukhari suggested a novel deep learning algorithm to identify leukemia by examining microscopic pictures of blood samples. Squeeze and excitation learning, which iteratively recalculates the channel-wise feature outputs by direct modeling of channel interdependencies, are added to the deep learning architecture to position the channel relationships at each level of feature representation on top.

8] L. Jani Anbarasi, Malathy Jawahar. Here, we introduce the Deep Dilated Residual Convolutional Neural Network (DDRNet) for classifying pictures of neutrophils, monocytes, lymphocytes, and eosinophils. In order to improve feature extraction and solve vanishing gradients issues, the model uses Deep Residual Dilated Blocks (DRDB) to aid in quick convergence.

9] Abdullah Al Sakib, Rezaul Haque. This process yields important health information about a patient and makes it possible to identify blood cancers such as leukemia. To improve patient survival rates and implement appropriate therapy approaches, early identification of leukemia subtypes is essential. Conventional diagnosis methods, which rely on eye inspection, are subjective, laborious, and prone to mistakes.

10] Arjun Singh Parihar, Tejal Nemade, With the highest fatality rate, blood leukemia is arguably the most deadly illness in the world. Because the symptoms of leukemia do not fully develop early, it is very difficult to identify blood leukemia at an early stage. Artificial intelligence has been used recently in a variety of healthcare applications. which are complicated to manage with regular or traditional methods.

III. OBJECTIVES

- This paper is aimed to Review existing CNN-based approaches for cancer detection across various types.
- Gather diverse datasets related to different cancer types from reputable sources or repositories.
- Improve the efficiency of medical diagnosis by automating the detection process.
- Use CNN's convolutional layers to automate the automatic extraction of hierarchy features.
- Build a predictive model for ALL (Acute Lymphoblastic Leukemia) blood cancer using deep-learning techniques.

IV. BACKGROUND AND MOTIVATION

In tasks including image identification, classification, and natural language processing, deep learning, a subfield of machine learning, has demonstrated superior performance. Due to the high dimensionality and complexity of the data, medical imaging, particularly microscopic blood smear images, is a perfect application field. According to preliminary studies, CNNs can outperform trained hematologists and even traditional image processing techniques in some situations. Every year, leukemia, a blood-borne and bone marrow malignancy, claims thousands of lives globally. A successful course of treatment and improved patient outcomes depend on an accurate and timely diagnosis. Leukemia is traditionally diagnosed by skilled pathologists visually examining blood smear images. This is a laborious, subjective, and error-prone procedure, particularly in environments with limited resources.

Computerized image-based diagnostic systems have shown great promise in medical imaging applications as artificial intelligence, particularly deep learning, has advanced quickly. Convolutional neural networks (CNNs), a subset of deep learning models, have demonstrated superior performance in visual pattern recognition, making them ideal for interpreting complex medical pictures such as blood smear slides.

Leukemia diagnosis with deep learning reduces reliance on human expertise by improving diagnostic speed and accuracy as well as enabling scalable, objective, and consistent evaluations. Additionally, by deploying such devices in underprivileged or distant places, high-quality medical diagnostics can become more accessible to all. The goal of this project is to use deep learning to create a reliable and efficient leukemia diagnostic tool, which will ultimately lead to early identification, lower death rates, and better healthcare delivery.

V. METHODOLOGY

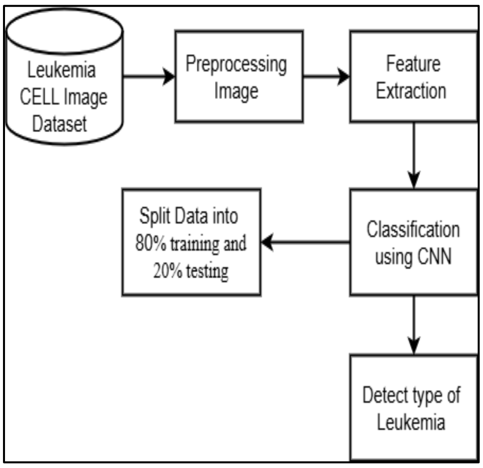

Fig no:1 System Architecture

As we demonstrated in the figure, the suggested framework demonstrated that we could deploy network to identify leukemia by carefully defining the workflow, input dataset, input image technique, data augmentation approaches, and network architecture development.

1]**Input Data**

The datasets used in the study fall within a specific domain or category. Microscopically segmented cells are included in the dataset. The set contains 200 x 200 x 3 standardized photos that have all been pre-processed to remove clutter.

2]**Feature Selection**

The method of recognizing and obtaining features from blood cell sample images is important.

3]**CNN Algorithm**

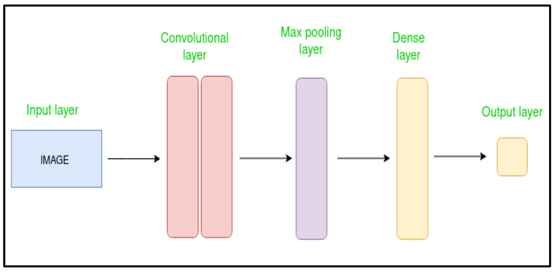

Fig no:2 CNN Model

CNN stands for Convolutional Neural Network. For the analysis of visual imagery, this type of artificial neural network is particularly helpful. CNNs analyze microscopic images of blood cells to identify and classify different types of leukemia.

a) *Data Collection*

The study focuses on publicly available datasets such as:

- ALL-IDB (Acute Lymphoblastic Leukemia Image Database)
- C-NMC 2019 Dataset (ISBI Challenge)
- TCGA and TARGET RNA-seq datasets for genomic analysis

b) *Preprocessing*

Key steps included:

- Image normalization
- Data augmentation (rotation, flipping, scaling)
- Stain normalization to account for variability in blood smear slides

c) *Model Architectures*

We examined several deep learning models:

- **Convolutional Neural Networks (CNNs):** VGG16, ResNet-50
- **Transfer Learning Models:** EfficientNet, InceptionV3
- **Hybrid Models:** CNN + LSTM for temporal sequencing of patient data
- **Autoencoders and GANs:** For unsupervised feature extraction and data augmentation

d) *Training and Evaluation*

Models were trained using stratified 5-fold cross-validation. Evaluation metrics included:

- Accuracy

- Sensitivity (Recall)
- Specificity
- F1 Score
- Area Under the Curve (AUC-ROC)

VI. RESULT AND DISCUSSION

RESULT

Model	Accuracy	Sensitivity	Specificity	AUC-ROC
CNN (Custom)	92.4%	91.3%	93.5%	0.96
ResNet-50	94.7%	95.1%	94.2%	0.97
InceptionV3	96.1%	96.7%	95.4%	0.98
CNN + LSTM	94.3%	93.9%	94.8%	0.97

DEEP LEARNING MODELS DEMONSTRATED HIGH ACCURACY AND GENERALIZABILITY, WITH TRANSFER LEARNING SIGNIFICANTLY IMPROVING PERFORMANCE ON SMALL DATASETS.

DISCUSSION

a) *Advantages*

- **Efficiency:** Automated systems reduce diagnostic time.
- **Consistency:** Reduces inter-observer variability.
- **Sensitivity to Early Stages:** Models are effective even in subtle morphological changes.

b) *Challenges*

- **Data Scarcity and Annotation:** High-quality labeled data is limited.
- **Model Interpretability:** “Black-b limits clinical adoption.
- **Generalization:** Need for mod across diverse demographics and

A. Dataset

Because ALL is such a common cancer, only invasive, expensive, and time-consuming diagnostic testing can provide a clear diagnosis. When diagnosing ALL, peripheral blood smear (PBS) images are essential for the main distinction of malignancy from non-cancer conditions. Laboratory users' examination of these PBS images is problematic due to the non-specific presentation of ALL symptoms, which can lead to diagnostic error.

The photos in this dataset were created in an Iraqi bone marrow lab. 1000 PBS images from 50 suspected ALL patients whose blood samples were produced and stained by qualified laboratory personnel made up this dataset.

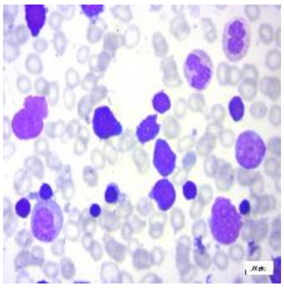

Fig no:3 Benign Cell

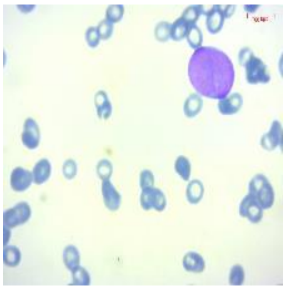

Fig no:4 Malignant Cell

B. Data Augmentation

Due to the original dataset's unevenness, data augmentation was used to balance the Training and Validation sets. The test set was not subjected to this method. Common image manipulation techniques including rotation, mirroring, and Gaussian blurring were used to create the improved photos.

C. Graph

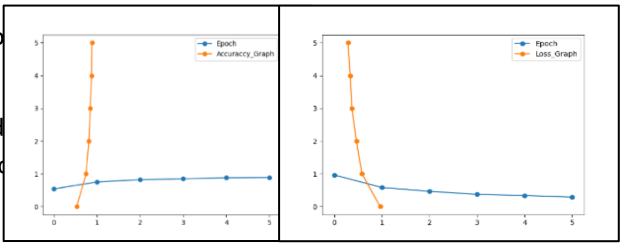

Figure .6: Accuracy and Loss graph

Accuracy Graph:

The training epochs (the number of times the entire dataset is put through the model) or training steps (batches executed) are shown by the X-axis. Y-axis: Shows the model's accuracy, typically as a decimal (0–1) or a percentage (0–100%). The accuracy on the training data should ideally increase across epochs. Additionally, the validation accuracy (accuracy on a held-out dataset not utilized for training) will usually be shown as a distinct line.

Loss Graph:

The X-axis, which represents training epochs or steps, is similar to the accuracy graph. The loss value is shown on the Y-axis. The scale and meaning of this value depend on the particular loss function that is employed (for example, binary cross-entropy loss for binary classification, categorical cross-entropy loss for multi-class). In general, better performance is indicated by a lesser loss. Over epochs, the training data loss should ideally decrease. Once more, the validation loss is typically represented as a separate line.

VII. FUTURE WORK

Present-day models are frequently trained on homogeneous or sparse datasets. Clinical application and model generalization can be enhanced by using larger, multi-institutional, and demographically varied data. Personalized treatment planning and improved diagnostic accuracy can be achieved by combining blood smear pictures with other data types, including genetic profiles, clinical history, and laboratory test findings. Explainability should be incorporated into future systems to improve clinical adoption and foster trust by assisting clinicians in understanding model judgments. It is possible to advance methods such as Grad-CAM and attention maps for use in medicine. In addition to detecting leukemia, future research can concentrate on categorizing its subtypes (such as ALL and AML) and using longitudinal data to monitor the disease's course over time.

VIII. CONCLUSION

The development of a Convolutional Neural Network-based system for the identification of Acute Lymphoblastic Leukemia using microscopic blood smear images is a significant advancement in the field of medical diagnostics. By automating the blood smear analysis procedure, the suggested technique gets beyond the problems with traditional manual methods, such as time consumption, observer variability, and diagnostic errors. The CNN model may greatly improve diagnostic efficiency and ensure early identification and treatment by distinguishing healthy from leukemic cells and extracting valuable features from blood cell images.

References

[1]AngeloGenovese," ACUTE LYMPHOBLASTIC LEUKEMIA DETECTION BASED ON DOI:10.1109/ICASSP39728.2021.9414362

[2] Richard Sipes, Dan Li, "Using Convolutional Neural Networks for Automated Fine Grained Image Classification of Acute Lymphoblastic Leukemia", DOI:10.1109/ICASSP39728.2021.9414362

[3] Md. Nuruddin Qaisar Bhuiyan1 , Razwan Ahmed Tanvir, "Automatic Acute Lymphoblastic Leukemia Detection and Comparative Analysis from Images", CoDIT'19 | Paris, France- April 23-26, 2019

[4] Angelo Genovese, "Histopathological Transfer Learning for Acute Lymphoblastic Leukemia Detection",DOI:10.1109/CIVEMSA52099.2021.949367

[5] Guguloth Ramesh, "Study of Machine Learning Algorithms on Early Detection of Leukemia", doi.org/10.1038/s41598024-58580-z

[6] M. Akter Hossain, M. Islam Sabik, I. Muntasir, A. K. M. Muzahidul Islam, S. Islam, and A. Ahmed, ``Leukemia detection mechanism through microscopic image and ML techniques,'' in Proc. IEEE REGION 10 Conf.n(TENCON), Nov. 2020, pp. 61_66.

[7] F. T. Zahra, N. A. Nahid, M. R. Islam, M. M. A. Al-Mamun, M. N. H. Apu, Z. Nahar, A. L. Kabir, S. K. Biswas, M. U. Ahmed, M. S. Islam, and A. Hasnat, ``Pharmacogenetic variants in MTHFR gene are signi_cant predictors of methotrexate toxicities in Bangladeshipatients with acute lymphoblastic leukemia,'' *Clin. Lymphoma MyelomaLeukemia*, vol. 20, no. 2, pp. e58_e65, Feb. 2020. [Online]. Available: https://www.sciencedirect.com/

[8]Y. Takahashi, M. Ueki, M. Yamada, G. Tamiya, I. N. Motoike, D. Saigusa, M. Sakurai, F. Nagami, S. Ogishima, S. Koshiba, K. Kinoshita,M. Yamamoto, and H. Tomita, ``Improved metabolomic data-based prediction of depressive symptoms using nonlinear machine learning with feature

selection,'' *Transl. Psychiatry*, vol. 10, no. 1, pp. 1_12, Dec. 2020.

[9] Y. R. Park, Y. J. Kim,W. Ju, K. Nam, S. Kim, and K. G. Kim, ``Comparison of machine and deep learning for the classi_cation of cervical cancer based on cervicography images,'' *Sci. Rep.*, vol. 11, no. 1, pp. 1_11, Dec. 2021.

[10] M. A. Hossain, M. I. Sabik, M. M. Rahman, S. N. Sakiba, A. M. Islam, S. Shatabda, S. Islam, and A. Ahmed, ``An effective leukemia predictiontechnique using supervised machine learning classification algorithm,'' in *Proc. Int. Conf. Trends Computer. Cognit. Eng.*, Cham, Switzerland:Springer, 2021, pp. 219_229.

[11] N. M. Deshpande, S. Gite, and R. Aluvalu, ''A review of microscopic analysis of blood cells for disease detection with AI perspective,'' *PeerJComput. Sci.*, vol. 7, pp. 1–27, Apr. 2021, doi: 10.7717/peerj-cs.460.

[12] M. O. Aftab, M. J. Awan, S. Khalid, R. Javed, and H. Shabir, ''Executing spark BigDL for leukemia detection from microscopic images usingtransfer learning,'' in *Proc. 1st Int. Conf. Artif. Intell. Data Anal. (CAIDA)*, Apr. 2021, pp. 216–220, doi: 10.1109/CAIDA51941.2021.9425264.

[13] K. K. Anilkumar, V. J. Manoj, and T. M. Sagi, ''Automated detection of B cell and T cell acute lymphoblastic leukaemia using deep learning,'' *IRBM*, vol. 43, no. 5, pp. 405–413, 2021, doi: 10.1016/j.irbm.2021.05.005.

[14] N. Ouyang, W. Wang, L. Ma, Y. Wang, Q. Chen, S. Yang, J. Xie, S. Su, Y. Cheng, Q. Cheng, L. Zheng, and Y. Yuan, ''Diagnosing acutepromyelocytic leukemia by using convolutional neural network,'' *Clinica Chim. Acta*, vol. 512, pp. 1–6, Jan. 2021, doi: 10.1016/j.cca.2020.10.039.

[15] M. J. Ahmed and P. Nayak, ''Detection of lymphoblastic leukemia using VGG19 model,'' in *Proc. 5th Int. Conf. I-SMAC (IoT Social, Mobile,Analytics Cloud) (I-SMAC)*, Nov. 2021, pp. 716–723, doi: 10.1109/ISMAC52330.2021.9640955.

[16] C. Matek, S. Krappe, C. Münzenmayer, T. Haferlach, and C. Marr, ''Highly accurate differentiation of bone marrow cell morphologies using deep neural networks on a large image data set,'' *Blood*, vol. 138, no. 20, pp. 1917–1927, Nov. 2021, doi: 10.1182/blood.2020010568.

[17] Genovese, M. S. Hosseini, V. Piuri, K. N. Plataniotis, and F. Scotti, ''Histopathological transfer learning for acute lymphoblasticleukemia detection,'' in *Proc. IEEE Int. Conf. Comput. Intell. Virtual Environ. Meas. Syst. Appl. (CIVEMSA)*, Jun. 2021, pp. 1–6, doi:10.1109/CIVEMSA52099.2021.9493677.

[18] M. A. Hamza, A. A. Albraikan, J. S. Alzahrani, S. Dhahbi, I. Al-Turaiki, M. A. Duhayyim, I. Yaseen, and M. I. Eldesouki, ''Optimal deep transfer learning-based human-centric biomedical diagnosis for acute lymphoblasticleukemia detection,'' *Comput. Intell. Neurosci.*, vol. 2022, pp. 1–13, May 2022, doi:10.1155/2022/7954111.

[19] M. L. Claro, R. D. M. S. Veras, A. M. Santana, L. H. S. Vogado, G. B. Junior, F. N. S. D. Medeiros, and J. M. R. S. Tavares, ''Assessingthe impact of data augmentation and a combination of CNNs on leukemia classification,'' *Inf. Sci.*, vol. 609, pp. 1010–1029, Sep. 2022,doi: 10.1016/j.ins.2022.07.059.

[20] R. Gupta, S. Gehlot, and A. Gupta, ''C-NMC: B-lineage acute lymphoblastic leukaemia: A blood cancer dataset,'' *Med. Eng. Phys.*, vol. 103, May 2022, Art. no. 103793, doi: 10.1016/j.medengphy.2022.103793.

[21] C.-W. Wang, S.-C. Huang, Y.-C. Lee, Y.-J. Shen, S.-I. Meng,and J. L. Gaol, ''Deep learning for bone marrow cell detection and classification on whole-slide images,'' *Med. Image Anal.*,vol. 75, Jan. 2022, Art. no. 102270, doi: 10.1016/j.media.2021.102270.

[22] L. D. C. Magpantay, H. D. Alon, Y. D. Austria, M. P. Melegrito, andG. J. O. Fernando, ''A transfer learning-based deep CNN approach forclassification and diagnosis of acute lymphocytic leukemia cells,'' in *Proc. Int. Conf. Decis. Aid Sci. Appl. (DASA)*, Mar. 2022, pp. 280–284,doi: 10.1109/DASA54658.2022.9765000.

[23] P. K. Das,V. A. Diya, S. Meher, R. Panda, and A. Abraham, ''Asystematic review on recent advancements in deep and machine learning based detection and classification of acute lymphoblastic leukemia,'' *IEEE Access*, vol. 10, pp.

81741–81763, 2022, doi: 10.1109/ACCESS.2022.3196037.

[24] K. Talukdar, K. Bora, L. B. Mahanta, and A. K. Das, ‘‘A comparative assessment of deep object detection models for blood smearanalysis,’’ *Tissue Cell*, vol. 76, Jun. 2022, Art. no. 101761, doi:10.1016/j.tice.2022.101761.

[25] P. K. Das, B. Nayak, and S. Meher, ‘‘A lightweight deep learning system for automatic detection of blood cancer,’’ *Measurement*, vol. 191,Mar. 2022, Art. no. 110762, doi: 10.1016/j.measurement.2022.110762.

[26] Genovese, ‘‘ALLNet: Acute lymphoblastic leukemia detection using lightweight convolutional networks,’’ in *Proc. IEEE 9th Int. Conf. Comput.Intell. Virtual Environ. Meas. Syst. Appl. (CIVEMSA)*, Jun. 2022, pp. 1–6, doi: 10.1109/civemsa53371.2022.9853691.

[27] Anita and A. Yadav, ‘‘An intelligent model for the detection of white blood cells using artificial intelligence,’’ *Comput. Methods Programs Biomed.*, vol. 199, 2021, Art. no. 105893, doi:10.1016/j.cmpb.2020.105893

[28] D. R. Putri, A. Jamal, and A. A. Septiandri, ‘‘Acute lymphoblastic leukemia classification in nucleus microscopic images using convolutionalneural networks and transfer learning,’’ in *Proc. 2nd Int. Conf. Artif. Intell. Data Sci. (AiDAS)*, Sep. 2021, pp. 1–6, doi: 10.1109/AiDAS53897.2021.9574176.

[29] X. Zheng, P. Tang, L. Ai, D. Liu, Y. Zhang, and B. Wang, ‘‘White blood cell detection using saliency detection and CenterNet: A two?stageapproach,’’ *J. Biophotonics*, 2022, Art. no. e202200174, doi: 10.1002/jbio.202200174.

[30] D. R. Putri, A. Jamal, and A. A. Septiandri, ‘‘Acute lymphoblastic leukemia classification in nucleus microscopic images using convolutionalneural networks and transfer learning,’’ in *Proc. 2nd Int.nConf. Artif. Intell. Data Sci. (AiDAS)*, Sep. 2021, pp. 1–6, doi: 10.1109/AiDAS53897.2021.9574176.