

Utilizing Deep Neural Networks for the Recognition and Classification of Hematologic Cancers

Himanshi Arora¹, Akansha Yadav², Purvi Rastogi³ and Rajesh Thakur⁴

¹⁻³Student, Department of Information Technology, Meerut Institute of Engineering and Technology, Meerut, India
Email: himanshi.arora.it.2020@miet.ac.in, akansha.yadav.it.2020@miet.ac.in, purvi.rastogi.it.2020@miet.ac.in

⁴Assistant Professor, MIET, Department of Information Technology, Meerut Institute of Engineering and Technology, Meerut, India
Email: rajesh.thakur@miet.ac.in

Abstract— Leukemia, another name for bloodstream cancerous cells illness, is a potentially fatal illness which needs prompt and accurate assessment to be able to effectively cured. Throughout the study, it offers a novel strategy that can identify and categorize bloodstream cancerous cells illness utilizing deep learning approach. By employing the Python adaptation of the MobileNetV2 Architecture, we were able to attain impressively precise outcomes. The dataset utilized is "Blood Cells Cancer (ALL) dataset," categorized into four classes: benign, (malignant) early Pre-B, (malignant) Pre-B, and (malignant) Pro-B. Comprising 3242 images of Peripheral Blood Smears (PBS), the collection plays a crucial role in properly determining Acute Lymphoblastic Leukemia (ALL) in prevention of cancer.

Using the MobileNetV2 Architecture as its foundation, the deep learning model exhibits good performance. It obtained an excellent precision of 98.00% during training, demonstrating its capacity to accurately discriminate among various kinds of bloodstream cancerous cells. Additionally, the Validation precision of 96.00% underscores the model's robustness and generalization capabilities.

This endeavor exclusively showcases the efficacy of neural networks within the domain of healthcare picture processing, it likewise leads towards the prompt identification and categorization of bloodstream cancerous cells, thereby enhancing the clinical experience. Leveraging MobileNetV2 and Python enables us to develop a productive and easily available method for medical practitioners, paving the way for improved bloodstream cancerous cells illness assessment and prevention.

Index Terms— Machine Learning (ML), Classification, Detection, Segmentation, MobileNetV2 Architecture, Acute Lymphoblastic Leukaemia (ALL), Convolutional Neural Network (CNN)

I. INTRODUCTION

Neural networks allows computational models that integrate multiple computing tiers to learn interpretations of information at various tiers of encapsulation. These methods have greatly advanced the latest developments across multiple field. Neural networks reveals intricate patterns within huge amounts of information by applying the back propagation algorithm which provides proposal for modifications for the mechanism's inner settings which can be employed for determining structure of information in every tier based on illustration of tier immediately prior to

it. Neural networks succeeded in achieving great strides within the interpretation of footage, pictures, language, and recordings.

ML technology powers many aspects of modern society, including browsing the internet, networking activity screening, and online store recommendations. It is also increasingly found in consumer electronics like video equipment and cellphones. Algorithms for ML have been employed to identify objects within images, convert voice into written form, correlate products, job posts, and headlines to consumer preferences, and choose appropriate result sets for users' queries. An increasing number of such programs make use of a family of techniques called neural networks. Raw natural data processing was beyond the capabilities of conventional machine-learning approaches. For many years, a critical step in creating a ML system was creating a tool for extracting characteristics that transformed raw input, for instance the pixel values of a picture, into an acceptable inner visualization, which the neural subsection, which is usually a classifier, may use to identify or classify designs in input. When a machine is fed raw data, it may automatically recognize the representations needed for classification or identification thanks to a set of methods called representation learning. Many layers of representation are included into representation-learning techniques called deep learning approaches. They do this by constructing simple yet nonlinear modules, each of which raises the representation of the raw input at a level and makes it somewhat more abstract. Learning of very complicated functions is possible if enough of these transformations are couple.

A. Background History

The term "machine learning" was coined in 1959 by Arthur Samuel, IBM innovator in the domains of AI. A classic work of ML study from 1960s was Nilsson's Learning Machines, that focused primarily around ML for predicting patterns. Pattern recognition remained a popular topic in the 1970s. In 1981, a research was presented on the use of educational methodologies to train an artificial brain to comprehend forty distinct characters from an operating system platform.

An explicit explanation of procedures examined this area of ML was given by Tom M. Mitchell, which is frequently cited: If a computer program's performance at tasks in T , as measured by P , increases with experience E , then it is said to have learned from experience E with regard to some group of tasks T and performance measure P . Rather than describing the area in terms of neural mechanisms, this description of the activities that are under the purview of ML provides an essentially functional explanation. This is in line with the suggestion made by Alan Turing in his work Computing Machinery and Intelligence, where he suggests asking Can machines accomplish what we (as thinking creatures) can do instead of Can machines think?

Two objectives drive contemporary ML: first, organizing information using built algorithms while second is, forecasting upcoming occurrences using such algorithms.

Recognizing and treating bloodstream cancerous cells-related illnesses poses significant challenges, often requiring a considerable amount of time. Over the past decade, various practices have been developed for determining, assessing, and dividing bloodstream cancerous cells illnesses in individuals. Despite these advancements, there currently exists no approach or paradigm capable of automating examination of human red-blood cells to ascertain the existence of bloodstream cancerous cells

B. IEEE Base Paper Abstract

An approach possessing such capabilities possesses the capacity to enhance illness recognition and avoidance, leading to quicker healthcare treatment. This work demonstrates the development of neural network research for bloodstream cell aberration recognition. In this work, a Hybrid Ensemble Deep Learning approach was used for recognizing cancer of the bloodstream, with a precision score that exceeded 95%.

The present method for detecting blood cancer was built with an Ensemble Voting Classifier, a strong and extensively used machine learning approach. This approach was primarily concerned with categorizing blood cell pictures into two categories: "Normal" and "Leukaemia." The major goal was to discriminate between healthy and leukemia-infected blood cells, assisting in the early detection of this potentially fatal illness. The ensemble method used numerous base classifiers, each of which contributed to the final prediction. The system used the collective wisdom of these classifiers to obtain a 95% accuracy rate. This high accuracy highlights the system's ability to properly distinguish instances of leukemia from normal blood cell pictures.

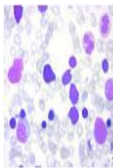
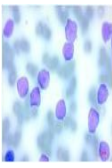
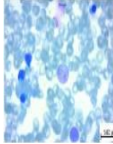
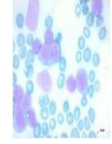
The ensemble approach amalgamated multiple base classifiers, each contributing to the final prediction. The system harnessed the collective intelligence of these classifiers to achieve an impressive accuracy rate of 95%. This high accuracy underscores the system's effectiveness in correctly identifying cases of leukemia from normal blood cell images.

The approach shortened the diagnosis procedure by limiting the categorization to only two classifications, "Normal" and "Leukaemia," simplifying decision-making for medical practitioners. This binary classification method not only improves the system's interpretability, but it also speeds up the detection of possible leukemia patients.

C. Supported Technologies

Dataset: In the first module of Utilizing Deep neural networks for the Recognition and classification of Hematologic Cancers, our team constructed a perspective for obtaining throughput information collection. Process of gathering data is the initial stage in the actual building of a neural network algorithm. This is an important stage that will impact the algorithm's overall success; the more accurate information we collect, the more effective the algorithm will be. There are several ways to collect data, such as personal involvement and web scraping. Our dataset is placed in the project and it's located in the model folder. The dataset is referred from the popular standard dataset repository kaggle where all the researchers refer it. The dataset consists of 3242 images. This thorough dataset contains 3,242 Peripheral Blood Smear (PBS) pictures acquired from 89 individuals suspected of having Acute Lymphoblastic Leukaemia (ALL). Highly competent laboratory technicians methodically processed and stained these photos. The two divisions of the information in the database are benign and malignant. Hematogenous cells belong in the benign group, while ALL, which additionally breaks down under Early Pre-B, Pre-B, and Pro-B ALL variants of malignant lymphoblasts, belongs in the malignant group. A professional used the sophisticated flow cytometry technology to determine the definitive categorization of cell kinds and subtypes, assuring the correctness and reliability of the dataset's annotations.

TABLE I. ILLUSTRATIONS FOR EVERY SET OF INFORMATION GROUPS

Class Name	Benign	Early Pre-B	Pre-B	Pro-B
Sample				
				

Importing the necessary libraries: In this project, we have chosen to utilize the Python programming language as our primary tool. Our initial step involves importing essential libraries to facilitate various tasks. These libraries include Keras, which will aid in constructing the core model, as well as Scikit-Learn (sklearn), which we will employ to partition the data into training and testing sets. Additionally, we will utilize the Python Imaging Library (PIL) to convert the images into numerical arrays. To assist with data manipulation and analysis, we will also make use of other widely-used collection of functions.

Acquiring the pictures: In this module our team acquire pictures from the dataset then convert them into a format that can be used for training and testing the model. This involves reading the images, resizing them, and normalizing the pixel values. Our team acquire pictures along with captions. Next, adjust the pictures to (224, 224) so that they are all identical in dimension during identification. After that, pictures get transformed as a numpy array.

Dividing the information in the database: Image operating data will be separated into training and testing sets in this lesson. Separate the train and test portions of the operating data. There are 20% of test data and 80% of train data. This will be done to train the model on a subset of the data, Verify the algorithm's functionality and assess its correctness using hypothetical information.

Building the model: Convolutional neural networks are a highly effective approach for recognition of pictures. The major distinction between CNN and classic neural networks is the convolution process. With a picture as input, CNN scans it several times to seek for certain characteristics. The freshly constructed frameworks obtained from the first convolution technique are displayed in the second column (layer). Relevant data regarding a single element

and its existence in the scanned picture is contained in every single frame. Where an element is clearly apparent, the consequent frame will have bigger values; where there aren't any or few such characteristics, the values will be smaller. Later then, the procedure is carried out amount of repetitions for each of the collected frames. I decided to use a traditional MobileNetV2 model for this project, which has just two convolution layers.

The deeper the layer we merge, the greater number of abstract characteristics the algorithm explores. This procedure mimics how individuals see things. Consider a detailed image that shows the characteristics that are sought at different CNN level as an example. The algorithm is being used in this instance for spotting faces. The way the algorithm chooses which traits to prioritize is a possible point of contention. When developing the CNN initially, the sought characteristics are arbitrary. The CNN may gradually detect properties that correspond with the preset objective, such correctly identifying photographs from the experimental collection, by incrementally altering the weights that are distributed among synapses throughout instructing phase.

D. MobileNetV2| CNN model Architecture

A convolutional neural network architecture called MobileNetV2 was created specifically for mobile devices. Its foundation is an inverted residual structure with bottleneck level residual linkages. Lightweight depthwise convolutions are used in the intermediate expansion layer filters as a source of nonlinearity. The MobileNetV2 design consists of 19 residual bottleneck layers after a fully convolution layer with 32 filters.

II. PROPOSED WORK PLAN

A. Utilizing Deep Neural Networks for the Recognition and Classification of Hematologic Cancers

Leukemia, another name for bloodstream cancerous cells illness, is a potentially fatal illness which needs prompt and accurate assessment to be able to effectively cured. Throughout the study, it offers a novel strategy that can identify and categorize bloodstream cancerous cells illness utilizing deep learning approach. By employing the Python adaptation of the MobileNetV2 Architecture, we were able to attain impressively precise outcomes.

The dataset utilized is "Blood Cells Cancer (ALL) dataset," categorized into four classes: benign, (malignant) early Pre-B, (malignant) Pre-B, and (malignant) Pro-B. Comprising 3242 images of Peripheral Blood Smears (PBS), the collection plays a crucial role in properly determining Acute Lymphoblastic Leukemia (ALL) in prevention of cancer.

Using the MobileNetV2 Architecture as its foundation, the deep learning model exhibits good performance. It obtained an excellent precision of 98.00% during training, demonstrating its capacity to accurately discriminate among various kinds of bloodstream cancerous cells. Additionally, the Validation precision of 96.00% underscores the model's robustness and generalization capabilities.

This endeavor exclusively showcases the efficacy of neural networks within the domain of healthcare picture processing, it likewise leads towards the prompt identification and categorization of bloodstream cancerous cells, thereby enhancing the clinical experience. Leveraging MobileNetV2 and Python enables us to develop a productive and easily available method for medical practitioners, paving the way for improved bloodstream cancerous cells illness assessment and prevention.

Multi-Class Classification: Unlike the previous binary classification method, the suggested system allows for multi-class classification, allowing for the detection and categorization of numerous blood cell abnormalities, including distinct subtypes of leukaemia. This gives healthcare workers a better grasp of the disease's complexities.

Enhanced Accuracy: Leveraging the MobileNetV2 architecture, the proposed system is expected to achieve a higher level of accuracy in Utilizing Deep Neural Networks for the Recognition and Classification of Hematologic Cancers. Neural network approach like MobileNetV2 are known for their ability to handle complex image data effectively, resulting in improved diagnostic precision.

Detailed Insights: The proposed system aims to provide detailed insights into the detected abnormalities, including information about the type, severity, and specific characteristics of the identified blood cell abnormalities. This level of detail empowers healthcare professionals who customize treatment programs to meet the demands of each unique person being treated.

III. EXPERIMENTAL RESULT ANALYSIS

A. MobileNetV1

t is the expansion factor, c is the number of output channels, n is the repeating number, and s is the stride. Spatial convolution is performed using 3x3 kernels.

TABLE I. FORMULA TABLE FOR CALCULATION OF INPUT-OUTPUT OF MOBILENETV1 ARCHITECTURE

Input	Operator	Output
$h \times w \times k$	1x1 conv2d, ReLU6	$h \times w \times (tk)$
$h \times w \times tk$	3x3 dwse=s, ReLU6	$\frac{h}{s} \times \frac{w}{s} \times (tk)$
$\frac{h}{s} \times \frac{w}{s} \times tk$	linear 1x1 conv2d	$\frac{h}{s} \times \frac{w}{s} \times k'$

TABLE II. CALCULATIONS TABLE

Input	Operator	t	c	n	s
$224^2 \times 3$	conv2d	-	32	1	2
$112^2 \times 32$	bottleneck	1	16	1	1
$112^2 \times 16$	bottleneck	6	24	2	2
$56^2 \times 24$	bottleneck	6	32	3	2
$28^2 \times 32$	bottleneck	6	64	4	2
$14^2 \times 64$	bottleneck	6	96	3	1
$14^2 \times 96$	bottleneck	6	160	3	2
$7^2 \times 160$	bottleneck	6	320	1	1
$7^2 \times 320$	conv2d 1x1	-	1280	1	1
$7^2 \times 1280$	avgpool 7x7	-	-	1	-
$1 \times 1 \times 1280$	conv2d 1x1	-	k	-	-

B. MobileNetV2

TABLE III. CALCULATIONS TABLE

Size	MobileNetV1	MobileNetV2	ShuffleNet (2x,g=3)
112x112	64/1600	16/400	32/800
56x56	128/800	32/200	48/300
28x28	256/400	64/100	400/600K
14x14	512/200	160/62	800/310
7x7	1024/199	320/32	1600/156
1x1	1024/2	1280/2	1600/3
max	1600K	400K	600K

Further investigation is done into the trade-offs between performance for width multipliers and input resolutions. The network computational cost can reach 585M MAdds, with model sizes ranging from 1.7M to 6.9M parameters. 16 GPUs with a batch size of 96 are used to train the network.

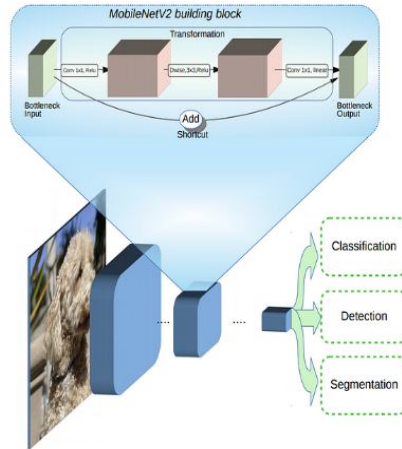


Figure 1 Illustration in regards to the MobileNetV2 Infrastructure. Blue blocks indicate the composite convolutional construction blocks seen above

IV. CONCLUSION

The project "Utilizing Deep Neural Networks for the Recognition and Classification of Hematologic Cancers" has culminated in the development of an innovative system that holds promise in the domain of blood cancer diagnosis. Leveraging the power of the MobileNetV2 architecture and Python programming.

The project has made significant strides in improving the accuracy and granularity of blood cancer detection. The earlier binary classification system, which categorized blood cell images as "Normal" or "Leukemia," served as a foundation upon which the proposed system was built. However, the limitations of this earlier system, including its lack of granularity and limited classification scope, prompted the development of the proposed system.

Proposed system, with its multi-class classification capability, has expanded diagnostic horizon by distinguishing various subtypes of leukemia, such as "Benign," "[Malignant] early Pre-B," "[Malignant] Pre-B," and "[Malignant] Pro-B." This expanded classification provides healthcare professionals with more comprehensive understanding of disease's complexity, ultimately leading to more informed and tailored treatment decisions.

The core of the proposed system, the MobileNetV2 architecture, has demonstrated its efficacy in handling complex and diverse blood cell images. With an emphasis on accuracy, the system has the potential to significantly improve the precision of blood cancer diagnoses, contributing to earlier interventions and improved patient outcomes.

Furthermore, the system's ability to provide detailed insights into the detected abnormalities, including type and severity, empowers healthcare professionals with the information they need to make informed clinical decisions.

Adaptability has been a key consideration in the project's design, ensuring that the system can evolve alongside advancements in blood cancer research and diagnostic criteria. This adaptability is crucial in maintaining the system's relevance and effectiveness over time.

Finally, the initiative marks a significant advancement in the identification and categorization of blood cancer.

REFERENCES

- [1] K. Kessenbrock, V. P. Laks, and Z. Werb, "Matrix metalloproteinases: Regulators of the tumor micro environment," *Cell*, vol. 141, no. 1, pp. 52–67, Apr. 2010.
- [2] A. Rehman, N. Abbas, T. Saba, S. I. U. Rahman, Z. Mehmood, and H. Kolivand et al., "Classification of acute lymphoblastic leukemia using deep learning," *Microsc. Res. Techn.*, vol. 81, no. 11, pp. 1310–1317, Nov. 2018.
- [3] S. Shafique and S. Tehsin, "Acute lymphoblastic leukemia detection and classification of its subtypes using pre-trained deep convolutional neural networks," *Technol. Cancer Res. Treatment*, vol. 17, Jan. 2018, Art. no. 153303381880278.
- [4] M. Hallek, P. Leif Bergsagel, and K. C. Anderson, "Multiple myeloma: Increasing evidence for a multistep transformation process," *Blood*, vol. 91, no. 1, pp. 3–21, Jan. 1998.
- [5] J. G. Kelton, A. R. Giles, P. B. Neame, P. Powers, N. Hageman, and J. Hirsch et al., "Comparison of two direct assays for platelet-associated IGG (Paigg) in assessment of immune and non-immune thrombocytopenia," *Blood*, vol. 55, no. 3, pp. 424–429, 1980.
- [6] M. Perkonig, J. Hofmanninger, B. Menze, M.-A. Weber, and G. Langs, "Detecting bone lesions in multiple myeloma patients using transfer learning," in *Data Driven Treatment Response Assessment and Preterm, Perinatal, and Pediatric Image Analysis*. Cham, Switzerland: Springer, 2018, pp. 22–30.
- [7] A. Jemal, R. Siegel, E. Ward, Y. Hao, J. Xu, and M. J. Thun et al., "Cancer statistics, 2009," *CA, A Cancer J. Clin.*, vol. 59, no. 4, pp. 225–249, Jul. 2009.
- [8] Y. Liu and F. Long, "Acute lymphoblastic leukemia cells image analysis with deep bagging ensemble learning," in *CNMC Challenge: Classification in Cancer Cell Imaging*. Singapore: Springer, 2019, pp. 113–121.
- [9] A. B. Kul'chyn'skyi, V. M. Kyjenko, W. Zukow, and I. L. Popovych, "Causal neuro-immune relationships at patients with chronic pyelonephritis and cholecystitis. Correlate ions between parameters EEG, HRV and white blood cell count," *Open Med.*, vol. 12, no. 1, pp. 201–213, Jul. 2017.
- [10] S. Kant, "Leukonet: Dct -based CNN architecture for the classification of normal versus leukemic blasts in b-all cancer," 2018, arXiv:1810.07961. [Online]. Available: <https://arxiv.org/abs/1810.07961>
- [11] Gao Huang, "Densely connected convolutional networks," in *CVPR*, 2017, vol. 1, p. 3.
- [12] Tsung-Yu Lin, "Bilinear cnn models for fine-grained visual recognition," in *Proceedings of ICCV*, 2015, pp. 1449–1457.
- [13] Karen Simonyan and Andrew Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.
- [14] Anubha Gupta and Rahul Duggal, "P -TELU: Parametric tanh hyperbolic linear unit activation for deep neural networks," in *Computer Vision Workshop (ICCVW)*, 2017 IEEE International Conference on. IEEE, 2017, pp. 974–978.
- [15] Kaiming He, "Delving deep into rectifiers: Surpassing human-level performance on image net classification," in *Proceedings of ICCV*, 2015, pp. 1026–1034.
- [16] Arun Mohan "Review On Mobile Net v2" in *Data Driven Investor*, jun 2020
- [17] Blood Cancer Dataset, Nov 2023 <https://www.kaggle.com/datasets/jayaprakashpondy/blood-cancer-dataset>

- [18] Andrew Srisuwananukorn, Mohamed E Salama and Alexander T. Pearson "Deep learning applications in visual data for benign and malignant hematologic conditions: a systematic review and visual glossary" Published online 2023 Jan in *Haematologica*
- [19] Yingyu Liang "Machine Learning Basics Lecture 1: Linear Regression" in Princeton University COS 495
- [20] Md. Wahidur Rahman, Mohammad Gazi Golam Faruque, Kaniz Roksana, A.H.M. Saifullah Sadi, Mohammad Motiur Rahman and Dr Mir Mohammad Azad et al. "Multiclass blood cancer classification using deep CNN with optimized features" in *Array*, ResearchGate Publication, *Array* 18(4):100292, May 2023
- [21] Anilkumar K.K., Manoj V.J, Sagi T.M. "Automated Detection of B Cell and T Cell Acute Lymphoblastic Leukaemia Using Deep Learning" in *IRBM*, Science Direct, *IRBM* Volume 43, Issue 5, October 2022, Pages 405-413
- [22] Namrata Dhanda, Stuti Shukla Datta and Mudrika Dhanda "Machine Learning Algorithms" in *Computational Intelligence in the Internet of Things*, 2019 |Pages: 24
- [23] P. Yadav, Deepak Kumar, Anand Singh Jalal, Ankit Kumar, Kamred Udham Singh & Mohd Asif Shah et al. "Morphological diagnosis of hematologic malignancy using feature fusion-based deep convolutional neural network" in *Scientific Reports* volume 13, Article number: 16988, 2023
- [24] Farkhondeh Asadi, Azamossadat Hosseini, Davood Bashash, Hassan Abolghasemi and Arash Roshanpour, "Machine Learning in Detection and Classification of Leukemia Using Smear Blood Images: A Systematic Review" in *Hindawi Journals*, June 2021
- [25] Ahmad Ibrahim, Hoda K. Mohamed, Ali Maher and Baochang Zhang "A Survey on Human Cancer Categorization Based on Deep Learning" in *Frontiers Artificial Intelligence*, Volume 5, June 2022
- [26] Noortaz Rezaoana, Mohammad Shahadat Hossain and Karl Andersson "Detection and Classification of Skin Cancer by Using a Parallel CNN Model" in *IEEE International Women in Engineering (WIE) Conference on Electrical and Computer Engineering in 2020*
- [27] Victor Ikechukwu Agughasi, "i-Net: a deep CNN model for white blood cancer segmentation and classification" in *International Journal of Advanced Technology and Engineering Exploration*, October 2022
- [28] Q. Arbab, M.Q. Khan, H.J.T.S. Ali "Automatic Detection and Classification of Acute Lymphoblastic Leukemia Using Convolution Neural Network", 3 (2022)
- [29] N. Ahmed "Identification of leukemia subtypes from microscopic images using convolutional neural network", 2019, p. 104
- [30] Sampathila N "Customized deep learning classifier for detection of acute lymphoblastic leukemia using blood smear images". *Healthcare*; 2022, [MDPI].
- [31] Pallegama R "Acute lymphoblastic leukemia detection using convolutional neural network". 2020, 26529. 10(6).
- [32] Abd El-Ghany S, Elmogy M, El-Aziz A.J.D. "Computer-Aided Diagnosis System for Blood Diseases Using EfficientNet-B3 Based on a Dynamic Learning Algorithm", 2023;13(3):404.
- [33] Mishra S "Texture feature based classification on microscopic blood smear for acute lymphoblastic leukemia detection", 2019;47:303–11.