

Intelligent Data Processing and Decision Systems for Leukemia Identification via Machine Learning

Kiran¹ Anil Sangwan²
M.Tech, ECE Associate Professor, ECE

Department of Electronic and Communication Engineering, University
Institute of Engineering and Technology (UIET), M.D.University,
Rohtak0124001(Haryana)

Abstract - Conventional leukemia diagnosis depends on labour-intensive manual analysis of blood samples and bone marrow aspirations, which often proves slow, inconsistent, and error-prone. This study evaluates a machine learning-based approach for automated classification of leukemia - related microscopic blood-smear images. The study focuses on machine learning and deep learning methods for classifying microscopic blood smear images, with experimental evaluation conducted on the C-NMC 2019 dataset. The proposed experimental pipeline consists of image preprocessing, model-based feature representation, weighted ensemble classification, and test-time augmentation followed by evaluation using accuracy, precision, F1-score, and confusion-matrix analysis. A weighted ensemble approach with Test-Time Augmentation (TTA) was evaluated on a held-out test set to distinguish Acute Lymphoblastic Leukemia (ALL) from hematogones (HEM). The evaluated model achieved an overall accuracy of 93.51%, with a macro precision of 0.9201, macro recall of 0.9347, and macro F1-score of 0.9268. On the held-out Fold 0 test set of 3,527 images, 2,243 ALL images and 1,055 HEM images were correctly classified. The findings demonstrate the potential of machine-learning-based microscopic image analysis as a decision-support approach; however, the results are experimental and require validation on independent clinical datasets before clinical deployment.

Keywords: Leukemia, Machine Learning, Deep Learning, C-NMC 2019, ALL, HEM, Microscopic Blood-Smear Images, Image Classification, Test-Time Augmentation

INTRODUCTION

Accurate and timely diagnosis is a vital goal in healthcare. However, leukemia poses some challenges for healthcare providers due to the variety of abnormal blood cells and the difficulty in distinguishing them from similar-looking cells. Leukemia is a group of cancers that affect the blood, specifically the production and accumulation of abnormal blood cells. Examining blood smears under a microscope provides valuable morphological information about the cells. However, this process is tedious and subjective.

Artificial intelligence advances, such as machine learning and deep learning, offer opportunities for automated medical imaging analysis. Several studies have explored the use of machine learning algorithms to automate the analysis of leukemia-related blood smear images. However, the performance of these algorithms varies according to several factors. Despite some successes with automation, there are challenges in automating leukemia detection. These include staining and image acquisition differences, class imbalance, and generalization issues. The results from automated systems may be specific to the datasets used to train them. Consequently, reproducible evaluation of automated systems is crucial to developing better screening and decision support systems.

Acute lymphoblastic leukemia is a type of leukemia that affects many children. Therefore, identifying abnormal cells in these patients is vital for hematological image analysis research. However, there is variability in the reported performance of these systems. For this reason, the results must be viewed within the context of the research study.

This research study evaluates a machine learning algorithm to identify acute lymphoblastic leukemia from hematogones (HEM) using images from the C-NMC 2019 dataset. Performance is evaluated using various metrics and test-time augmentation. Historical methods for classifying leukemia, such as the FAB system, use morphological classification of abnormal blast cells. Modern classification systems use morphological, immunophenotypic, cytogenetic, and molecular data. ML image classification systems are not a complete diagnostic system; they are designed for image classification tasks.

2. LITERATURE REVIEW

A number of recent studies have put machine and deep learning to the test for leukemia analysis on various forms of data. In a systematic review of medical images for the diagnosis of acute lymphoblastic leukemia (ALL), Shah et al. (2026) [1] covered the ground from preprocessing and feature extraction to CNN and Vision Transformer classifications, making the case for systems that can be deployed in a clinical setting and are externally validated. Tande et al. (2025) [2] turned their attention to peripheral blood smear images, surveying ML and DL techniques such as MobileNet, DenseNet, ResNet and CNN. They were quick to point out the obstacles in the field: stain variability, an imbalance in datasets, sparse annotations and a dearth of clinical validation. For rapid classification of acute leukemia, Zhang et al. (2025) [3] put forward a SERS approach aided by deep learning. By employing a transformer model on CSF samples, they showed how AI and SERS can be combined to tell apart the various subtypes. Kandi et al. (2026) [4] have also made contributions with microarray gene-expression data for 13-class leukemia classification. Their work evaluated a range of models from RNN and MLP to a hybrid CNN–LSTM–MLP; after using SMOTE to correct for class imbalance, the latter was the clear winner with 96.6% accuracy and an F1-score of 96.5%. Haque et al. (2024) [5] put forward a thorough method for the automated diagnosis of leukemia that draws on image processing as well as machine and transfer learning. The research was centred on microscopic images of peripheral blood smears and addressed the problem of classification in both binary and multiclass terms. After subjecting the data to preprocessing and augmentation, the authors turned to convolutional neural networks to extract deep features. They then put a number of machine-learning and transfer-learning models to the test. What their findings show is that the latter are capable of strong performance; the Inception-ResNet model, for instance, delivered an F1-score of 96.07 per cent on the binary task and 95.89 per cent for multiclass. Overall, the study makes a case for the value of integrating deep learning with image processing in this field. Hendrawan and colleagues (2025) [6] offered a different take on ALL detection with a hybrid of pre-trained ResNet-50 and ViT. On the C-NMC 2019 dataset, the ResNet handled local cellular features while the ViT took in the broader context of the microscopic image, resulting in 98% accuracy and an AUC of 0.99. Altalhi et al. (2026) [7] introduced an explainable framework that pairs LIME and Grad-CAM with transfer-learning and CNN. With a corpus of over 66,000 images of healthy and leukemic cells (ALL, AML, CLL, CML), their MobileNetV2 model hit 97.9% accuracy and gave clinicians a meaningful visual of the cell regions. Other notable work includes that of Abbasi et al. (2024) [8] who compared several algorithms in an integrated multi-omics study and found the RNN to be particularly effective at 98% accuracy. Ilyas et al. (2025) [9] distilled 22,283 gene features down to 25 from the CuMiDa dataset for subtype classification; an LSTM model proved superior, posting perfect scores across the board. Talaat et al. (2023/2024) [10] used fuzzy optimization to fine-tune CNN hyperparameters on the C-NMC_Leukemia dataset, arriving at 99.99% accuracy in separating normal from abnormal images. Finally, Afsar et al. (2026) [11] provided an overview of the state of deep learning in this area, noting the promise of multimodal AI but cautioning against issues like dataset bias and poor generalizability.

3. SYSTEMATIC METHODOLOGY

To ensure an unbiased and systematic selection of pertinent research, this review has been carried out in accordance with established guidelines. We conducted a thorough search of published articles on IEEE Xplore, PubMed, Scopus, ScienceDirect, SpringerLink, Google Scholar and MDPI, using terms including Leukemia, Artificial Intelligence, Deep Learning, CNN and Blood Smear Images. The inclusion criteria were strict: only peer-reviewed work applying ML or DL to leukemia detection with reported metrics (e.g. precision, recall, specificity, F1-score) was considered. There was no place for review papers, duplicates, non-English pieces or studies lacking experimental backing.

From the studies that made the cut, we extracted the key details – the datasets, the algorithms, the evaluation methods and the findings. This allowed for a comparative analysis of the different AI approaches and an assessment of where the challenges and opportunities lie in leukemia diagnosis.

4. DATASETS

One cannot have a successful machine learning model for leukemia without access to well-annotated, high-quality data. Whether it is haematological results, flow cytometry or clinical records, these resources are what allow for the proper training and testing of classification algorithms. Public datasets are therefore the norm, offering a standardised way to benchmark performance.

The C-NMC 2019 [12] is perhaps the most popular of these for leukemia classification. It is a suitable benchmark for any deep learning model, containing 15,135 microscopic images of lymphocytes from 118 patients sorted into HEM (healthy) and ALL (leukemic) categories. Of the 10,661 images in the training set, 7,272 are ALL and 3,389 HEM, with a further 1,867 reserved for the preliminary test.

5. DATA PREPROCESSING

Before model training, the microscopic images undergo preprocessing to obtain consistent input and improve model generalization. The images are resized to 224×224 pixels and normalized according to the required input distribution. Data augmentation techniques, including horizontal and vertical flipping, rotation, brightness, and contrast variations, are applied to increase image diversity and reduce overfitting.

6. FEATURE EXTRACTION

Feature extraction is performed to capture significant morphological properties of leukaemia cells. With the help of deep learning models such as CNN, ResNet, DenseNet, and MobileNet, features such as shape, texture, intensity, and structure of the cells are automatically extracted from preprocessed images. In the hybrid approach, features can be combined using various architectures to have a better feature representation for leukemia classification. In the reference study, ResNet-50 and Vision Transformer features are combined to achieve a more informative feature representation for microscopic cells.

7. MODEL TRAINING

Post feature extraction, the extracted features and preprocessed images are fed to selected machine learning and deep learning models. Dataset is divided into training, validation, and testing splits. Training split is used for learning morphological characteristics of normal and leukemic cells. Validation split is used for checking the performance of the model and avoiding overfitting. Testing split is used for final model evaluation purposes. In this study, various models such as CNN, ResNet, DenseNet, MobileNet, Random Forest, Naïve Bayes, Decision Tree, and Logistic Regression can be used for training and evaluation. While deep learning models will learn features automatically from the data, conventional machine learning models will use the extracted features as inputs. In the process of training, model parameters are optimized based on classification loss function. Various regularization techniques like dropout, data augmentation, and early stopping can be applied.

Models are then applied to predict which of the cells belong to either ALL (leukemia) or HEM (healthy). Final predictions are then compared to the ground truth labels and evaluated on metrics such as accuracy, precision, recall, F1 score and specificity

8. RESULT ANALYSIS

8.1 Overall Test Performance

We have put the proposed leukemia classification pipeline through its paces on the C-NMC 2019 dataset, with a view to reporting results in terms of accuracy, precision, recall, F1-score and the confusion matrix. Such metrics are necessary to give a full picture of the system's overall classification capabilities and how well it can be relied upon to identify ALL and HEM cells. For this evaluation we used a held-out Fold 0 as our independent test set

(3,527 images). The remaining Folds 1 and 2 were put together and split into 5,708 for training and 1,426 for validation. A weighted ensemble with test-time augmentation (TTA) was employed for the final run.

8.2 Class-wise Performance

The table below gives an overview of the performance on the C-NMC 2019 Fold 0 test set.

Metric	Result
Test set	C-NMC 2019 Fold 0
Number of test images	3,527
Accuracy	93.51%
Macro precision	0.9201
Macro recall	0.9347
Macro F1-score	0.9268

With the TTA and weighted ensemble, an overall accuracy of 93.51 per cent was attained on the held-out test set. The macro-average F1-score of 0.9268 is a sign of good balance between precision and recall for the two classes. In all, the test set was made up of 2,397 ALL and 1,130 HEM images.

8.3 Confusion Matrix Analysis

Looking at the numbers on a class-by-class basis, the model proved robust for both ALL and HEM.

Class	Precision	Recall	F1-score	Support
ALL	0.9676	0.9358	0.9514	2,397
HEM	0.8726	0.9336	0.9021	1,130
Macro Average	0.9201	0.9347	0.9268	3,527

Recall figures are comparable across the board, but one will note the lower precision for HEM (0.8726), which means some healthy-cell images were put down as ALL. By contrast, ALL registered a precision of 0.9676 and an F1-score of 0.9514.

8.4 Interpretation of the Results

As the confusion matrix shows, the vast majority of ALL and HEM cells were put in the right category. There were 2,243 correctly classified ALL and 1,055 for HEM. However, 154 ALL cells were misclassified as HEM and 75 HEM as ALL; the latter figure is what drags down the HEM precision.

Actual / Predicted	ALL	HEM	Total
ALL	2,243	154	2,397
HEM	75	1,055	1,130
Total	2,318	1,209	3,527

In short, 3,298 of the 3,527 test images were predicted correctly, accounting for the 93.51% accuracy figure.

8.5 Summary of Experimental Findings

It is important to regard these as the findings of an experimental exercise rather than proof of clinical diagnostic ability. The work is based on output from the supplied notebook using the C-NMC 2019 data. One would need to see independent validation from other patient populations, laboratories and scanners before any claims of generalization or clinical utility could be made. That said, the experiment does show that the weighted ensemble with TTA is capable of learning the morphological patterns in microscopic blood-smear images for binary classification. With both classes posting F1-scores over 0.90, the 0.9268 macro score demonstrates that the results are not skewed by a single class. The results of this study point to the viability of machine-learning-based microscopic image analysis as a form of decision support in leukemia screening research. That said, one should not view the performance reported here as a substitute for the judgment of a hematopathologist or for established clinical diagnostic protocols.

Finding	Observed result
Overall classification accuracy	93.51%
Macro F1-score	0.9268
ALL recall	0.9358
HEM recall	0.9336
ALL F1-score	0.9514
HEM F1-score	0.9021
Correct predictions	3,298 / 3,527
Incorrect predictions	229 / 3,527

9. CONCLUSION

By way of example, the data from the C-NMC 2019 dataset is telling. In our evaluation of the proposed classification pipeline on the held-out Fold 0 test set, the model put up an overall accuracy of 93.51 per cent and a macro F1-score of 0.9268. A look at the confusion matrix reveals that of the 3,527 cases, 3,298 were predicted correctly and 229 incorrectly; specifically, 2,243 ALL images and 1,055 HEM images were properly identified, with 154 ALL and 75 HEM images being misclassified. This was done using a weighted ensemble with Test-Time Augmentation (TTA) to distinguish Acute Lymphoblastic Leukemia (ALL) from healthy (HEM) cells after preprocessing the microscopic blood-smear images. The fact that the pipeline can so effectively pick up on morphological characteristics in these cell images speaks to the potential of such automated techniques.

In sum, while the findings make a case for AI-based image analysis in the classification and screening of leukemia, they remain experimental in nature. They are predicated on the supplied notebook evaluation and the C-NMC 2019 dataset and cannot stand in for a clinical diagnosis. Any move toward practical deployment will first require validation against independent clinical datasets.

10. FUTURE SCOPE

There is clearly scope to build on what has been demonstrated here in terms of robustness and clinical applicability. For instance, the binary task of separating ALL from hematogones (HEM) could be broadened to encompass a wider array of clinically relevant subtypes, offering a more thorough classification than the two classes we have focused on. We also see value in incorporating Explainable AI methods like Grad-CAM or LIME into the framework. This would add a layer of transparency by highlighting the cellular regions driving the model's predictions, which is of use to both researchers and medical staff. Another avenue for development is multimodal AI. Rather than relying on image analysis in isolation, future systems might integrate flow cytometry data, molecular information and other clinical parameters with the blood-smear images for a fuller picture of the patient. Finally, to guard against any bias inherent to the C-NMC 2019 dataset, it would be prudent for subsequent work

to test the pipeline on independent material from various hospitals and laboratories, with their own imaging devices and staining procedures.

There is scope for future work to probe more sophisticated hybrid architectures, be they a blend of convolutional neural networks and Vision Transformers or the like. One might also look at refining feature-fusion and ensemble methods in order to see if they yield sturdier feature representations and classification results when put through a systematic comparison with the existing pipeline. It would be prudent for subsequent experiments to offer a fuller account of model configurations, validation protocols and training parameters. The scope of evaluation could be broadened as well; where it makes sense, one should go past the usual accuracy, precision, recall and F1-score to report on specificity, sensitivity, ROC curves and ROC-AUC. Independent external testing must also figure prominently in any future assessment.

Once the necessary technical and clinical validation has been done, these approaches may be put forward for use in automated leukemia screening or as part of clinical decision-support. This is of particular interest in settings that lack ready access to specialist diagnostic expertise. That said, such systems are intended to be supportive in nature and to augment the established procedures and expert judgment of the clinic, not to take their place.

REFERENCES

- [1] S. R. F. R. J.-R. D. A.-S. W. Hussain Shaha, A systematic review of machine and deep learning techniques for acute, 2026. <https://doi.org/10.1016/j.artmed.2026.103393>
- [2] A. Tande, Artificial intelligence and datasets for leukemia diagnosis: A scoping review of machine learning and deep learning approaches, 2025. <https://doi.org/10.1016/j.mex.2025.103722>
- [3] Z. C. Dongjie Zhang, Rapid and sensitive acute leukemia classification <https://doi.org/10.1016/j.xcrm.2025.102320>
- [4] D. NithinKandia, Analysis of LEUKEMIA classification using micro-array Data with Deep Learning models, 2026.
- [5] A. A. S. 2. Rezaul Haque 1, Advancing Early Leukemia Diagnostics: A Comprehensive Study Incorporating Image Processing and Transfer Learning, 2024. <https://doi.org/10.3390/biomedinformatics4020054>
- [6] T. F. H. J. B. G. Jason Hendrawan, Hybrid deep learning model for acute lymphoblastic leukemia (ALL) detection using pre-trained ResNet-50 and Vision Transformer architecture, 2025. DOI:10.1016/j.procs.2025.09.036
- [7] S. H. A. a. M. *, Demystifying Deep Learning Decisions in Leukemia Diagnostics Using Explainable AI, 2026. <https://doi.org/10.3390/diagnostics16020212>
- [8] E. Y. Abbasi, A machine learning and deep learning-based integrated multi-omics technique for leukemia prediction, 2024. <https://doi.org/10.1016/j.heliyon.2024.e25369>
- [9] M. Ilyas, An efficient leukemia prediction method using machine learning and deep learning with selected features, 2025. <https://doi.org/10.1371/journal.pone.0320669>
- [10] F. M. T. . S. A. Gamel, Machine learning in detection and classification of leukemia using C-NMC_Leukemia, 2023. <https://doi.org/10.1007/s11042-023-15923-8>
- [11] P. A. a. T. M. Navamani*, Recent advances in deep learning for leukemia diagnosis: a scoping review of diagnostic modalities and fusion-based approaches, 2026. DOI 10.3389/fdgth.2026.1874695
- [12] S. K. S. K. P. G. A. & G. R. Mourya, C-NMC 2019 | C_NMC_2019 Dataset: ALL Challenge dataset of ISBI 2019, 2019. <https://doi.org/10.7937/tcia.2019.dc64i46r>