



## Research Paper

# Generative-Augmented Deep Feature Representation Learning for Accurate Blood Cancer Diagnosis from Histopathological Images

Geetha Kandukuri<sup>1</sup> , Adapa Mahendar<sup>2</sup> , Shanmugam Kolangiammal<sup>3</sup> , Kumareson Anish Pon Yamini<sup>4</sup> , Subramanian Famila<sup>5\*</sup>

1. Department of Computer Science and Engineering, CVR College of Engineering, Ibrahimpatnam, Ranga Reddy, Telangana, India.
2. Department of Computer Science and Engineering, CMR Technical Campus, Hyderabad-501401, Telangana, India.
3. Electronics and Communication Engineering, SRM Institute of Science and Technology, Kattankulathur, Chennai 603203, Tamil Nadu, India.
4. Electronics and Communication Engineering, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Chennai, India.
5. Electronics and Communication Engineering, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Chennai, India.

**Citation** Kandukuri G, Mahendar A, Kolangiammal S, Yamini KAP, Famila S. Generative-Augmented Deep Feature Representation Learning for Accurate Blood Cancer Diagnosis from Histopathological Images. *International Journal of Medical Toxicology and Forensic Medicine*. 2026; 16:E53421.

<https://doi.org/10.22037/ijmtfm.v16.53421>

### Article info:

**Received:** 10 September, 2026

**Accepted:** 22 September, 2026

**Published:** 06 October, 2026

### Keywords:

Histopathological images, Leukemia, Convolutional Neural Networks, feature extraction, Filtration

## ABSTRACT

**Background:** Leukemia is a cancer that impacts the blood and bone marrow. Detection and classification are traditionally performed using labor-intensive and specialized techniques. Detecting blood cancer through histopathological analysis of bone marrow tissue is a critical diagnostic process that requires accuracy and efficiency. In this study, we propose an automated system for detecting blood cancer (Leukemia) from bone marrow histopathological images using a hybrid approach combining Generative Adversarial Networks and Convolutional Neural Networks (GAN+CNN).

**Methods:** The system processes high-resolution images to identify malignancies by learning complex patterns from pixel-level features. Our model incorporates image preprocessing techniques such as stain normalization and noise reduction to improve image consistency and quality. This automated approach can speed diagnosis and reduce subjectivity in blood cancer detection, providing a robust tool to support clinical decision-making.

**Results:** The proposed GAN+CNN achieves 99.96% accuracy, 98.3% precision, 98.4% recall, and a 97.53% F1-score.

**Conclusion:** The proposed GAN+CNN method achieved better performance in leukemia detection and can support rapid, reliable diagnosis.

### \* Corresponding Author:

S. Famila, PhD

Associate Professor, Electronics and Communication Engineering, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Chennai, India.

E-mail: [familas1999@gmail.com](mailto:familas1999@gmail.com)



Copyright © 2026 The Author(s);

This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0);

<https://creativecommons.org/licenses/by-nc/4.0/>, which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

## Introduction

**B**lood cancer denotes a category of malignancies that impact the blood, bone marrow, and lymphatic system. These malignancies encompass several types of leukemia, lymphoma, and myeloma [1].

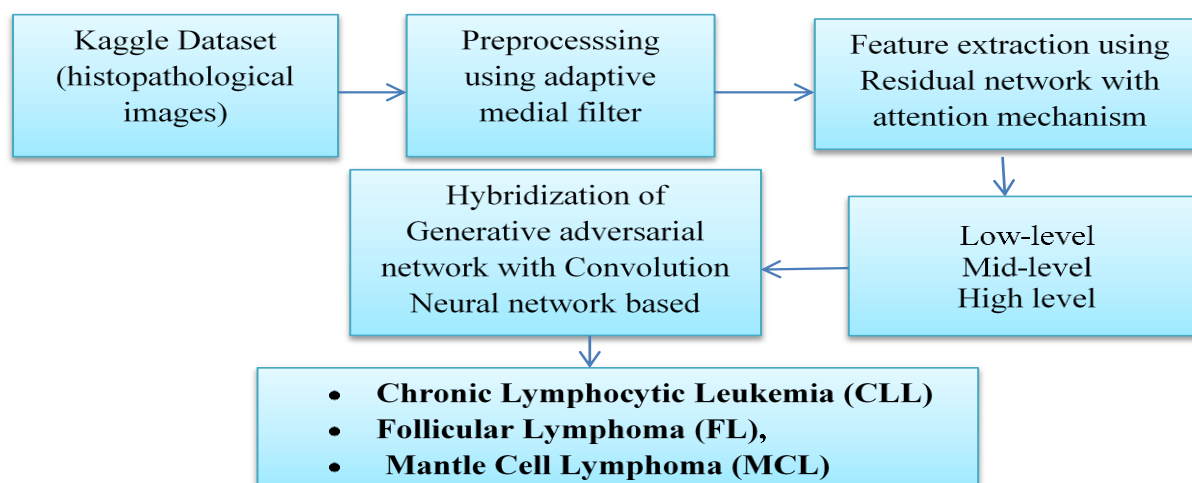
Effective blood cancer diagnostics is essential for early disease detection [2]. If you exhibit signs and symptoms of blood cancer, it is crucial to obtain medical consultation and get testing [3]. Flow cytometry is an effective method for examining individual cells to detect atypical traits. This procedure entails extracting a blood sample and filtering it through biocompatible nanosized particles to identify biochemical markers of malignancy [4]. Uncontrolled growth of lymphocytes, a type of leukocyte, causes lymphoma. Lymphocytes in the blood and lymphatic tissues protect the organism against several diseases [5]. Thus, computerized diagnosis of malignant lymphoma is needed to address existing challenges. Despite advances, conventional machine learning methods for H&E-stained histopathology image analysis have three limitations [6-9]. The initial phase involves unsupervised extraction of handcrafted properties that are uncorrelated with subsequent WSI analytical assignments [10]. Unlike traditional machine learning approaches, deep learning can extract meaningful features for specific WSI analysis tasks by mapping input directly to output. Moreover, heterogeneous patterns in WSIs can induce variability across samples, limiting the generalization of handcrafted features [14]. Deep learning algorithms can describe complex patterns when trained on sufficient WSI data. Recently, high-throughput tissue banking and the preservation of digitized whole slide images have made deep learning algorithms more scalable because they can analyze large amounts of data and perform complex calculations efficiently in terms of cost and time.

The authors in [15] offer a sophisticated automated approach employing deep neural networks to detect ALL blast cells in tiny blood smear images. It achieves an impressive 97% success rate with MobileNetV2. In [16], convolutional neural networks are used to automatically detect Acute Lymphoblastic Leukemia. In [17], random forests, KNN, naive Bayes, decision trees, VGG16, ResNet50, and DenseNet201 were employed. In [18], deep transfer learning models VGG16, ResNet50, AlexNet, MobileNet, GoogLeNet, and EfficientNetB0 were used for feature extraction and classification. [19] offers a decision-assistance system, the Histogram Threshold Segmentation Classifier (HTsC). It uses blood smear image color and brightness fluctuations for assessment. A hybrid model utilizing geometric transformation (GT) and a deep convolutional autoencoder (DCAE) is proposed [20]. Based on hybrid features acquired using LBP, GLCM, and FCH, [21] uses ANN, FFNN, and SVM. AlexNet + SVM achieved 100% accuracy, GoogLeNet 98.1%, and ResNet-18 100%.

In this study, we propose an automated system for detecting blood cancer (Leukemia) from bone marrow histopathological images using Hybridization of Generative Adversarial Networks and Convolutional Neural Networks (GAN+CNN). The system processes high-resolution images to identify malignancies by learning complex patterns from pixel-level features. The proposed architecture is optimized to handle variations in cell morphology, staining, and image quality, achieving superior accuracy, sensitivity, and specificity.

## Materials and Methods

Figure 1 outlines a pipeline for classifying histopathological images into specific lymphoma subtypes. The process begins with the Kaggle Dataset containing histopathological images, which undergo



**Figure 1.** Overall block diagram for leukemia classification.

preprocessing using an adaptive median filter to remove noise and enhance image quality. Subsequently, feature extraction is performed using a Residual Network (ResNet) integrated with an attention mechanism to capture multi-level features (low-level, mid-level, and high-level) for detailed analysis. The extracted features are fed into a hybrid model combining a Generative Adversarial Network (GAN) and a Convolutional Neural Network (CNN) for classification. This hybrid approach leverages GANs to enhance training and CNNs for accurate feature-based classification. Finally, the system classifies the images into one of three lymphoma subtypes: Chronic Lymphocytic Leukemia (CLL), Follicular Lymphoma (FL), or Mantle Cell Lymphoma (MCL).

### Dataset description

The proposed methods assessed histological images of malignant lymphomas obtained from biopsies and stained with H&E. Figure 2 below shows sample images from the malignant lymphoma dataset. The dataset was obtained from Kaggle and consisted of 15,000 images uniformly distributed across three types of malignant lymphoma: CLL, FL, and MCL. All images in the collection were in JPEG format, and each image was  $512 \times 512$  pixels.

### Preprocessing using adaptive median filter

AMF is a category of filters whose filtering efficacy is adjusted according to local image statistics. AMF is

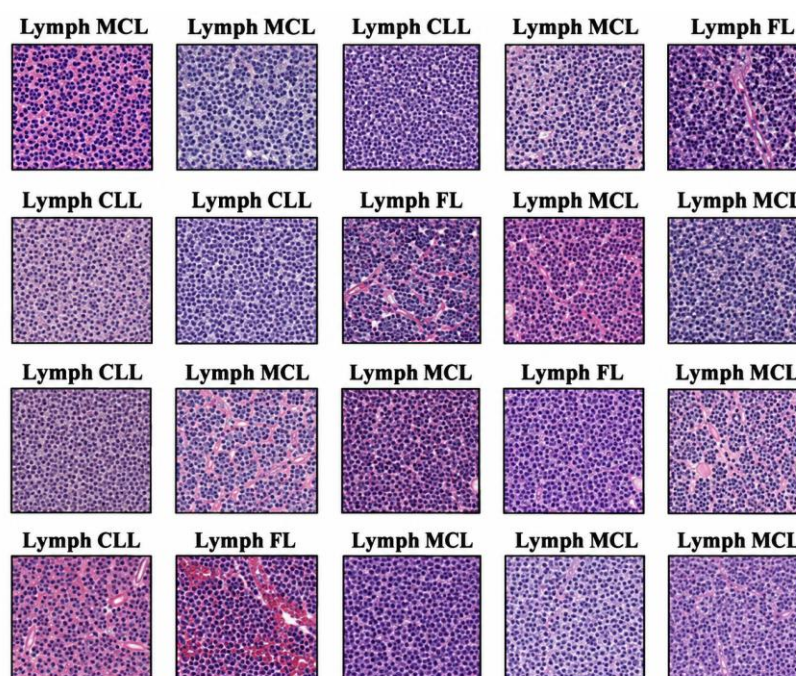
a spatial model that uses a dynamic kernel. The initial picture ( $I(x, y)$ ) is processed utilizing Equation (1),

$$I_{AMF}(x, y) = I_{MF}(x, y) + \frac{\sigma_L^2 - \sigma_G^2}{\sigma_L^2} (I(x, y) - I_{MF}(x, y)) \quad (1)$$

The global variance of picture noise is  $\sigma_G^2$  while the local variance is  $\sigma_L^2$ . In homogeneous areas,  $\sigma_L^2$  is tiny and the filter operates aggressively, resulting in an equation that approaches  $I_{MF}(x, y)$ . In the edge area,  $\sigma_L^2$  is large and the filter decreases noise, resulting in an equation close to  $I(x, y)$ . Although AMF pictures have minimal noise, spatial resolution is compromised.

### Feature extraction using Residual network

Using a residual network integrated with an attention mechanism to extract features from histological images of blood cancer significantly improves diagnostic precision. Residual networks, such as ResNet, mitigate the vanishing gradient issue and improve the training efficacy of deep neural networks through skip connections that promote information transfer across layers. These networks excel at extracting hierarchical features from histopathology images, capturing both low- and high-level patterns essential for identifying malignant areas. The geometric characteristics were determined using area ( $A$ ), perimeter ( $P$ ), roundness (meter  $P$ , roundness  $R$ , and segmented cells.  $L$ . The number of segmented pixels for each nucleus is  $x$  sub  $k$ , and start equation cap



**Figure 2.** Sample dataset of malignant lymphoma.

A is its area; perimeter is  $P$ ,  $tA$  area is  $is$ 's roundness  $i$   
 $R$ .  $k$  is the number of pixels at the cell boundary of  
 each nucleus,

$$f_1: L \quad (2)$$

$$f_2: A = \frac{1}{L} \sum_{k=1}^L x_k \text{ for } k = 1, 2, \dots, L \quad (3)$$

$$f_3: P = \frac{1}{L} \sum_{k=1}^L P_k \quad (4)$$

$$f_4: \frac{A}{P} \quad (5)$$

$$f_5: R = \frac{4A\pi}{P^2} \quad (6)$$

Similar to residual learning, a soft mask unit created  
 as a specific mapping should perform as well as its  
 unattended equivalent. Thus, Attention Module output  
 $R$  is modified as follows:

$$R_{i,c}(y) = (1 + G_{i,c}(Y)) * D_{i,c}(y) \quad (7)$$

$R(y)$  will approximate original features  $D(y)$  if  
 $G(y)$  approximates 0.

### Classification using hybrid GAN+CNN

To classify blood cancer histopathological images into  
 three malignant lymphoma classes—Chronic  
 Lymphocytic Leukemia (CLL), Follicular Lymphoma  
 (FL), and Mantle Cell Lymphoma (MCL)—a hybrid  
 approach using GANs for data augmentation and CNNs

for classification is highly effective, as shown in Figure 3.

The generator generates images from latent space  
 coordinates. Most latent spaces are 100-dimensional  
 vectors. During training, the generator learns to map  
 discrete points to images. After retraining, the model  
 will have enhanced mapping skills. Mastering GANs  
 involves real-time training of both the discriminator  
 and generator networks, which is complex and requires  
 min-max modifications for both components. This  
 procedure is illustrated in Equation (1),

$$\begin{aligned} \min_G \max_D V_{GAN}(D, G) \\ = E_{x \sim p_{data}(x)} [\log D(x)] \\ + E_{z \sim p_z(z)} [\log (1 - D(G(z)))] \end{aligned}$$

Where the expected value of authentic instances is  
 $E_{x \sim p_{data}(x)}$ , whereas the expected value of counterfeit  
 instances is  $E_{z \sim p_z(z)}$ ; the stochastic noise variable for a  
 typical normal distribution is  $p_z(z)$ . The original data  
 is  $x$ , the generator function that maps it to space is  $G(z)$ ,  
 and the probability of the original data  $x$  vs. the  
 generated data is  $D(X)$ . V3 is a convolutional neural  
 network architecture from the Inception family.

### Results

All experiments were conducted using Google  
 Colaboratory, a cloud-based platform that supports  
 Jupyter Notebooks using Python 3.12.0 on Google  
 Compute Engine. Colab Pro offers prolonged runtimes.

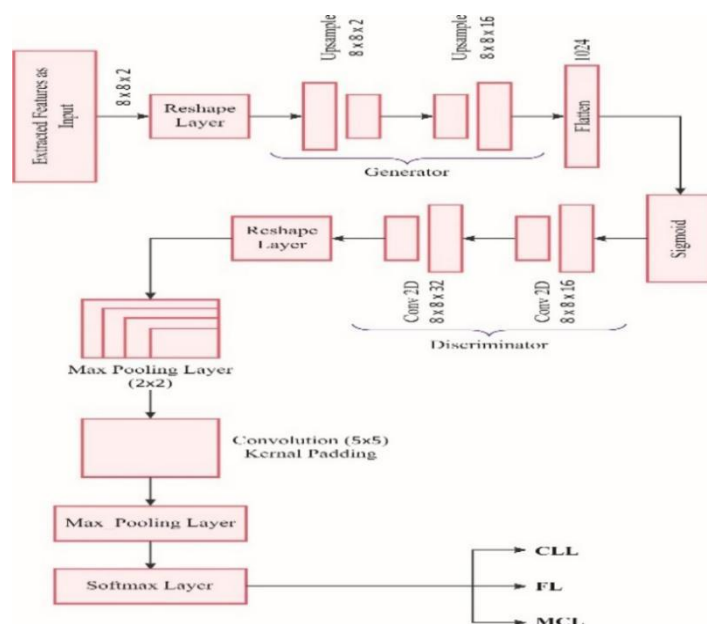


Figure 3. Hybrid GAN\_CNN architecture.

The runtime used a T4 GPU, 15 GB of RAM, 12.7 GB of system RAM, and 201.2 GB of disk space. Colab makes it easy to integrate Google Drive, which was used to store and access the dataset, model weights, and log files. Table 1 shows the parameter settings involved in the proposed model.

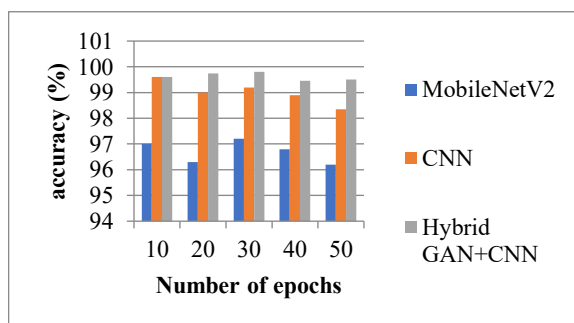
**Table 1.** Parameter settings.

| Method    | Parameter              | Value                      |
|-----------|------------------------|----------------------------|
| ResNet    | Feature levels         | 3                          |
|           | Feature types          | Area, perimeter, roundness |
| Attention | Convolution layers     | 2                          |
|           | Convolution size       | 1 × 1                      |
|           | Activation             | Sigmoid                    |
| GAN       | Latent vector size     | 100                        |
|           | Architecture           | Inception V3               |
| CNN       | Factorized convolution | 7 × 7                      |
|           | Auxiliary classifier   | 1                          |
|           | Label smoothing        | Used                       |
|           | Output classes         | 3                          |
| Training  | Learning Rate          | 0.001                      |
|           | Epochs evaluated       | 10, 20, 30, 40, 50         |

International Journal of  
Medical Toxicology & Forensic Medicine

We evaluated DL models using numerous metrics in these trials. Key performance criteria are precision, accuracy, F1 score, and recall. These parameters are analyzed for the existing MobileNetV2 [15], CNN [16], and the proposed Hybrid GAN+CNN. Here's a breakdown of each assessment measure and related mathematical equation,

Figure 4 indicates the overall comparison of accuracy. As the number of epochs increases from 10 to 50, the proposed hybrid GAN+CNN method achieves the best accuracy of 99.96%, which is 2.96% better than MobileNetV2 and 0.3.5% better than CNN.

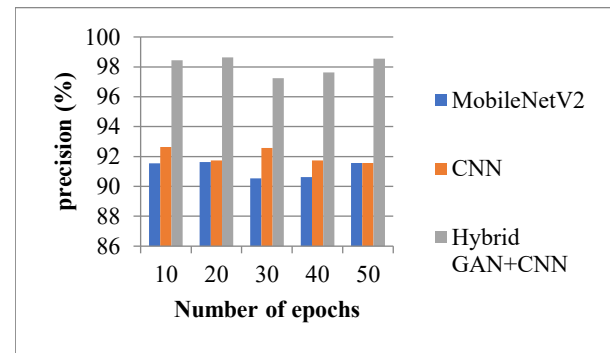


International Journal of  
Medical Toxicology & Forensic Medicine

**Figure 4.** Analysis of accuracy.

Figure 5 shows the overall precision comparison. As the number of epochs increases from 10 to 50, the

proposed hybrid GAN+CNN method achieves the best precision of 91.6%, which is 7.3% better than

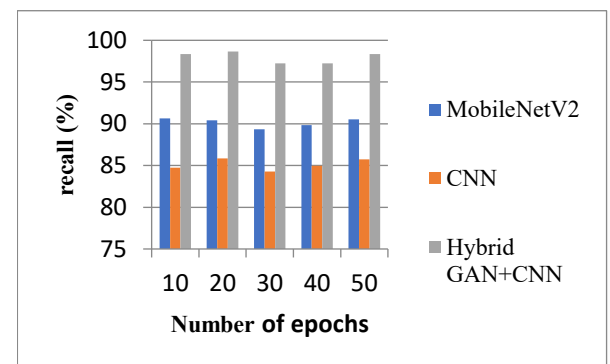


International Journal of  
Medical Toxicology & Forensic Medicine

**Figure 5.** Analysis of precision.

MobileNetV2 and 7.3% better than CNN.

Figure 6 demonstrates the overall comparison of recall. As the number of epochs increases from 10 to 50, the proposed hybrid GAN+CNN method achieves the best recall of 98.4%, which is 8.15% better than

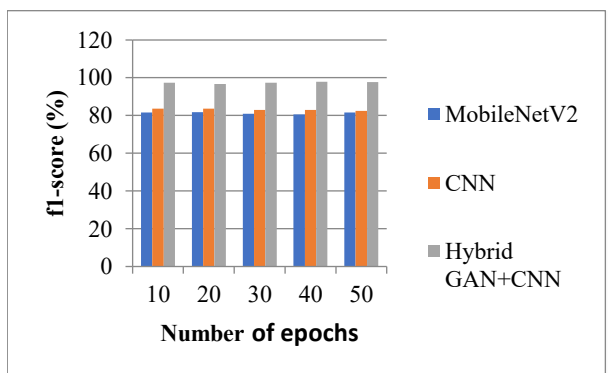


International Journal of  
Medical Toxicology & Forensic Medicine

**Figure 6.** Analysis of recall.

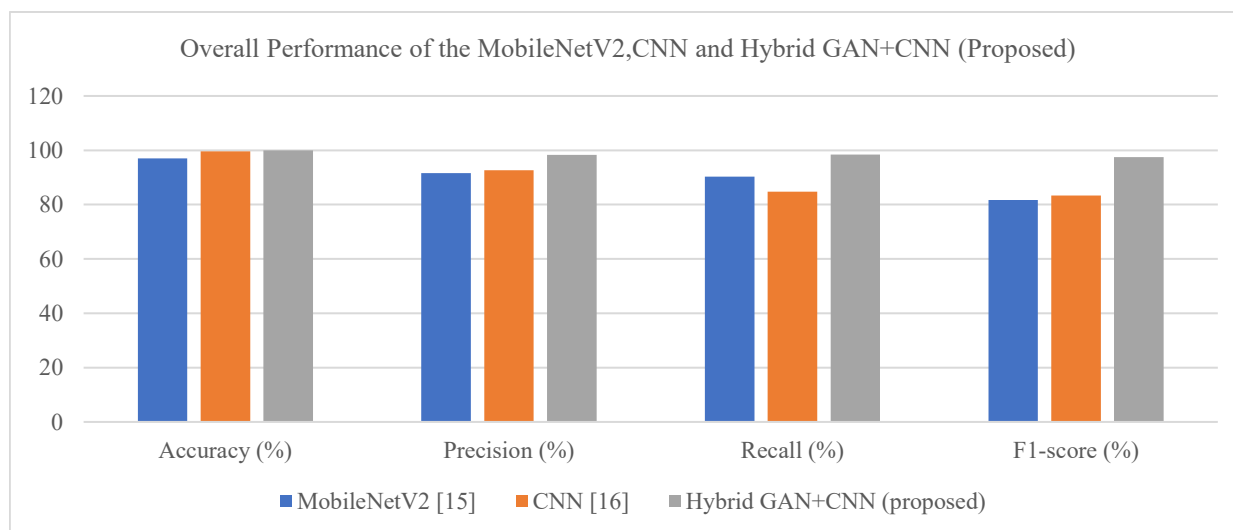
MobileNetV2 and 14.32% better than CNN.

Figure 7 provides an overall comparison of the F1-score. As the number of epochs increases from 10 to 50, the proposed Hybrid GAN+CNN achieves the best F1-score of 97.53%, which is 16.12% better than



International Journal of  
Medical Toxicology & Forensic Medicine

**Figure 7.** Analysis of F1-score.



**Figure 8.** Overall performance of the MobileNetV2, CNN and Hybrid GAN+CNN (proposed).

MobileNetV2 and 14.42% better than CNN.

## Discussion

Figure 8 indicates the Overall performance of MobileNetV2, CNN, and Hybrid GAN+CNN (proposed) and lists the overall performance of the Hybrid GAN+CNN (proposed) method. The proposed method shows a 0.34% increase in accuracy, a 5.23% increase in Precision, an 8.4% increase in Recall, and an 11.23% increase in F1-Score.

Table 2 clearly presents the ablation study analysis of the proposed model in terms of different measures. The experimental results indicate that the proposed Attention-ResNet + GAN + CNN model achieves better performance than the other models. It is also observed that the CNN model achieves lower performance, and the inclusion of GAN and ResNet

along with CNN leads to improved results.

## Conclusion

This study developed attention-mechanism-based feature extraction and hybrid deep learning and image-processing methods to accurately identify and classify leukemia cells. GANs enhance the training dataset by generating synthetic yet realistic images, addressing class imbalance and improving generalization. CNNs, on the other hand, provide powerful feature extraction and classification capabilities, particularly when enhanced with attention mechanisms for multi-level feature representation. The proposed technique may accurately and efficiently identify and categorize leukemia cells, improving diagnosis and therapy. Future research can incorporate multi-modal data such as genetic profiles, clinical data, and other imaging modalities (e.g., MRI or CT scans) to enhance the diagnostic accuracy and provide a holistic understanding of blood cancers.

## Ethical Approval

Not applicable.

## Acknowledgment

None.

## Funding

No external funding was received for this research.

## Conflicts of Interest

**Table 2.** Ablation results of the proposed model.

| Methods                      | Accuracy | Precision | Recall | F1-Score |
|------------------------------|----------|-----------|--------|----------|
| CNN                          | 99.61    | 91.00     | 84.08  | 83.11    |
| ResNet + CNN                 | 99.72    | 93.46     | 90.82  | 88.94    |
| GAN + CNN                    | 99.86    | 96.42     | 95.76  | 94.86    |
| Attention-ResNet + GAN + CNN | 99.96    | 98.30     | 98.40  | 97.53    |

There are no competing interests to declare.

## References

- [1] Sinha K, Uddin Z, Kawsar HI, Islam S, Deen MJ, Howlader MMR. Analyzing chronic disease biomarkers using electrochemical sensors and artificial neural networks. *TrAC Trends Anal Chem.* 2022;158:116861. [DOI: 10.1016/j.trac.2022.116861]
- [2] Billah ME, Javed F. Bayesian convolutional neural network-based models for diagnosis of blood cancer. *Appl Artif Intell.* 2022;36:2011688. [DOI: 10.1080/08839514.2021.2011688]
- [3] Karar ME, Alotaibi B, Alotaibi M. Intelligent medical IoT-enabled automated microscopic image diagnosis of acute blood cancers. *Sensors.* 2022;22(6):2348. [DOI: 10.3390/s22062348]
- [4] Atteia G, Alhussan AA, Samee NA. Bo-allcnn: Bayesian-based optimized CNN for acute lymphoblastic leukemia detection in microscopic blood smear images. *Sensors.* 2022;22(15):5520. [DOI: 10.3390/s22155520]
- [5] Capobianco N, Meignan M, Cottureau AS, Vercellino L, Sibille L, Spottiswoode B, et al. Deep-learning 18F-FDG uptake classification enables total metabolic tumor volume estimation in diffuse large B-cell lymphoma. *J Nucl Med.* 2021;62(1):30-6. [DOI: 10.2967/jnumed.120.242412]
- [6] Patil AM, Patil MD, Birajdar GK. White blood cell image classification using deep learning with canonical correlation analysis. *IRBM.* 2021;42(5):378-89. [DOI: 10.1016/j.irbm.2020.08.005]
- [7] Malagi AV, Kandasamy D, Pushpam D, Khare K, Sharma R, Kumar R, et al. IVIM-DKI with parametric reconstruction method for lymph node evaluation and characterization in lymphoma: A preliminary study comparison with FDG-PET/CT. *Results Eng.* 2023;17:100928. [DOI: 10.1016/j.rineng.2023.100928]
- [8] Hasani N, Paravastu SS, Farhadi F, Yousefirizi F, Morris MA, Rahmim A, et al. Artificial intelligence in lymphoma PET imaging: A scoping review (current trends and future directions). *PET Clin.* 2022;17(2):145-74. [DOI: 10.1016/j.cpet.2021.11.002]
- [9] Steinbuss G, Kriegsmann M, Zgorzelski C, Brobeil A, Goeppert B, Dietrich S, et al. Deep learning for the classification of non-Hodgkin lymphoma on histopathological images. *Cancers.* 2021;13(10):2419. [DOI: 10.3390/cancers13102419]
- [10] Mao B, Zhang L, Ning P, Ding F, Wu F, Lu G, et al. Preoperative prediction for pathological grade of hepatocellular carcinoma via machine learning-based radiomics. *Eur Radiol.* 2020;30(12):6924-32. [DOI: 10.1007/s00330-020-07056-5]
- [11] Cosatto E, Laquerre PF, Malon C, Graf HP, Saito A, Kiyuna T, et al. Automated gastric cancer diagnosis on H&E-stained sections; training a classifier on a large scale with multiple instance machine learning. *Proc SPIE.* 2013;8676:867605. [DOI: 10.1117/12.2007047]
- [12] Komura D, Ishikawa S. Machine learning approaches for pathologic diagnosis. *Virchows Arch.* 2019;475(2):131-8. [DOI: 10.1007/s00428-019-02594-w]
- [13] Jimenez-del-Toro O, Otálora S, Andersson M, Eurén K, Hedlund M, Rousson M, et al. Analysis of histopathology images: From traditional machine learning to deep learning. In: *Biomedical Texture Analysis.* Amsterdam: Elsevier; 2017. p. 281-314. [DOI: 10.1016/B978-0-12-812133-7.00010-7]
- [14] Pasquini G, Arias JER, Schäfer P, Busskamp V. Automated methods for cell type annotation on scRNA-seq data. *Comput Struct Biotechnol J.* 2021;19:961-9. [DOI: 10.1016/j.csbj.2021.01.015]
- [15] Tusar MTHK, Islam MT, Sakil AH, Khandaker MHN, Hossain MM. An Intelligent telediagnosis of acute lymphoblastic leukemia using histopathological deep learning. *J Comput Theory Appl.* 2024;2(1):1-12. [DOI: 10.62411/jcta.10358]
- [16] Saeed U, Kumar K, Khuhro MA, Laghari AA, Shaikh AA, Rai A. DeepLeukNet—A CNN-based microscopy adaptation model for acute lymphoblastic leukemia classification. *Multimed Tools Appl.* 2024;83(7):21019-43. [DOI: 10.1007/s11042-023-16191-2]
- [17] Özgür E, Saygılı A. A new approach for automatic classification of non-Hodgkin lymphoma using deep learning and classical learning methods on histopathological images. *Neural Comput Appl.* 2024;36(32):20537-60. [DOI: 10.1007/s00521-024-09865-9]

- [18] Davamani KA, Jawahar M, Anbarasi LJ, Ravi V, Al Mazroa A, Robin CR. Deep transfer learning technique to detect and classify white blood cells in regular clinical practice using histopathological images. *Multimed Tools Appl.* 2024. [DOI: [10.1007/s11042-024-19133-8](https://doi.org/10.1007/s11042-024-19133-8)]
- [19] Veeraiah N, Alotaibi Y, Subahi AF. Histogram-Based Decision Support System for Extraction and Classification of Leukemia in Blood Smear Images. *Comput Syst Sci Eng.* 2023;46(2):1621-36. [DOI: [10.32604/csse.2023.034658](https://doi.org/10.32604/csse.2023.034658)]
- [20] Elhassan TA, Mohd Rahim MS, Siti Zaiton MH, Swee TT, Alhaj TA, Ali A, et al. Classification of atypical white blood cells in acute myeloid leukemia using a two-stage hybrid model based on deep convolutional autoencoder and deep convolutional neural network. *Diagnostics.* 2023;13(2):196. [DOI: [10.3390/diagnostics13020196](https://doi.org/10.3390/diagnostics13020196)]
- [21] Abunadi I, Senan EM. Multi-method diagnosis of blood microscopic sample for early detection of acute lymphoblastic leukemia based on deep learning and hybrid techniques. *Sensors.* 2022;22(4):1629. [DOI: [10.3390/s22041629](https://doi.org/10.3390/s22041629)]