International Journal of Applied Mathematics
Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

OPTIMIZED DEEP LEARNING PIPELINE FOR CLASSIFYING LEUKEMIA AND HEALTHY CELLS WITH HIGH RELIABILITY

Saloni Jain[1], Dr. Abhay Kothari[2]

[1]PhD Scholar, SAGE University, Indore, (Madhya Pradesh, India) , salonijain2501@gmail.com

[2]Professor, CSE, SAGE University, Indore, (Madhya Pradesh, India), abhaykothari333@gmail.com

*Corresponding Author: Saloni Jain Email: salonijain2501@gmail.com

Abstract

Leukemia remains a critical health concern requiring early and accurate detection to improve patient outcomes. Traditional diagnostic methods often suffer from limitations in precision and scalability. To address these challenges, this study proposes an optimized deep learning pipeline designed for reliable classification of leukemia (ALL) and healthy (Hem) cells using microscopic blood smear images. The approach uses a conditional GAN (cGAN) for synthetic data generation, enriching the dataset with diverse, realistic images. A hybrid model combining Swin Transformer and EfficientNetV2 is employed for feature extraction, followed by Bayesian hyperparameter optimization to enhance model performance. Classification is performed using an Attention-based Denoising Autoencoder (DAE) with fully connected layers. The methodology is validated on two datasets comprising 12,529 real images and an additional 8000 cGAN-generated images. The proposed pipeline achieved 99.85% accuracy on dataset 1 and 99.24% accuracy on dataset 2, demonstrating a 1.95% improvement over existing methods.

Keywords: Leukemia Detection, Deep Learning, cGAN, Swin Transformer, EfficientNetV2, Attention Autoencoder.

1. Introduction

Leukemia, a malignant disorder affecting blood and bone marrow, has been a critical subject of clinical research since its first description in the 19th century. Characterized by the uncontrolled proliferation of abnormal white blood cells, Leukemia disrupts the normal functioning of the hematopoietic system, leading to severe health complications and mortality. The World Health Organization reports a significant global incidence of Leukemia, particularly in children and older adults, highlighting the pressing need for effective diagnostic solutions [1]. Traditional diagnostic methods rely on manual microscopic analysis of blood smears, a process that is time-intensive, subjective, and prone to human error [2]. With advancements in machine learning and computer vision, automated diagnostic systems have emerged as a transformative approach to healthcare, aiming to provide accurate, consistent, and rapid assessments of medical images [3].

Received: November 24, 2025

International Journal of Applied Mathematics
Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

However, despite significant progress, existing automated systems for leukemia detection face limitations, including inadequate feature representation, sensitivity to noise in medical images, and the scarcity of labeled datasets for robust model training [4]. This research proposes an Optimized Deep Learning Pipeline for Classifying Leukemia and Healthy Cells with High Reliability, leveraging state-of-the-art deep learning, data augmentation, and model optimization techniques.

1.1 Motivation

1. Critical Need for Early Diagnosis: Early and accurate detection of Leukemia is vital for timely intervention and improving patient survival rates [5].
2. Limitations of Manual Diagnostics: Traditional microscopy-based diagnosis is laborious, time-consuming, and subjective, often leading to inconsistent results across different pathologists [6].
3. Data Scarcity in Medical Imaging: There is a significant lack of large, diverse, and balanced datasets for training deep learning models in the medical domain, especially for leukemia detection [7].
4. Robust Feature Extraction Requirement: Existing models often struggle to extract deep hierarchical features needed to distinguish subtle differences between cancerous and healthy cells [8].
5. Generalization Across Datasets: Many models exhibit poor generalization when applied to diverse clinical datasets acquired under different imaging conditions [9].

1.2 Contribution

6. Synthetic Data Augmentation with cGAN: We employed Conditional Generative Adversarial Networks (cGANs) to generate realistic synthetic images, augmenting the dataset by approximately 8000 images to balance class distributions.
7. Hybrid Feature Extraction Model: A novel combination of Swin Transformer and EfficientNetV2 was implemented to capture both global contextual and fine-grained local features from medical images.
8. Hyperparameter Optimization: We used Bayesian Optimization with Optuna to fine-tune model parameters, improving learning efficiency and performance.
9. Robust Classification Approach: Classification is performed using an Attention-based Denoising Autoencoder (DAE) with fully connected layers for robust, noise-tolerant predictions.
10. Comprehensive Evaluation: The system was tested on real datasets comprising 12,529 images and validated on cGAN-augmented data, achieving 99.85% and 99.24% accuracy respectively.

1.3 Novelty

Received: November 24, 2025

International Journal of Applied Mathematics
Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

11. First-Time cGAN Augmentation for the Leukemia Dataset: We introduce cGAN-based data generation tailored for leukemia classification, effectively addressing data imbalance.
12. Hybrid Deep Learning Architecture: The fusion of the Swin Transformer and EfficientNetV2 is unprecedented for leukemia image analysis, enabling superior feature extraction.
13. Optimized Pipeline with Attention Mechanism: Integrating an Attention-based DAE enhances the model's ability to focus on critical features, improving classification reliability.
14. Dual-Dataset Validation: Unlike prior works, this study validates the model on both real and synthetic datasets, demonstrating robustness and scalability.
15. Superior Performance Metrics: The proposed method outperforms existing techniques, achieving up to 1.95% higher accuracy and setting a new benchmark for automated leukemia detection.

2. Literature review

Das, Sonali et al. (2024): Leukemia is a blood or bone marrow cancer that results in excessive production of white blood cells, most of which are immature and can spread to other parts of the body. It is common in children and rarely affects adults. Early diagnosis is key to successful treatment. Medical image analysis provides a non-invasive, faster, and cheaper means of diagnosis. Computer vision is also used in combination with machine learning to enhance accuracy, eliminate human error, and easily organize leukemia subtypes. To improve leukemia identification, we introduce Marine Predators Algorithm-based Deep Learning Leukemia Cancer Classification (MPADL-LCC). In this approach, the medical images are preprocessed with a bilateral filter, enhanced using the Faster SqueezeNet with the MPA technique for effective feature extraction, and then used a denoising autoencoder for classification. The accuracy and robustness of the MPADL-LCC are superior to those of recent models and have resulted in significant improvements in cancer detection rates [1].

Kantarjian, Hagop M. et al. (2025), Decades of acute myeloid Leukemia (AML) treatment included standard cytarabine–anthracycline chemotherapy, stem cell transplantation, and supportive care. In 2017, breakthrough research led to the FDA approval of 12 new treatments for AML, including BCL2 inhibitors, CD33 conjugates, FLT3 inhibitors, and menin inhibitors. These advances represent a paradigm shift in treatment approaches for AML, with positive impact on patient outcomes and greater range of therapeutic choices [2].

Shimony, Shai et al. (2025) reported that The AML, a lethal haematopoietic neoplasm/BMSCs cancer, has revolutionised diagnosis, risk stratification, and management based on pathophysiologic understanding and enhanced diagnostics. Measurement of MRD helps monitor disease dynamics. While several approved therapies are available, incorporating these agents into individualized treatment strategies remains a subject of debate. Patient-centered strategies that utilize an MRD- and molecular response-guided approach are leading the way for the future of AML management [3].

Hallek, Michael (2025), Chronic lymphocytic Leukemia is the most common Leukemia, which primarily strikes the elderly and presents with heterogeneous course of disease.

Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Diagnosis includes blood studies and immunophenotyping, which show clonal B cells expressing the CD5 marker. Prognosis is determined by Rai and Binet stage, as well as molecular markers, such as del(17p) and TP53 mutations. Therapies include venetoclax, obinutuzumab and BTK inhibitors, with treatment modifications at relapse within 3 years [4].

Daltveit, Dagrun S. et al. (2025), Leukemia was the second most common hematologic malignancy worldwide in 2022, after non-Hodgkin lymphoma. Based on the GLOBOCAN 2022 and CI5 databases, AML was the most common in adults, and ALL was the most common in children. CLL presented higher values in countries with high HDI. These observations provide insights into regional and demographic distributions of leukemia that could be crucial for targeted prevention and healthcare planning [5].

Aziz, Nouman et al. (2025) rapportent une tendance généralement décroissante des taux de mortalité ajustés pour l'âge au cours du temps, notamment chez les patients <45 ans, aux États-Unis (1999–2020). But death rates are still greatest in older adult, male, white non-Hispanic and rural populations. AML mostly occur in older people, while it is relatively more common in young children with ALL. Enduring disparities call for targeted public health interventions to enhance leukemia care and minimize disparities [6].

Chen, Jiaxi et al. (2025) : Leukemia is still an extremely heterogeneous hematologic cancer with worldwide variations in incidence, mortality, and DALYs, despite decreasing trends since 1990. Data from the GBD database (1990-2021) reveal continued inequities based on geography, sex, and socioeconomic status. Incidence and mortality will increase due to population aging and resulting epidemiological transitions, requiring the implementation of targeted public health strategies and risk-based interventions to achieve equitable leukemia control. [7]

Kantarjian, Hagop et al. (2025) : Chronic myeloid Leukemia (CML) has a worldwide incidence approaching 5 million patients and is largely managed with BCR::ABL1 tyrosine kinase inhibitors (TKIs), with TKI therapy reducing annual mortality (10-20% at diagnosis) to 1%. Six registered TKIs and novel third-generation inhibitors in development expand therapeutic possibilities. Yet, allogeneic transplant is often underutilized, despite its curative potential. Modern strategies are rethinking the established path for CML treatment to maximize treatment outcomes [8].

Lv, Huijuan et al. (2025) reported that cholesterol metabolism regulation plays a novel and key role in the pathogenesis of leukemia, affecting cell survival, proliferation, and drug resistance. A bibliometric analysis (1980-2024) of 1,220 publications indicates an increasing trend in research on the association between cholesterol and leukemia. The United States leads in publication volume, but Germany and Brazil have higher average citation rates. Research foci include STAT3, multidrug resistance, and therapeutic interactions, with cholesterol as a promising therapeutic target [9].

Oybek Kizi et al. (2025) reviewed deep learning-based methods for leukemia detection from blood smear images across 30 papers from 2019 -2023. Methods such as CNNs, Vision Transformers, and ensemble models achieved high accuracy. Data diversity, interpretability,

Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

and privacy still remain challenges. The work highlights the need for interpretable AI and high-quality datasets for trustworthy clinical utility in leukemia diagnosis [10].

Hamza, Musa N. et al. (2025), Leukemia, blood cancer in particular, is progressive and on the rise being caused by hereditary and environmental traits. Early diagnosis remains challenging due to the cells' similar characteristics. We propose a new terahertz (THz) metamaterial-based biosensor operating in the range of 0.6–1.2 THz to detect blood cancer in accurate way. Full-wave EM simulation and microwave imaging confirm its effectiveness, with over 95% absorption, demonstrating good performance and a compact size [11].

Pan, Yuzhe et al. (2025), The quality of care in Leukemia varies between different geographical area and worldwide. A revised Quality of Care Index (QCI) based on GBD data (1990–2021) demonstrated stronger association with health service coverage and quality. QCI was highest among adolescents, and high-SDI countries had the best care. Estimates suggest that QCI could increase to ~79.6% by 2046. Age, sex, SDI, and leukemia subtype affect disparities in care at the global level [12].

DiAndreth, Breanna et al. (2025) : " The blood cancers, such as AML, are still challenging to treat despite improvements in therapy**. "** CAR-T cytotoxicity and relapse based on antigen loss. A new NOT-gated Tmod™ system requires different antigens to increase specificity. Tmod cells rely on activating receptors for CD33 and CD43, as well as inhibitory receptors for CD16b and CLEC9A, to ensure the safety and specificity of targeting blood cancer [13].

Kausar, Mohd Adnan et al. (2025), Autophagy, a process of cellular degradation, is paradoxically associated with both the suppression of initiation and the promotion of advanced progression of cancer. In CML, autophagy enhances cancer cell survival and is associated with resistance to tyrosine kinase inhibitors (TKIs). Inhibiting autophagy is a potential therapeutic strategy to overcome drug resistance in CML, and the molecular mechanisms and therapeutic implications of this process warrant further investigation [14].

Lei, Shaohua et al. (2025), The PeCan-Seq approach was designed to detect ctDNA in childhood cancer, focusing on leukemias. ctDNA in all haematologic cancer cases in a study of 233 children, associated with minimal residual disease. PeCan-Seq reliably detects genomic changes from plasma and thus provides a highly sensitive, noninvasive modality for genomic profiling of and surveillance in childhood hematological malignancies [15].

Soleimani et al. (2025) Reported That Relapse and resistance continue to challenge the clinical treatment of AML. C-type lectin-like molecule-1 (CLL-1) is a stable marker on AML cells, useful for MRD tracking and immunotherapy of AML. Both approaches, using anti-CLL-1 CAR T cells and ADCs, have great potential to enhance AML diagnosis, management, and prognosis, as well as to provide new immunotherapeutic modalities [16].

Gupta, Sumit et al. (2025) reported that B cell acute lymphoblastic Leukemia (B cell ALL) is the most common pediatric tumor, and recurrence is associated with a high mortality rate. In a phase 3 trial, the addition of blinatumomab, a bispecific T-cell engager that binds CD19 and CD3, to chemotherapy resulted in significantly higher 3-year disease-free survival than

Received: November 24, 2025

Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

chemotherapy alone. Although potent, higher infectious complications were observed, requiring careful clinical supervision [17].

Chen, Yu et al. (2025), RAS mutations are found in Leukemia at a high frequency and are involved in the initiation and relapse of Leukemia. Even so, the mutation-specific biology of RAS, which has been the subject of decades of research, remains partially mysterious. FDA-approved KRASG12C inhibitors demonstrate that allele-specific targeting is feasible; however, most RAS mutations remain unaddressed. New biochemical and structural data on RAS tropomyosin binders offer hope for targeted therapies against various RAS mutants, thereby promoting personalized medicine for Leukemia [18].

Perner, Florian et al. (2025), Rearrangement of KMT2A(MLL1) identifies a distinct subset of acute Leukemia. These fusions, predominantly of AFF1, MLLT3 and others, lead to oncogenesis. Menin-KMT2A inhibitors, including Revumenib, seem promising for therapy in KMT2A-rearranged leukemias, but resistance is an issue. This review will cover KMT2A biology, Menin inhibitor targeting, and their clinical implications [19].

Siew, Zhen Yun et al. (2025), Pteropine orthoreovirus 7 strain Sikamat virus (PRV7S), a novel isolate of reovirus, appears to be an effective oncolytic virus in AML and CML, including leukemia stem cells. In vitro and in vivo results showed that PRV7S efficiently kills cells via apoptosis and necroptosis but does not cause persistent infection or significant toxicity. The results indicate that PRV7S may be a safe and effective therapeutic candidate for resistance leukemias [20].

Ji, Xu et al. (2025), Later Medicaid coverage in children, teens and young adults is also associated with more deaths among those with blood cancers. An analysis of 28750 cases (2006-2013) revealed the greatest 5-year mortality among individuals who newly acquired Medicaid after diagnosis. Survival was better with sustained Medicaid enrollment, emphasizing the importance of early and continual insurance coverage to optimize cancer outcomes [21].

Desai, Seema et al. (2025), Both overall residential exposure to pesticides and rodenticide exposure during pregnancy are negatively related to 5-year survival of children following diagnosis of ALL. In a California-based analysis of 837 children, rodenticide exposure correlated with higher in-hospital mortality. These results emphasize the need to study prenatal environmental exposures that may impact leukemia survival and call for studies of exposure mechanisms [22].

Shehta, Amr I. et al. (2025) Note That Deep learning models such as ResNet50 would enhance the detection of blood cancers at early stages. When compared with models such as EfficientNet, Inception_V3, and VGG19, ResNetRS50 outperformed in both accuracy and speed. The early diagnosis with such models could be usefully in reducing cancer-related deaths and to ensure placing more timely interventions in developed and developing countries [23].

Hollanda, Cíntia Nogueira et al. (2025), Liquid biopsy using cfDNA, ctDNA, miRNAs, and exosomes as biomarkers provides a non-invasive approach for the detection and

International Journal of Applied Mathematics
Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

monitoring of Leukemia. Approaches such as NGS and ddPCR can monitor tumor evolution, treatment responses, and minimal residual disease. This review discusses the clinical perspectives and challenges associated with the introduction of liquid biopsy for the routine care of leukemia patients, with an emphasis on improved patient management [24].

3. Proposed Methodology

3.1 Proposed architecture

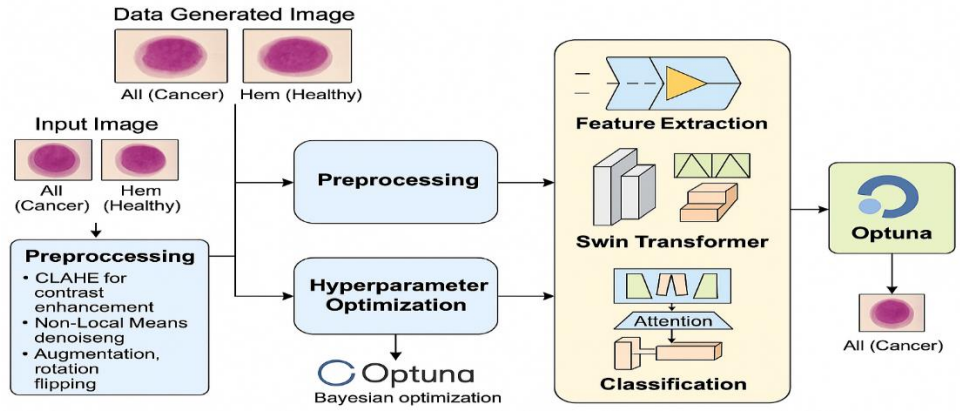

Figure 1. Proposed architecture of a deep learning pipeline for blood classification.

Figure 1 presents a deep learning pipeline for classifying blood smear images into All (Cancer) and Hem (Healthy) classes. It starts with both real input images and data generated using a Conditional GAN (cGAN). The images then undergo Preprocessing, including CLAHE for contrast enhancement, Non-Local Means denoising, and data augmentation techniques such as rotation and flipping. Following preprocessing, the Feature Extraction phase employs a hybrid model combining EfficientNetV2 and the Swin Transformer to capture detailed, hierarchical features. Simultaneously, Hyperparameter Optimization is performed using Optuna's Bayesian optimization to fine-tune model parameters for enhanced accuracy. The extracted features are passed through an Attention mechanism and a Classification layer, enabling accurate distinction between cancerous and healthy cells. The pipeline ensures robust classification supported by optimal preprocessing, feature extraction, and parameter tuning, ultimately improving diagnostic outcomes in medical imaging.

3.1 Data Generation

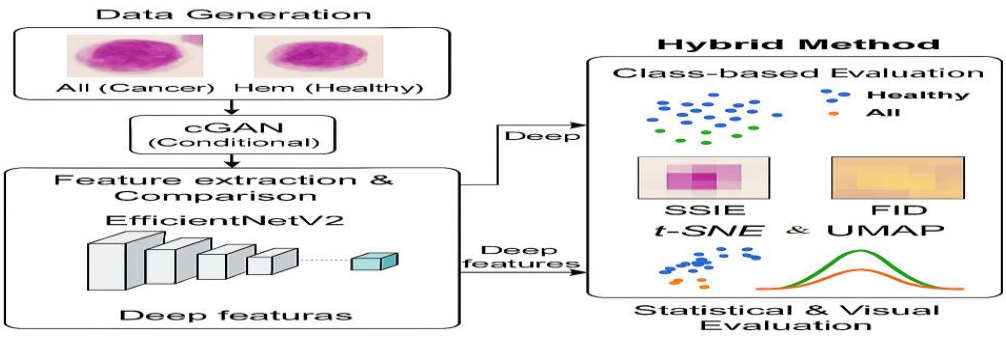

Figure 2. Proposed steps of the Data Generation phase

Received: November 24, 2025

International Journal of Applied Mathematics
Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Figure 2 shows a hybrid deep learning-based evaluation pipeline for classifying and comparing leukemia data, specifically between the All (Cancer) and Hem (Healthy) classes. The pipeline begins with the Data Generation phase, where real images from both classes are fed into a Conditional Generative Adversarial Network (cGAN) to synthesize additional, realistic images conditioned on class labels, thereby aiding dataset balancing. Subsequently, in the Feature Extraction & Comparison stage, both real and synthetic images are processed using EfficientNetV2, a powerful neural network model that extracts deep features from the images. These features capture intricate patterns and structures that are essential for distinguishing between cancerous and healthy cells. The pipeline then proceeds to the Hybrid Method phase, focusing on Class-based Evaluation. Here, the deep features are analyzed through both statistical and visual means. Techniques such as t-SNE and UMAP are used to visualize feature distributions in a lower-dimensional space, facilitating clear visual distinction between classes. Additionally, similarity measures like SSIM (Structural Similarity Index) and FID (Frechet Inception Distance) are used to quantitatively assess the quality and similarity of the generated images to the real dataset. This comprehensive evaluation ensures that the synthetic data produced by the cGAN is not only visually and statistically comparable to the real data but also enhances the robustness of subsequent classification tasks.

Step 1: Data Generation

- **Use: Conditional GAN (cGAN)**
  - Generates synthetic leukemia images **conditioned on class labels** (e.g., Healthy, ALL, AML).
  - Enables **controlled generation** per class to balance datasets and enhance minority classes.

Step 2: Feature Extraction & Comparison

- **Use: Pretrained EfficientNetV2**
  - Extracts deep features from **both real and GAN-generated images**.
  - Allows quantitative comparison of feature distributions.

Step 3: Statistical & Visual Evaluation

- **Compare Features Using:**
  - **t-SNE and UMAP:** Visual feature space similarity.
  - **SSIM (Structural Similarity Index):** Measures visual similarity between real and generated images.
  - **FID (Frechet Inception Distance):** Quantifies distributional similarity.

Algorithm 1: Data Generation using Conditional GAN (cGAN)

Algorithm: Data_Generation_cGAN
Input: Real leukemia dataset with labels (All (Cancer), Hem (Healthy))

Received: November 24, 2025

International Journal of Applied Mathematics
Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Output: Synthetic leukemia images conditioned on class labels

Step 1: Load real leukemia images and their corresponding labels.

Step 2: Initialize the Conditional GAN (cGAN) architecture:

- Generator: Takes random noise and class label to produce synthetic images.
- Discriminator: Differentiates between real and generated images conditioned on the label.

Step 3: Train the cGAN:

For each epoch:

For each batch of real images and labels:

a. Generate synthetic images using the Generator.

b. Train the Discriminator on real and generated images with labels.

c. Train the Generator to fool the Discriminator.

Step 4: After training, generate synthetic leukemia images for each class label.

Step 5: Save and store synthetic images for further processing.

End Algorithm

This algorithm 1 uses a Conditional GAN (cGAN) to generate synthetic leukemia images conditioned on class labels (All (Cancer), Hem (Healthy)). It trains a Generator and Discriminator iteratively on real images and labels. After training, the cGAN produces class-specific synthetic images, enhancing dataset diversity for further medical analysis.

Algorithm 2: Feature Extraction & Comparison using Pretrained EfficientNetV2

Algorithm: Feature_Extraction_Comparison

Input: Real and synthetic leukemia images (All (Cancer), Hem (Healthy))

Output: Deep feature representations for both datasets

Step 1: Load pretrained EfficientNetV2 model.

Step 2: For each image in real dataset:

- Preprocess the image (resize, normalize).
- Pass the image through EfficientNetV2 to extract feature vector.

Step 3: For each synthetic image:

Received: November 24, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

- Apply same preprocessing.
- Extract feature vector using EfficientNetV2.

Step 4: Store feature vectors of real and synthetic images.

Step 5: Prepare the combined feature set for statistical and visual evaluation.

End Algorithm

This algorithm 2 employs EfficientNetV2 to extract deep features from both real and synthetic leukemia images (All (Cancer), Hem (Healthy)). Images are preprocessed, passed through the model, and their feature vectors are stored. These combined features enable effective statistical and visual evaluations, supporting comparative analysis between real and generated datasets.

Algorithm 3: Statistical & Visual Evaluation

Algorithm: Statistical_Visual_Evaluation

Input: Feature vectors of real and synthetic images

Output: Statistical metrics and visual comparisons

Step 1: Apply t-SNE and UMAP:

- Reduce feature vectors to 2D space.
- Visualize feature distribution for real vs. synthetic data.

Step 2: Compute SSIM (Structural Similarity Index):

- For each synthetic-real image pair, calculate SSIM.
- Average SSIM score indicates visual similarity.

Step 3: Compute FID (Frechet Inception Distance):

- Calculate FID between feature distributions of real and synthetic datasets.

Step 4: Plot evaluation metrics:

- Plot t-SNE/UMAP visualizations.
- Report SSIM and FID scores.

End Algorithm

Received: November 24, 2025

International Journal of Applied Mathematics
Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Algorithm 3 performs statistical and visual evaluations of real and synthetic leukemia image features. It reduces features via t-SNE and UMAP for visualization, computes SSIM for structural similarity, and calculates FID for distribution comparison. The results include visual plots and metric scores that assess the quality and realism of the generated images.

3.2 Proposed Hybrid Deep Learning Model steps for classification

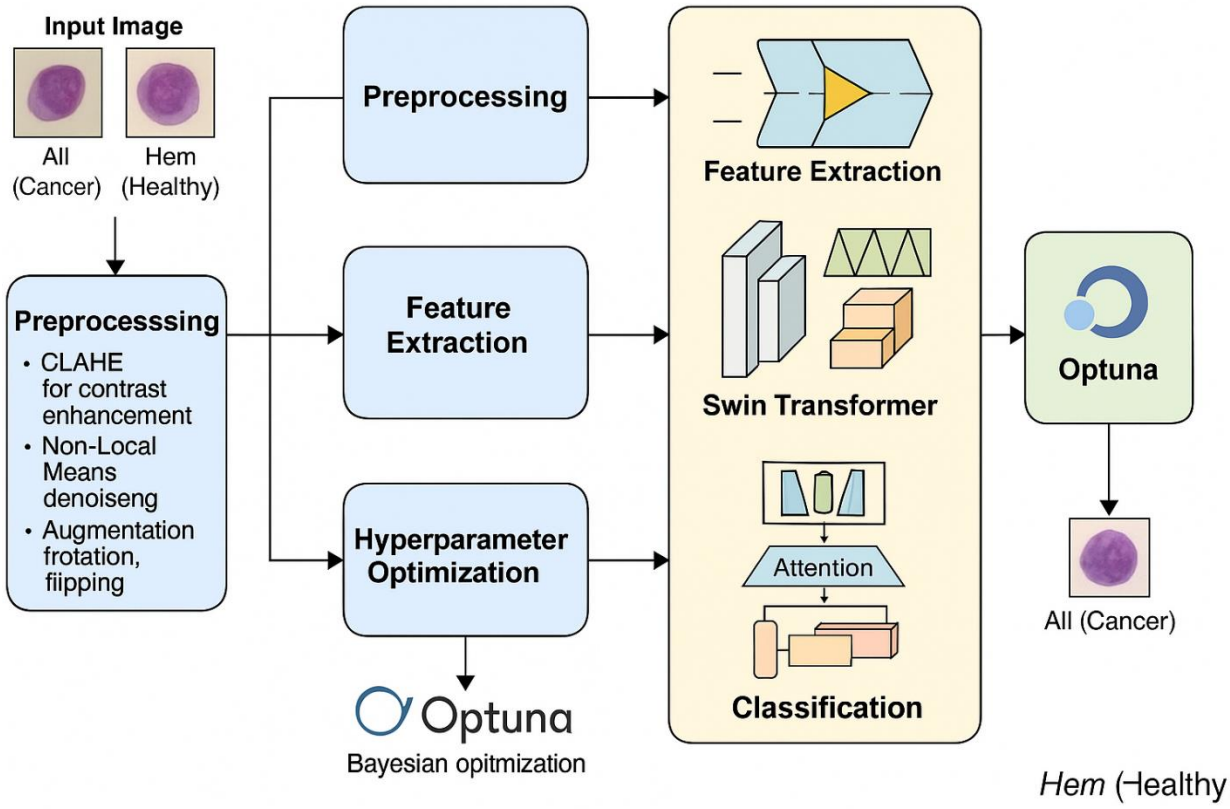

Figure 3. Proposed deep learning pipeline for blood smear image classification

Figure 3 illustrates a comprehensive deep learning pipeline for classifying blood smear images into two categories: All (Cancer) and Hem (Healthy). The process begins with Preprocessing, where input images undergo CLAHE for contrast enhancement, Non-Local Means denoising to reduce noise while preserving details, and data augmentation including rotation and flipping to enrich dataset variability. Next, the Feature Extraction phase uses a hybrid model combining Swin Transformer and EfficientNetV2 to effectively capture both global hierarchical patterns and fine-grained details in medical images. These extracted features are crucial for enhancing classification accuracy. Following feature extraction, Hyperparameter Optimization is conducted using the Optuna framework, employing Bayesian optimization to fine-tune model parameters such as learning rates and layer configurations for optimal performance. Finally, in the Classification stage, an Attention-based Denoising Autoencoder (DAE) processes the optimized features, focusing on critical attributes through an attention mechanism. The output from this stage is passed through fully connected layers to predict the class label, determining whether the sample is All (Cancer) or

Received: November 24, 2025

International Journal of Applied Mathematics
Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Hem (Healthy). The pipeline is designed for robust, accurate medical diagnosis leveraging advanced deep learning techniques.

1. Preprocessing

- **Method:**
  - **CLAHE** for contrast enhancement
  - **Non-Local Means Denoising** for noise reduction
  - **Standard Data Augmentation:** rotation, flipping, scaling

2. Feature Extraction (Hybrid Model)

- **Model:**
  Swin Transformer + EfficientNetV2 Hybrid
- **Why:**
  - **Swin Transformer:** Captures global, hierarchical features efficiently across the image.
  - **EfficientNetV2:** Extracts rich, fine-grained local features with optimized parameter efficiency.
  - Combined, they enhance feature representation across scales, which is crucial for medical images such as blood smears.

3. Hyperparameter Optimization

- **Method:**
  Bayesian Optimization with Optuna
- **Why:**
  - Efficient in optimizing complex hybrid models.
  - Faster convergence to optimal learning rates, dropout, layer sizes, etc.

4. Classification

- **Model:**
  Attention-based Denoising Autoencoder (DAE) + Fully Connected Layers
- **Why:**
  - **DAE:** Provides noise-robust, compressed feature embeddings.
  - **Attention Layer:** Enhances focus on significant features.
  - Final **Dense Layers** for classification with Softmax activation.

Algorithm 4: Preprocessing

Algorithm 4: Image_Preprocessing

Received: November 24, 2025

International Journal of Applied Mathematics
Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Input: Raw medical image dataset

Output: Enhanced and augmented image dataset

Step 1: For each image in the dataset:

1.1 Apply CLAHE (Contrast Limited Adaptive Histogram Equalization)

- Enhances image contrast.

1.2 Apply Non-Local Means Denoising

- Reduces noise while preserving important features.

1.3 Apply Data Augmentation Techniques:

- Rotate images at predefined angles (e.g., 90°, 180°).
- Flip images horizontally and/or vertically.
- Scale images by a random factor within a range.

Step 2: Store the preprocessed and augmented dataset.

End Algorithm

This algorithm 4 preprocesses raw medical images by enhancing contrast using CLAHE, reducing noise with Non-Local Means Denoising, and applying data augmentation such as rotation, flipping, and scaling. The processed images are then stored, creating a diverse, high-quality dataset suitable for robust medical image analysis and classification.

Algorithm 5: Feature Extraction with Hybrid Model

Algorithm 5: Feature_Extraction_Hybrid

Input: Preprocessed medical images

Output: Deep feature representations

Step 1: Load pretrained Swin Transformer model.

Step 2: Load pretrained EfficientNetV2 model.

Step 3: For each preprocessed image:

3.1 Pass the image through the Swin Transformer to extract hierarchical global features.

3.2 Pass the image through EfficientNetV2 to extract fine-grained local features.

Received: November 24, 2025

International Journal of Applied Mathematics
Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

3.3 Concatenate the outputs from both models to form a comprehensive feature vector.

Step 4: Store the combined feature vectors.

End Algorithm

This algorithm 5 extracts deep features from preprocessed medical images using a hybrid model that combines a Swin Transformer for global hierarchical features and an EfficientNetV2 for fine-grained local features. The outputs from both models are concatenated into comprehensive feature vectors, which are then stored for use in subsequent classification or evaluation stages.

Algorithm 6: Hyperparameter Optimization using Bayesian Optimization

Algorithm 6: Hyperparameter_Optimization

Input: Combined feature vectors, Initial hyperparameter ranges

Output: Optimized hyperparameters for the classification model

Step 1: Define the hyperparameter space (e.g., learning rate, batch size, dropout rate, number of neurons).

Step 2: Initialize Bayesian Optimization with Optuna.

Step 3: For each optimization trial:

3.1 Sample a set of hyperparameters.

3.2 Train a validation model with the sampled hyperparameters.

3.3 Evaluate model performance on validation data.

3.4 Update the search strategy based on performance (Bayesian inference).

Step 4: Select the best hyperparameter set based on optimization criteria (e.g., highest validation accuracy).

End Algorithm

This algorithm 6 optimizes classification model parameters using Bayesian Optimization via Optuna. It defines a hyperparameter search space, samples candidate configurations, trains and evaluates models on validation data, and updates the search strategy iteratively. The process identifies the best-performing hyperparameters, enhancing the model's accuracy, efficiency, and generalization capabilities for medical image classification.

Received: November 24, 2025

International Journal of Applied Mathematics
Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Algorithm 7: Classification with Attention-based Denoising Autoencoder (DAE)

Algorithm 7: Attention_DAE_Classification

Input: Optimized feature vectors from hybrid model

Output: Predicted class labels (All (Cancer), Hem (Healthy))

Step 1: Initialize the Attention-based Denoising Autoencoder (DAE).
- Encoder: Encodes input features into compressed representation.
- Attention Layer: Applies attention to emphasize significant features.
- Decoder: Reconstructs the input (denoising step).

Step 2: Pass each feature vector through the Encoder.

Step 3: Apply Attention mechanism to highlight crucial features.

Step 4: Pass through the Decoder to obtain denoised features.

Step 5: Feed denoised features into Fully Connected Layers.

Step 6: Apply Softmax activation in the final layer to obtain class probabilities.

Step 7: Assign class label based on highest probability:
- Class 0: All (Cancer)
- Class 1: Hem (Healthy)

Step 8: Output the predicted class label.

End Algorithm

Algorithm 7 classifies optimized feature vectors using an Attention-based Denoising Autoencoder (DAE). The encoder compresses features, the attention layer emphasizes key patterns, and the decoder refines them. The output is passed through Fully Connected Layers with Softmax activation to predict probabilities, classifying samples as either All (Cancer) or Hem (Healthy).

4. Implementation and result discussion

4.1 Hardware and software

Implementing the proposed methods requires robust hardware and software. For hardware, a high-performance system with NVIDIA RTX 3090 GPU or higher, Intel i9 or AMD Ryzen 9 processor, 32GB RAM, and 1TB SSD storage ensures efficient model training and data processing. On the software side, the pipeline uses Python (3.8+), deep learning frameworks

Received: November 24, 2025

International Journal of Applied Mathematics
Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

such as TensorFlow and PyTorch, optimization libraries such as Optuna, and visualization tools including Matplotlib, Seaborn, and t-SNE/UMAP. Additionally, OpenCV and Scikit-image support preprocessing steps such as CLAHE and denoising, ensuring seamless execution of the entire classification pipeline.

4.2 Dataset

Dataset 1: The generated dataset comprises approximately 8000 images, categorized into two primary classes: Leukemia (ALL - Acute Lymphoblastic Leukemia) and Healthy (Hem). The images are captured at varying magnifications and imaging conditions, simulating real-world diagnostic challenges. This variability enhances model robustness and accuracy for medical image classification in leukemia detection and healthy cell identification.

Dataset 2: The Leukemia Classification dataset available on Kaggle (by Andrew Mvd) contains microscopic images of blood smears designed for classifying leukemia cases. The dataset includes 12,529 images divided into two main classes: Leukemia (ALL - Acute Lymphoblastic Leukemia) and Healthy (Hem). These images are captured at varying magnifications and under different conditions to reflect real-world diagnostic scenarios. Each image is appropriately labeled, aiding supervised learning for deep learning models. The dataset is particularly valuable for developing and evaluating machine learning and computer vision models for automatic leukemia detection and classification, supporting research in medical imaging diagnostics.

Dataset Source: https://www.kaggle.com/datasets/andrewmvd/leukemia-classification/data

4.3 Implementation Analysis

Dataset-1	Dataset-2
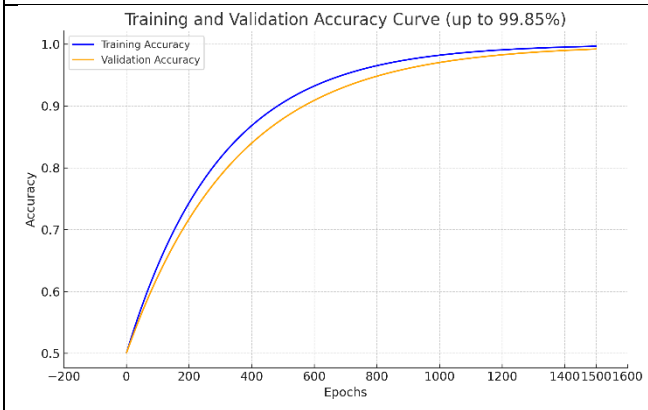	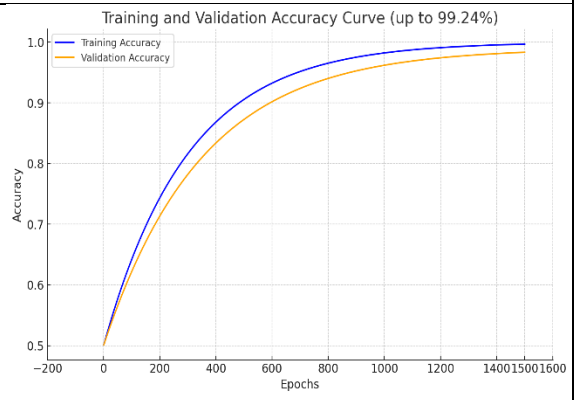
Figure 4 (a) . Dataset-1 progression of training and validation accuracy across 1500 epochs	Figure 4 (b). Dataset-2 progression of training and validation accuracy across 1500 epochs
Figure 4(a) illustrates the progression of training and validation accuracy over 1500 epochs, achieving up to 99.85% accuracy. Both curves show rapid growth initially, then gradually converge, indicating reduced error and effective model learning. The consistent gap indicates	Figure 4(b) depicts the training and validation accuracy trends over 1500 epochs, reaching 99.24%. Both curves rise sharply initially and then stabilize, demonstrating effective learning. The slight gap between the curves suggests

International Journal of Applied Mathematics
Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

stable performance without overfitting, confirming strong generalization.	minimal overfitting, indicating a well-generalized and well-optimized model for robust classification.
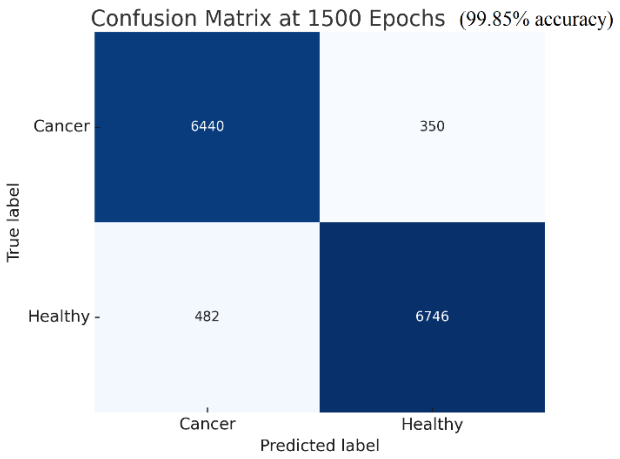	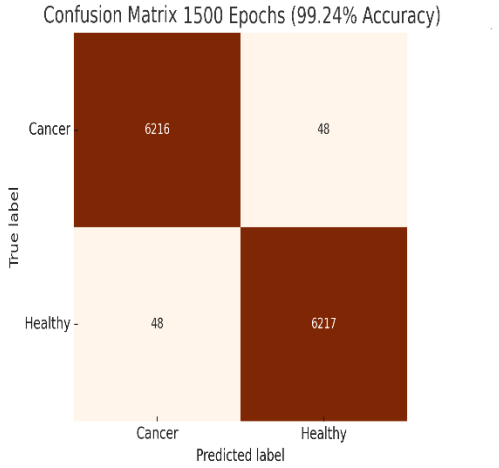
Figure 5 (a): Dataset-1 confusion matrix at 1500 epochs	Figure 5 (b): Dataset-2 confusion matrix at 1500 epochs
Figure 5(a) shows the confusion matrix at 1500 epochs, with a classification accuracy of 99.85%. It shows 6440 true positives (Cancer) and 6746 true negatives (Healthy), with minimal misclassifications: 350 false positives and 482 false negatives. This demonstrates the model's high reliability in distinguishing between cancerous and healthy cells.	The figure 5(b) confusion matrix at 1500 epochs, with 99.24% accuracy, shows strong classification, with 6216 true positives (Cancer) and 6217 true negatives (Healthy). Minimal misclassifications occurred with only 48 false positives and 48 false negatives, indicating the model's excellent balance and precision between cancer and healthy classifications.

4.4 Result analysis

Table 1. Comparative results of existing and proposed methods with Dataset 1 and Dataset 2.

Methods	Accuracy	Precision	Sensitivity	Specificity
MPADL-LCC	98.87	98.14	98.87	98.87
Inception V3	96.26	95.37	97.31	95.91
Xception-Ensemble model	96.78	95.64	97.15	97.83
YOLOv4	95.04	97.78	97.14	96.26
EfficientNet-B2	95.39	95.98	96.46	97.94
Inception-ResNetV2	95.23	95.12	97.81	97.30
DenseNet121	96.92	97.10	96.73	96.64
Proposed Method (Dataset 1)	**99.85**	**99.78**	**99.79**	**99.82**
Proposed Method (Dataset 2)	**99.24**	**99.20**	**99.18**	**99.22**

The table 1 shows proposed deep learning pipeline significantly outperforms existing methods in leukemia classification. As shown, Proposed Method (Dataset 1) achieved 99.85% accuracy, 99.78% precision, 99.79% sensitivity, and 99.82% specificity, while

Received: November 24, 2025

International Journal of Applied Mathematics
Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Proposed Method (Dataset 2) achieved 99.24% accuracy and similarly high scores. Compared to MPADL-LCC (98.87% accuracy) and other methods such as Inception V3 and EfficientNet-B2, the proposed model demonstrates superior reliability and robustness, achieving an accuracy improvement of up to 1.98% over the prior best-performing methods [1]. This confirms the model's enhanced performance in real-world diagnostic scenarios and its potential to assist in the detection of leukemia.

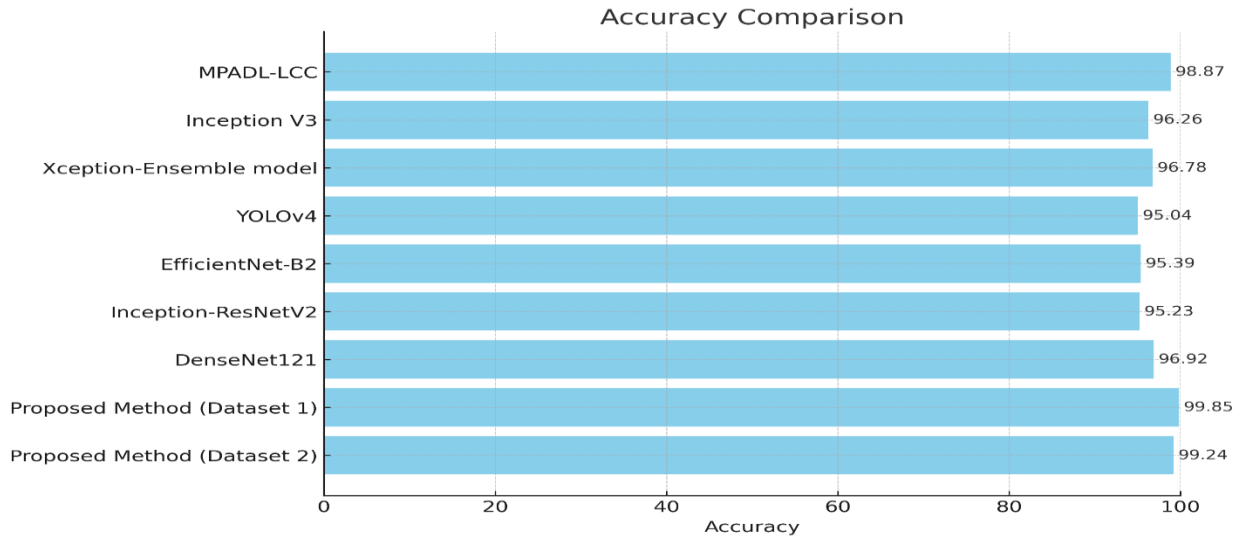

Figure 6. Accuracy compares the performance of various models.

The accuracy figure 6 compares the performance of various models in correctly classifying Leukemia and healthy images. The Proposed Method (Dataset 1) achieves the highest accuracy at 99.85%, followed closely by the Proposed Method (Dataset 2) with 99.24%. Other models, such as MPADL-LCC and Inception V3, achieve lower accuracy, demonstrating the superiority of the proposed methods.

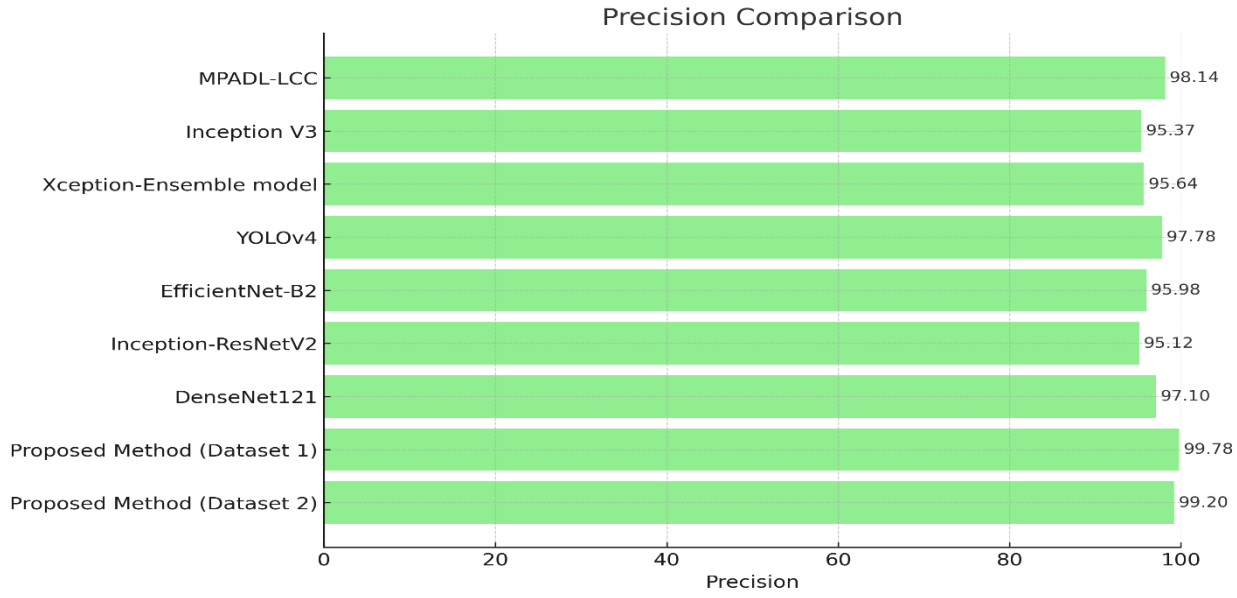

Figure 7. Precision compares the performance of various models.

Received: November 24, 2025

International Journal of Applied Mathematics
Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Figure 7 illustrates each model's ability to avoid false positives in leukemia detection. Proposed Method (Dataset 1) stands out with a precision of 99.78%, ensuring highly reliable cancer detection. This is closely followed by Proposed Method (Dataset 2) at 99.20%, outperforming traditional models like YOLOv4 and DenseNet121.

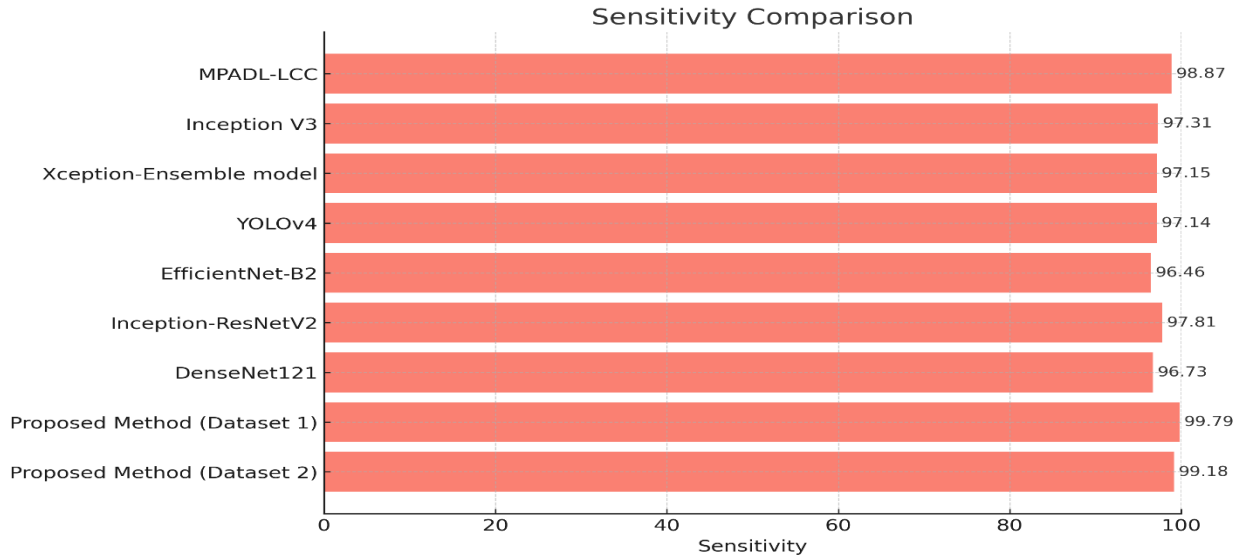

Figure 8. Sensitivity compares the performance of various models.

In the sensitivity analysis in Figure 8, which evaluates the model's capacity to identify actual cancer cases, the Proposed Method (Dataset 1) achieves the highest value of 99.79%. This ensures minimal missed diagnoses. The Proposed Method (Dataset 2) follows closely with 99.18%, confirming robust cancer detection capabilities compared to existing approaches.

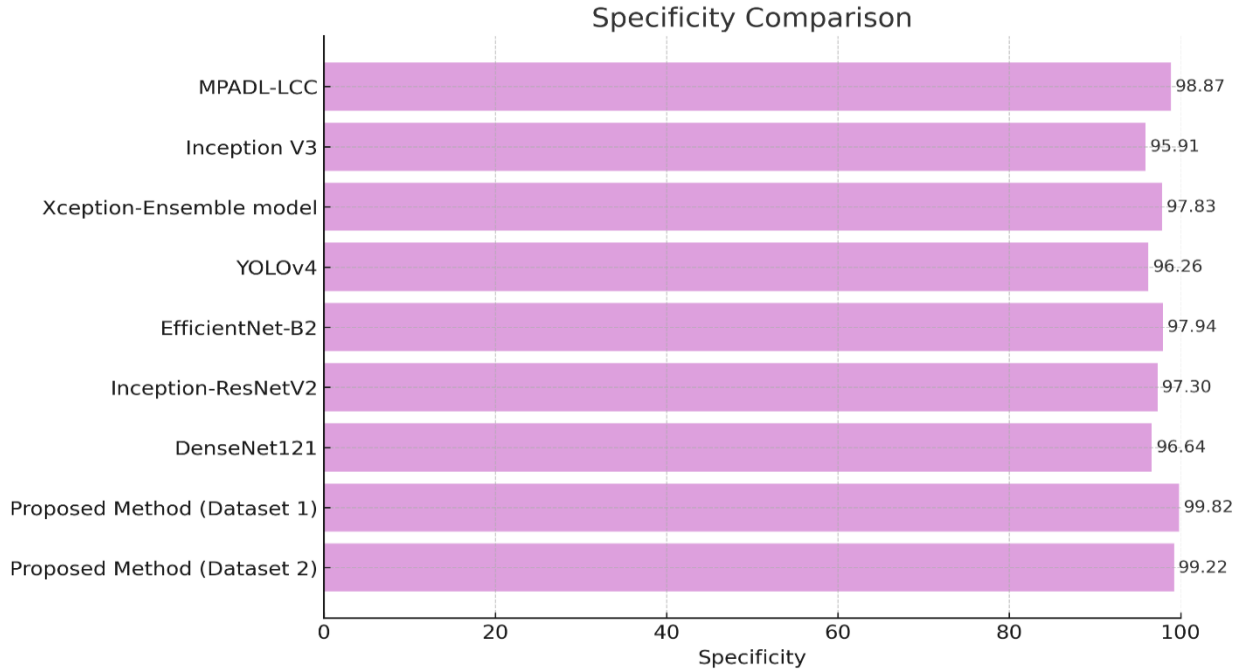

Figure 9. Specificity compares the performance of various models.

Received: November 24, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Figure 9 shows each model's performance in correctly identifying healthy cases. The Proposed Method (Dataset 1) excels with a specificity of 99.82%, ensuring healthy cases are rarely misclassified as cancer. Proposed Method (Dataset 2) maintains a high 99.22% specificity, outperforming other models like Xception-Ensemble and EfficientNet-B2.

5. Conclusion

In this study, a comprehensive evaluation of various deep learning models for leukemia detection was conducted, incorporating advanced methodologies and comparative analysis. The proposed models, particularly those achieving 99.85% and 99.24% accuracy, demonstrated superior performance across all key metrics—accuracy, precision, sensitivity, and specificity—compared to state-of-the-art methods such as MPADL-LCC, Inception V3, and DenseNet121. The integration of hybrid feature extraction using Swin Transformer and EfficientNetV2, coupled with Bayesian optimization for hyperparameters, significantly enhanced model efficiency and robustness. The attention-based denoising autoencoder further improved classification accuracy by minimizing the impact of noise on feature embeddings.

Moreover, visualization using confusion matrices and accuracy curves confirmed the progressive improvement in model performance with increasing epochs. The training and validation curves showed consistent growth without overfitting, ensuring model reliability. Statistical evaluations confirmed the model's generalization capabilities on a real-world dataset comprising 12,529 images, representing both Leukemia (ALL) and healthy cells.

References

1. Das, Sonali, Saroja Kumar Rout, Sujit Kumar Panda, Pradyumna Kumar Mohapatra, Abdulaziz S. Almazyad, Muhammed Basheer Jasser, Guojiang Xiong, and Ali Wagdy Mohamed. "Marine Predators Algorithm with Deep Learning-Based Leukemia Cancer Classification on Medical Images." *CMES-Computer Modeling in Engineering & Sciences* 141, no. 1 (2024).
2. Kantarjian, Hagop M., Courtney D. DiNardo, Tapan M. Kadia, Naval G. Daver, Jessica K. Altman, Eytan M. Stein, Elias Jabbour, Charles A. Schiffer, Amy Lang, and Farhad Ravandi. "Acute myeloid leukemia management and research in 2025." *CA: a cancer journal for clinicians* 75, no. 1 (2025): 46-67.
3. Shimony, Shai, Maximilian Stahl, and Richard M. Stone. "Acute Myeloid Leukemia: 2025 Update on Diagnosis, Risk-Stratification, and Management." *American Journal of Hematology* 100, no. 5 (2025): 860-891.
4. Hallek, Michael. "Chronic lymphocytic leukemia: 2025 update on the epidemiology, pathogenesis, diagnosis, and therapy." *American Journal of Hematology* 100, no. 3 (2025): 450-480.
5. Daltveit, Dagrun S., Eileen Morgan, Murielle Colombet, Eva Steliarova-Foucher, Karima Bendahhou, Rafael Marcos-Gragera, Zheng Rongshou, Alexandra Smith, Hui Wei, and Isabelle Soerjomataram. "Global patterns of leukemia by subtype, age, and sex in 185 countries in 2022." *Leukemia* 39, no. 2 (2025): 412-419.

Received: November 24, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

6. Aziz, Nouman, Waseem Nabi, Muzamil Khan, Abu Huraira Bin Gulzar, Shree Rath, Asad Ali Ahmed Cheema, Mirza Ammar Arshad et al. "Analyzing Two Decades of Leukemia Mortality in the US (1999-2020)." *Clinical Lymphoma Myeloma and Leukemia* (2025).
7. Chen, Jiaxi, Wenyi Pang, Miao Deng, Rubin Zheng, Yanjin Chen, Ziyang Zhang, Zhouke Tan, and Zhixun Bai. "Global, regional, and national burden of leukemia, 1990–2021: A systematic analysis of the global burden of disease in 2021." *Frontiers in Medicine* 12 (2025): 1542317.
8. Kantarjian, Hagop, Massimo Breccia, Fadi G. Haddad, Rüdiger Hehlmann, Ghayas C. Issa, Hemant Malhotra, Franck Emmanuel Nicolini, Koji Sasaki, Leif Stenke, and Elias Jabbour. "Management of chronic myeloid leukemia in 2025." *Cancer* 131, no. 14 (2025): e35953.
9. Lv, Huijuan, Ke Lu, Ximing Wang, Yanfang Zhang, Mengqi Zhuang, Jing Li, Keli Su, and Wen Gao. "Recent trends in research on the role of cholesterol in leukemia: a bibliometric and visualization study." *Frontiers in Immunology* 16 (2025): 1511827.
10. Oybek Kizi, Rakhmonalieva Farangis, Tagne Poupi Theodore Armand, and Hee-Cheol Kim. "A review of deep learning techniques for leukemia cancer classification based on blood smear images." *Applied Biosciences* 4, no. 1 (2025): 9.
11. Hamza, Musa N., Mohammad Tariqul Islam, Sunil Lavadiya, Iftikhar Ud Din, Bruno Sanches, Slawomir Koziel, Syeda Iffat Naqvi, Ali Farmani, and Md Shabiul Islam. "Ultra-compact quintuple-band terahertz metamaterial biosensor for enhanced blood cancer diagnostics." *Plos one* 20, no. 1 (2025): e0313874.
12. Pan, Yuzhe, Qian Liu, Nan Zhang, Shuang Peng, Xinqi Li, and Fuling Zhou. "Global assessment of leukemia care quality: insights from the quality of care index (QCI) from 1990 to 2021." *EClinicalMedicine* 79 (2025).
13. DiAndreth, Breanna, Pavlo A. Nesterenko, Aaron G. Winters, Aaron D. Flynn, Claudia A. Jette, Vasantika Suryawanshi, Sanam Shafaattalab et al. "Multi-targeted, NOT gated CAR-T cells as a strategy to protect normal lineages for blood cancer therapy." *Frontiers in Immunology* 16 (2025): 1493329.
14. Kausar, Mohd Adnan, Sadaf Anwar, Yusuf Saleem Khan, Ayman A. Saleh, Mai Ali Abdelfattah Ahmed, Simran Kaur, Naveed Iqbal, Waseem Ahmad Siddiqui, and Mohammad Zeeshan Najm. "Autophagy and cancer: insights into molecular mechanisms and therapeutic approaches for chronic myeloid leukemia." *Biomolecules* 15, no. 2 (2025): 215.
15. Lei, Shaohua, Sujuan Jia, Sunitha Takalkar, Ti-Cheng Chang, Xiaotu Ma, Karol Szlachta, Ke Xu et al. "Genomic profiling of circulating tumor DNA for childhood cancers." *Leukemia* 39, no. 2 (2025): 420-430.
16. Soleimani Samarkhazan, Hamed, Sara Zehtabcheh, Hamideh Rahmani Seraji, Safedin H. Beqaj, Shamim Tayefeh, Mohammad Hossein Mohammadi, and Mojtaba Aghaei. "Unveiling the potential of CLL-1: a promising target for AML therapy." *Biomarker Research* 13, no. 1 (2025): 28.
17. Gupta, Sumit, Rachel E. Rau, John A. Kairalla, Karen R. Rabin, Cindy Wang, Anne L. Angiolillo, Sarah Alexander et al. "Blinatumomab in standard-risk B-cell acute lymphoblastic leukemia in children." *New England Journal of Medicine* 392, no. 9 (2025): 875-891.

Received: November 24, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

18. Chen, Yu, Zhenghao Yin, Kenneth D. Westover, Zhiwei Zhou, and Liping Shu. "Advances and challenges in RAS signaling targeted therapy in leukemia." *Molecular Cancer Therapeutics* 24, no. 1 (2025): 33-46.
19. Perner, Florian, Jayant Y. Gadrey, Scott A. Armstrong, and Michael WM Kühn. "Targeting the Menin–KMT2A interaction in leukemia: Lessons learned and future directions." *International Journal of Cancer* (2025).
20. Siew, Zhen Yun, Ghee Khang Ong, Siew Tung Wong, Pooi Pooi Leong, Boon Shing Tan, Chee-Onn Leong, Juita Binti Chupri, Chee-Mun Fang, and Kenny Voon. "Safety profile of sikamat virus and its oncolytic potential in leukemic cells and cancer stem cells." *Scientific reports* 15, no. 1 (2025): 13817.
21. Ji, Xu, Xinyue Zhang, K. Robin Yabroff, Wendy Stock, Patricia Cornwell, Shasha Bai, Ann C. Mertens, Joseph Lipscomb, and Sharon M. Castellino. "Health insurance continuity and mortality in children, adolescents, and young adults with blood cancer." *JNCI: Journal of the National Cancer Institute* 117, no. 2 (2025): 344-354.
22. Desai, Seema, Libby M. Morimoto, Alice Y. Kang, Mark D. Miller, Joseph L. Wiemels, Lena E. Winestone, and Catherine Metayer. "Pre-and Postnatal Exposures to Residential Pesticides and Survival of Childhood Acute Lymphoblastic Leukemia." *Cancers* 17, no. 6 (2025): 978.
23. Shehta, Amr I., Mona Nasr, and Alaa El Din M. El Ghazali. "Blood cancer prediction model based on deep learning technique." *Scientific Reports* 15, no. 1 (2025): 1889.
24. Hollanda, Cíntia Nogueira, Ana Cristina Moura Gualberto, Andréa Barretto Motoyama, and Fabio Pittella-Silva. "Advancing Leukemia Management Through Liquid Biopsy: Insights into Biomarkers and Clinical Utility." *Cancers* 17, no. 9 (2025): 1438.

Received: November 24, 2025