RESEARCH ARTICLE

INTERNATIONAL RESEARCH JOURNAL OF MULTIDISCIPLINARY TECHNOVATION

Asian Research Association

DOI: 10.54392/irjmt2563

Detection and Classification of Genetic Acute Myeloid Leukemia Cells using Deep Learning Techniques

Hema Patel [a,*], Himal Shah [b], Gayatri Patel [c], Atul Patel [a]

[a] Smt. Chandaben Mohanbhai Patel Institute of Computer Applications, Charotar University of Science and Technology, CHARUSAT – Campus, Changa – 388421, Anand, Gujarat, India

[b] QURE Haematology Centre, Ahmedabad –380006, Gujarat, India

[c] Ramanbhai Patel College of Pharmacy, Charotar University of Science and Technology, CHARUSAT – Campus, Changa – 388421, Anand, Gujarat, India

* Corresponding Author Email: hemapatel.mca@charusat.ac.in

DOI: https://doi.org/10.54392/irjmt2563

Received: 16-07-2025; Revised: 29-09-2025; Accepted: 08-10-2025; Published: 16-10-2025

Abstract: Acute Myeloid Leukemia (AML) is a hematological disease that is defined by the fast growth of aberrant myeloid precursor cells in the blood and bone marrow, which disrupts normal hematopoiesis. Treatment and prognosis are influenced by the early detection of this deadly illness and its appropriate classification. Therefore, utilizing the Human Leukemia Cytomorphology Collection dataset, which comprises leukemic and normal single-cell images of Acute Myeloid Leukemia (AML) type, this research suggests a deep learning-based hybrid model for automated leukemia detection and classification. By taking morphological characteristics and genetic abnormalities into account, leukemic cells have been distinguished. The features in this study are extracted using MobileNetV2, ResNet-101, and VGG-16. Then, feature-level stacking is performed using the Support Vector Machine and Random Forest classifiers for final classification using Principal Component Analysis (PCA). Utilising image segmentation, normalisation, data augmentation, and data oversampling techniques, the pipeline improves data quality and corrects class imbalance. Additionally, t-distributed Stochastic Neighbour Embedding (t-SNE), which shows the extracted features used for the detection of leukemia subtypes, and Gradient-Weighted Class Activation Mapping (Grad-CAM) images help with interpretability by highlighting important decision areas. The suggested study achieved 98.35% accuracy, 95.87% precision, 95.84% sensitivity, 98.97% specificity, and 95.74% F1-Score. Along with the trial results, a comparison of the four separate frameworks, viz., MobileNetV2, ResNet-101, VGG-16, and Vision Transformer, has also been carried out. The comparison shows that the proposed model outperforms the other frameworks. The outcomes show that the suggested model has the capability to be used a reliable means for the prompt identification of AML and its subtypes.

Keywords: Acute Myeloid Leukemia, Genetic Abnormalities, Mutation, Hematological Disorder, Cytomorphology, Stacking, Feature Extraction.

1. Introduction

Acute Myeloid Leukemia (AML) is a rapidly progressing cancer in which the bone marrow of the human body proliferates an excess of abnormal white blood cells of a specific type named myeloblasts [1, 2]. Acute granulocytic leukemia, acute myeloblastic leukemia, acute myelogenous leukemia, and acute nonlymphocytic leukemia are some additional names, usually popular to refer to this lethal disease. The likelihood of contracting this ailment is accelerated by a number of factors, including age, lifestyle, gender environment, food choices, and geographic proximity. The main cause of the leukemia is the generation of undeveloped white blood cells (WBCs) in the abundant quantity [3]. This type of leukemia typically affects adults who are above 70 years of age. Patients under the age of 60 have an average chance of recovery. Still, in contrast, many individuals with AML have a worse prognosis or high chances of relapse, particularly as they age or have specific genetic abnormalities. Even though 5-year survival rates are close to 70% because of advancements in diagnosis and treatment, recurrence still happens in about 30% of patients [4]. As per the recent survey, an estimated 1.54% cases of acute myeloid leukemia (AML) occur for every 100,000 person-years worldwide. Particularly in the USA and Asian nations, the prevalence has been rising. Due to its acute nature, it spreads swiftly and gives the patient little time to recover. Also, compared to females, males have a higher chance of developing this leukemia.

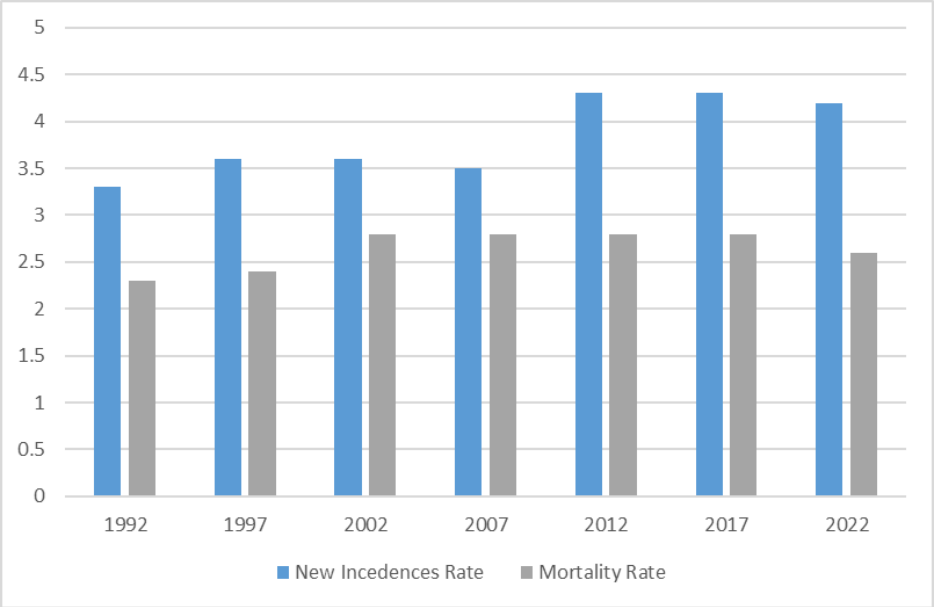


Figure 1. Data on New Instances and Mortality Rate

According to statistics from the National Cancer Institute, there have been 22,010 new cases of cancer recorded in the USA, with a mortality count of 11,090 [5]. The National Cancer Institute also provided data on the latest case rate, the increased rate, and the mortality rate per 100,000 individuals, including male and female genders, from 1992 to 2023, as illustrated in Figure 1.

This statistic shows the seriousness of the disease. Curing this illness and saving lives is therefore crucial, and this can be achieved with the aid of a timely and accurate diagnosis. Conventional procedures or manual diagnosis, such as flow cytometry [6], immunophenotyping [7], bone marrow biopsies, and cytogenetic analysis [8], can be used; however, they are laborious, time-consuming, and prone to mistakes due to factors including pathologist experience, camera performance, and the weariness of medical professionals. Due to these limitations, automatic detection and classification have become popular because of their accurate results [9, 10], which can be achieved with the aid of artificial intelligence. Artificial intelligence (AI)-based computerized microscopic analysis of blood smear images [11] has become a more dependable and quicker approach to this concern. Deep Learning (DL) methods and neural networks are capable of expertly classifying single blood and bone marrow cells [12, 13], predicting cancer patient survival [14, 15], differentiating between leukemia subtypes [16], and identifying cancer types by extracting the information from large image data sets with single cell resolution.

Many state-of-the-art techniques and DL techniques now in use struggle with a number of issues, including selecting important features for classification, accurately segmenting overlapping cells, removing unwanted and noisy cells, and class imbalance. Furthermore, single classifiers are frequently used in traditional machine learning frameworks, which might not completely utilise the variety of data available for effective categorisation. To address these limitations, a novel hybrid deep learning architecture is developed in this work to extract essential features using Principal Component Analysis (PCA) and perform final classification with an ensemble classifier that includes Support Vector Machine (SVM) and Random Forest (RF) classifiers through stacking. The contribution of this study is as follows:

- A novel leukemia detection and classification model is proposed, which provides accurate segmentation and admirable classification results.
- The implementation of oversampling techniques ensures equitable representation across categories, thereby addressing the issue of class imbalance in AML subtypes. This enabled the model to produce accurate predictions that were free from bias towards majority classes.
- Furthermore, our study provides interpretability through visualization techniques. Specifically, Gradient-Weighted Class Activation Mapping (Grad-CAM) illustrations emphasise key decision areas. At the same time, t-distributed Stochastic Neighbour Embedding (t-SNE) plots depict the extracted feature distributions employed for leukemia subtype detection.
- Ablation study emphasizes the importance of the distinct components that we have incorporated in the development of the proposed model.
- Following the extraction of vital and essential features using hybrid DL models, PCA-based feature selection is used. Finally, promising classification results are obtained by

implementing stacking with the help of ensemble classifiers SVM and RF.

- The proposed model has been compared with other four pre-trained networks viz., MobileNetV2, ResNet-101, VGGNet-16, and Vision Transformer using the same dataset. This comparison study also showed that the suggested approach works superior in identifying malignant cells with various subtypes.

Despite its immense potential, there are some limitations. Although the AML Cytomorphology dataset is larger and more diverse, all the tests in this study employed data from a single source, which may have introduced dataset-specific biases due to variations in staining techniques, imaging equipment, or demographic factors. This makes it challenging to determine whether the proposed paradigm can be applied in various labs or clinical settings. Nonetheless, the framework has significant potential to enhance current diagnostic practices, reduce clinical burden, and enable faster, more accurate AML identification. Using independent cohorts and multi-center datasets, future research should focus on verifying the model's robustness and real-world application.

This manuscript is organised into five sections, each addressing a specific purpose of the research. Section 2 provides background information on the classification of AML. Section 3 provides a literature review. Section 4 outlines the methodology of the proposed model, including the training policy. Section 5 illustrates the results obtained and analyze the performances. Section 6 discusses the conclusion and future directives of the research.

2. Background Updates on the Classification of Acute Myeloid Leukemia

Blood comprises various types of blood cells, one of which plays a crucial role in the progression of leukemia is myeloid. Myeloid is a sort of progenitor blood cell and are further differentiated into red blood cells (RBCs), basophils, granulocytes (eosinophils), neutrophils, monocytes, and megakaryocytes, which produce platelets [17]. The myeloblasts, a precursor cell of granulocytes, elevate in quantity with immature proliferation in the case of infection or leukemia. Especially in AML, an accelerated number of myeloblasts acts as a key indicator of being affected with this specific type of leukemia. AML is mostly diagnosed clinically based on myeloblast volume and genetic characteristics. Changes in the amount or structure of chromosomes due to translocations, deletions, insertions, inversions, monosomies, trisomies, polyploidy, and other abnormalities are referred to as cytogenetic disturbances. About 55% of AML patients often have one or more cytogenetic anomalies.

Three renowned systems categorise AML, namely, the French-American-British (FAB) classification, the International Consensus Classification (ICC), and the World Health Organisation (WHO) classification. The French-American-British (FAB) organization classified AML into its subtypes in the year 1976 on the basis of the blood cell's cytochemical and morphological features, as given in Table 1. Initial proposals for the six FAB subtypes (M1 through M6) were made in 1976; however, M7 and M0 were added in 1985 and 1987, respectively. Rare forms not covered by the FAB classification are frequently included in the morphologic subtypes of AML. For example, acute basophilic leukemia was suggested as a ninth subtype, M8, Acute Basophilic Leukemia in 1999 and not included in the FAB classification type of AML.

Table 1. Classification of AML (French-American-British)

Type of Leukemia	Name of Variant	Description
Acute Myeloid Leukemia	M0	AML without differentiation
	M1	AML with minimal maturation
	M2	AML with maturation
	M3	Acute Promyelocytic Leukemia (APL)
	M4	Acute Myelomonocytic Leukemia
	M4eo	Myelomonocytic Leukemia with Bone Marrow eosinophilia
	M5 (a)	Acute Monoblastic Leukemia
	M5 (b)	Acute Monocytic Leukemia
	M6	Acute Erythroid Leukemia
	M6 (a)	Erythroleukemia
	M6 (b)	Unusual Erythroid Leukemia
	M7	Acute Megakaryoblastic Leukemia

Table 2. Classification of AML (World Health Organization)

Classification	Subtype
AML with defining genetic abnormalities	AML with RUNX1-RUNX1T1 Fusion
	AML with CBFB-MYH11 Fusion
	APL with PML-RARA Fusion
	AML with KMT2A-MLLT3 Fusion
	AML with DEK-NUP214 Fusion
	AML RBM15-MKL1
	AML with BCR-ABL1 Fusion
	AML with KMT2A Rearrangement
	AML with MECOM Rearrangement
	AML with NUP98 Rearrangement
	AML with NPM1 Mutation
	AML with Biallelic Mutation of CEBPA
	AML with Myelodysplasia-related Changes
	AML with other Defined Genetic alterations
AML is defined by differentiation	AML with Minimal Differentiation
	AML without Maturation
	AML with Maturation
	Acute Myelomonocytic Leukemia
	Acute Basophilic Leukemia
	Acute Monoblastic
	Acute Monocytic Leukemia
	Acute Erythroid Leukemia
	Acute Megakaryoblastic Leukemia

A subtype of AML, M3 is also referred to as Acute Promyelocytic Leukemia (APL). The FAB system is based on cytochemistry and the morphological characteristics of blasts as observed under a microscope [18]. It separates AML into subtypes according to the leukemic cells' cell type and level of maturity. The type of cell lineage involved and the ratio of differentiation define these subtypes. M0, M5, and M6 are derived from immature white blood cells, immature red blood cells, and immature platelet-making cells, respectively.

Currently, the FAB classification system has been largely superseded by the World Health Organisation (WHO) classification system, which employs more contemporary methods for categorising AML, based on the genetic traits of the cells and immunophenotypic studies. The focus of this research is on the WHO classification. Although the FAB classification aids in the diagnosis of AML, the World Health Organization (WHO) classification has a stronger correlation with treatment outcomes. Actually, WHO categorises Myeloid Neoplasms, a cluster of cancers. The WHO has categorised myeloid neoplasms into a total of nine kinds, which include AML, Myelodysplastic Neoplasms (MDNs), Myeloproliferative Neoplasms (MPNs), and other Myeloid Neoplasms. The most recent edition, WHO 2022 [19], has been revised in light of new diagnostic standards and genetic discoveries, and classified AML into the following subtypes, as listed in Table 2.

The study of blood smear images requires skilled and knowledgeable medical personnel, and genetic testing of blood smear images is a procedurally challenging operation. In clinical practice, an automated AML classification system that could overcome these issues is required. In this research, the utilized dataset comprises of 81214 single cell images of four sub-types of AML with recurrent genetic abnormalities kind along with healthy cells analyzed have been analyzed. Genetic variations and the proportion of myeloblasts (leukemia cells) in bone marrow and blood are used to classify AML. The brief summary of these subtypes is follows:

I. ***Normal (Control):*** Collection of benign blood cells.

II. ***APL with PML-RARA:*** A particular genetic anomaly associated with Acute Promyelocytic Leukemia (APL) is known as PML-RARA. It is a fusion gene created via a chromosomal translocation, which involves the exchange of genetic arrangement between chromosomes 15 and 17. In particular, the RARA gene on chromosome 17 and the PML gene on chromosome 15 are fused [20]. APL is characterized by the accumulation of aberrant promyelocytes in the bone marrow, which is caused by a special protein produced by this fusion that interferes with normal cell proliferation and differentiation.

III. ***AML with NPM1 Mutation:*** A unique subtype of AML known as Nucleophosmin 1 (NPM1) - mutated AML is caused by a mutation in the Nucleophosmin 1 (NPM1) gene [21].

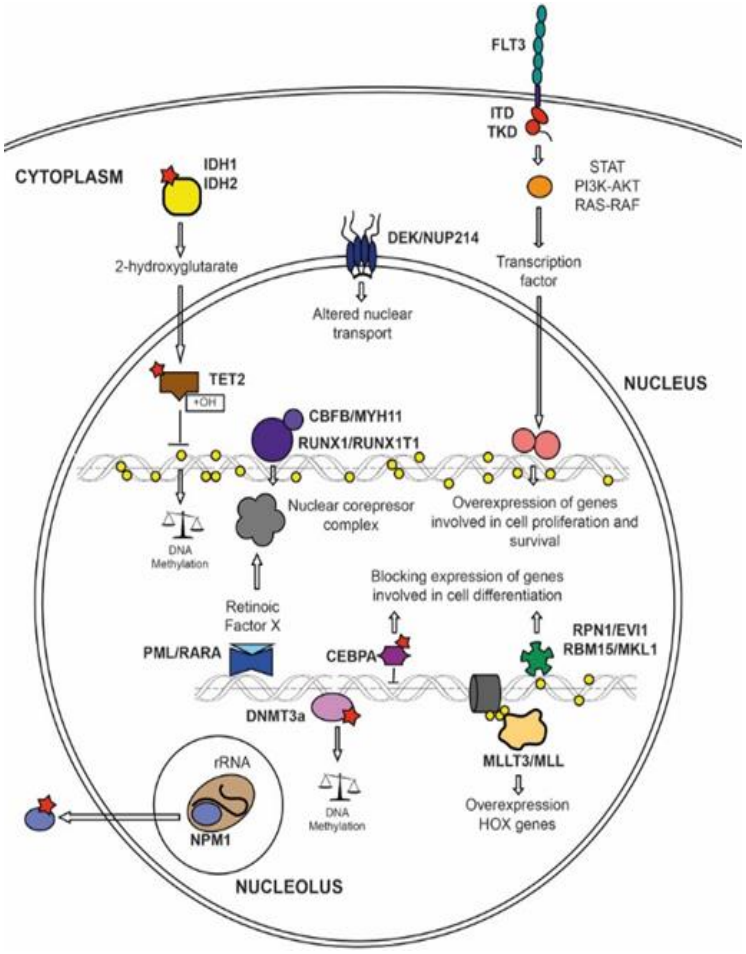


Figure 2. Diagram of the AML Genetic-Molecular Background

Mutations can lead to NPM1. NPM1 is kind of a protein, which is located in the nucleus that is responsible for ribosome synthesis and nucleolar activities, to be inappropriately localized in the cytoplasm. Exon 12 is where these mutations are generally discovered.

IV. ***AML with CBFB-MYH11 (without NPM1 mutation):*** It is caused by a translocation or inversion of a chromosome. The CBFB (Core Binding Factor beta) and MYH11 (Myosin Heavy Chain 11) genes unite to form the combined gene [22]. This fusion is caused by chromosomal rearrangement, usually a balanced translocation or an inversion of chromosome 16 (inv(16)(p13.1q22)). This fusion plays a crucial role in the onset and progression of AML, especially in younger people, when NPM1 mutations are absent.

V. ***AML with RUNX1-RUNX1T1:*** A reciprocal translocation between chromosomes 8 and 21 results in a fusion gene, which is its defining feature. Full of RUNX1 and RUNX1T1 is Runt-related Transcription factor 1. The t (8; 21) (q22; q22.1) translocation specifically causes RUNX1 and RUNX1T1 to fuse, producing a novel fusion protein [23]. AML patients normally exhibit this type of fusion, which is usually produced by this translocation and is associated with a favourable prognosis. Figure. 2 illustrates these fusions [24].

This research aims to develop a deep learning model that automatically identifies normal cells and subsequently classifies AML into four categories based on its genetic disorder and mutation.

3. Literature Review

A number of research studies have been thoroughly analysed in this particular section. This research is based on both - microscopic blood smear and bone marrow images using ML and DL techniques.

Acharya *et al.* [25] proposed a diagnosis methodology to segment peripheral blood smear images in order to identify and classify AML subtypes. To classify AML into its sub-categories using the ML-based RF classifier, morphologic characteristics of leukemic cells, including cytoplasm, Auer rods, and numerous lobes, were retrieved from a small dataset of 1000 images for training and 500 images for testing. A classification model was offered by Rawat *et al.* [26], which is responsible to categorize acute lymphoblasts, myeloblast cells, and their subtypes. After 420 blood

smear images were segmented, 331 attributes were collected and classified using Support Vector Machine (SVM) [27] using a genetic algorithm and multiple kernel functions. Although the proposed system obtained high accuracy, but was relied on hand-crafted features. Likewise, using 8637 blood smear images, Narayanan *et al.* [28] used RF, hybrid fuzzy C Means, and SVM to identify acute leukemia. Using the K-means algorithm, all images were segmented with an accuracy of 99.06%. Shaheen *et al.* [29] created a classification model based on AlexNet to identify AML in microscopic blood images. For research objectives, a dataset of 4,000 blood smear samples was used and the author compared the model with the LeNet-5 architecture. PRICoLBP, SVM, and RF were used by Zhao *et al*. [30] to categorize neutrophils, lymphocytes, and monocytes from peripheral blood smear images. The method was predicated on WBC morphological characteristics and color correlation. Similarly, Ke Liu *et al*. [31] created a framework to identify AML into its sub-categories M1 and M2. Using RF and the Broad Learning System (BLS), morphological, medical, and radiomics features were selected and categorised from a small dataset comprising only 50 bone marrow images.

Similar to this, Yu *et al.* [32] used employed DL networks to analyze distinct kinds of WBC images. Classifiers employed included ResNet50, VGG16 [33], Inception V3, Xception, and VGG19 [33] on the dataset without erythrocytes and platelets. “Mayfly optimization with Generative Adversarial Network (MayGAN)” was utilized by Veeraiah *et al*. [34] to optimise the classification process. For segmentation, morphological processes utilising geometric features were used. This model obtained an accuracy of 99.8% in the classification of abnormal WBCs, AML, ALL, CML, and CLL but not able to classify the images with overlap stains. For the detection of AML malignancy, Venkatesh *et al*. [35] created a unique framework known as "Enhanced Few-Shot Learning Technique Integrated Base Classifier (Feature Encoder)-EFLTBC." In order to classify the lineage of acute leukemia, Boldú *et al.* [36] chose the ALNet system that has two modules: one for identifying abnormal promyelocytes and the other for differentiating across lineage kinds. 16000 blood smear images were used in this investigation to identify the leukemia blast lineage and predict the difference between leukemia and infections. In order to identify distinct types of leukemia, Agrawal *et al.* [37] segmented the data using “Otsu Adaptive Thresholding and Gaussian Distribution”, and then used the K-Means approach for clustering. To classify 100 blood smear images into two groups - ALL (subtypes: L1, L2, and L3) and AML (subtypes: M2, M3, and M5), but worked on a small dataset.

In a similar vein, Qin *et al.* [38] implemented a deep residual neural network for the division of leukocytes. However, the authors required implementing a proper segmentation policy in order to improve the results. Using YOLOv5, Rohaziat *et al.* [39] trained the model using 2800 blood smear images to identify several types of white blood cells (eosinophil, lymphocyte, monocyte, and neutrophil). Convolutional Neural Network (CNN) layers were used to extract features, which took a long time to reach the final decision. Similarly, on the collection of 18,365 images of individual cells, Matek *et al.* [12] also employed a CNN to detect AML cells. They achieved this by merging morphological and taxonomic groups to produce 15 distinct categories for training and examination. Images were classified using the ResNeXt CNN architecture. In a similar vein, by using a multi-step DL approach, Eckardt *et al.* [40] are able to automatically segment cells from bone marrow images, differentiate between AML and healthy images, and use only image data from bone marrow smears to predict the mutation status of Nucleophosmin 1, a sort of AML mutation. This model lacks transferability to other datasets. To identify ALL and AML, Zare *et al*. [41] suggested a model that combined CNN and graph theory, featuring six graph Conv levels and a Softmax layer. By physically inspecting and evaluating the morphological characteristics of lymphocyte and monocyte cells, the leukemia categorization was established. Using substantial datasets of blood cell images, Ramesh *et al*. [42] developed two deep learning based models for AML detection. These models used up to 97% less parameters to detect AML because researchers employed a variety of kernel shapes to uncover intricate patterns from images. In order to identify AML, a Fractional Black Widow-based Neural Network (FBW-NN) was presented by Ramya *et al*. [43] to increase the efficacy of Artificial Neural Networks (ANN). To partition the AML region, the Adaptive Fuzzy Entropy (AFE) method was developed. The dynamic contour-based method and the fuzzy C-mean clustering technique are combined. The statistics and features at the image level are obtained after separation. The application to tumour locations that were growing was limited, and it also needs to work on the denoising process for the input images. Similarly, Aby *et al.* [44] implemented the Gaussian error linear unit (GELU) activation function, achieving 94.02% accuracy in classifying 669 malignant images with reduced training time. Table 3 summarizes various studies on acute myeloid leukemia detection, and classification on blood smear images using ML and DL techniques. For leukemia classification problems, this data offers a solid foundation for suggesting sophisticated CNN-based models, which are more accurate, more generalised, and have more useful applications.

4. Method

The hybrid deep learning architecture that combines image preprocessing and feature extraction from three deep CNN architectures, viz., MobileNetV2, ResNet-101, and VGG-16, which was followed by

Principle Component Analysis (PCA) [45] and final classification using an ensemble classifier, which includes Support Vector Machine (SVM) and Random Forest (RF) classifiers through stacking and implementation of Gradient-weighted Class Activation Mapping (GRAD-CAM) and t-distributed Stochastic Neighbour Embedding (t-SNE) on the single-cell blood smear images, is suggested as shown in Figure. 3.

In this research, the task of recognising healthy and malignant cells and classifying malignant cells into their subtypes based on their genetic composition has been carried out. By integrating the benefits of both CNN models, the stacking technique enhances the overall accuracy and resilience of the classification process. Additionally, it uses progressive training simulation for performance evaluation and Grad-CAM for visual interpretability. This research categorises the data into five distinct groups: AML with NPM1 mutation, APL with PML-RARA fusion, AML with CBFB-MYH11 fusion (excluding NPM1 mutation), ALL with RUNX1-RUNX1T1 fusion, and control (healthy cells), as per the 2022 WHO AML classification.

4.1 Dataset

Human Leukemia Cytomorphology Collection dataset has been used in this study which comprises of benign and leukemic cells. Leukemic single cell images are actually different variants of AML, namely, APL with PML::RARA fusion, AML with NPM1mutation, AML with CBFB::MYH11 fusion (without NPM1 mutation) and AML with RUNX1::RUNX1T1 fusion. As part of the research conducted at the Munich Leukemia Laboratory (MLL), 189 peripheral blood smears from the MLL database between 2009 and 2020 have been scanned using a 10x magnification to get an overview image. A segmentation threshold and logarithmic colour modification were applied to the obtained samples in order to automatically detect cells using the Metasystems Metafer platform.

Table 3. Studies Using Machine Learning And Deep Learning Methods For Detection And Classification of Acute Myeloid Leukemia

Sr. No.	Year	Author & References	Dataset	Data Volume	Methods	Accuracy
1.	2019	Agrawal *et al.* [37]	-	100	K-Means, Gray Level Co-occurrence Matrix	97.30%
2.	2019	Matek *et a*l. [12]	Munich University Hospital	18365	ResNext	99%
3.	2021	Shaheen *et al.* [29]	Hospital of Peshawar	4000	AlexNet	98.58%
4.	2021	Ramya *et al*. [44]	AML Morphology dataset and CPTAC-AML dataset	18,365 and 120	FBW-NN	96.56%
5.	2022	Venkatesh *et al.* [35]	-	4000 and 18384	ResNet-18 and Enhanced Few-Shot Learning Technique integrated Base Classifier	97.87%
6.	2023	Acharya *et al*. [25]	Kasturba Hospital	800	RF	99.48%
7.	2023	Veeraiah *et al*. [34]	-	1200	Mayfly optimization with Generative Adversarial Network and Generative Adversarial System	99.8%
8.	2024	Zare *et al*. [42]	Ghazi Tabriz Medical Sciences Center	670	generative adversarial networks (GANs)	99%
9.	2025	Ramesh *et al.* [43]	-	-	CNN	99.70%
10.	2025	Aby *et al.* [44]	ASH	669	GELU with InceptionV3	94.02%
11	2025	Das *et al.* [46]	ALLIDB and ASH	108 and 80	Grey Level Co-occurrence Matrix (GLCM), and PCA	99.07% 96.25%

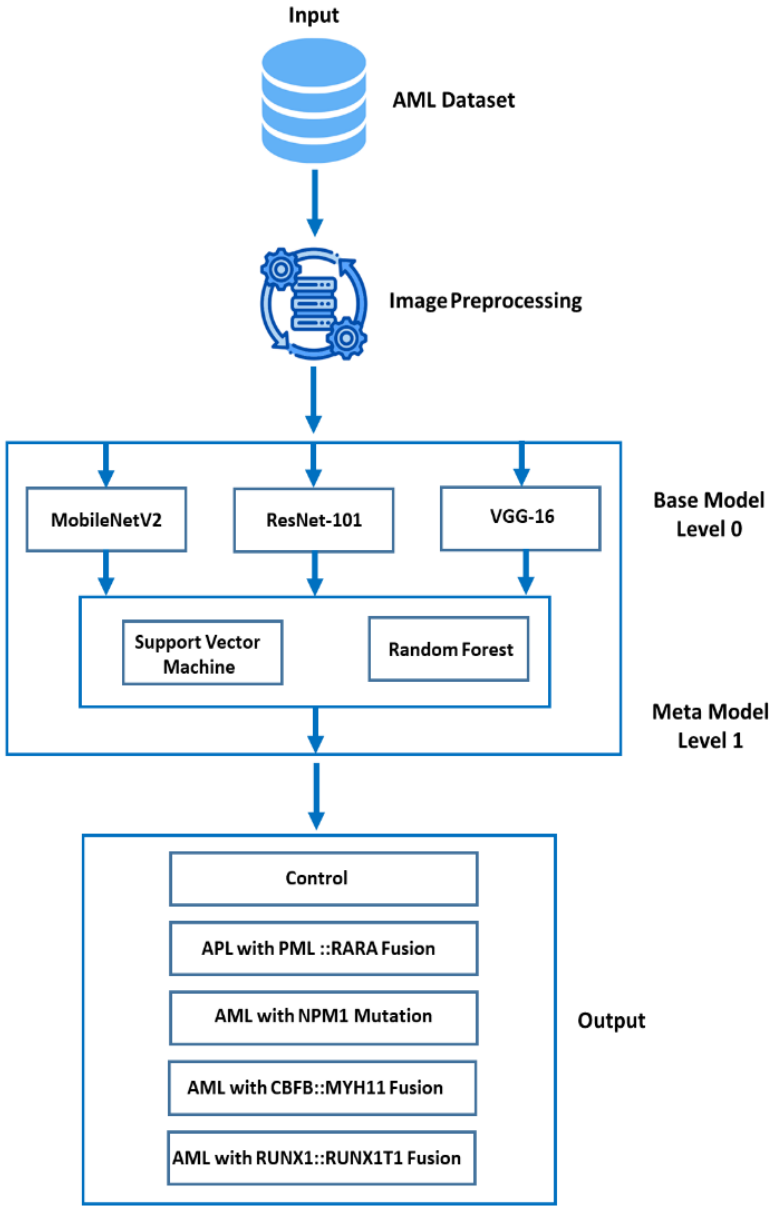


Figure 3. Schematic Diagram of Proposed Hybrid Deep Learning Model with Ensemble Stacking

Utilizing a MetaSystems CMOS Colour Camera with a resolution of 4096 x 3000 px and a pixel size of 3,45µm x 3,45µm, every subject's 99-500 white blood cells were scanned at 40x magnification and saved as images of the size 144 X 144 pixels with .TIF format and made available on Kaggle and National Cancer Institute, Cancer Imaging Program with AML-Cytomorphology_MLL_Helmholtz title.

This large dataset comprises a total of 81214 single-cell images, and all images from each folder and its subfolders have been utilised in this research. Some samples from each category are shown in the Figure. 4.

Approximately 12800 features were retrieved from every image [47]. The total count of 189 blood smears is divided into each group as depicted in Figure. 5 and Figure. 6 illustrate the division of images among five categories:

4.2 Image Pre-processing

4.2.1 Image Resize

The original images are square and somewhat small, measuring 144 x 144 pixels only. Resizing images

From 144×144 pixels to 224×224×3 is an essential pre-processing step in deep learning for leukemia identification in order to guarantee compatibility with common CNNs like VGG-16, ResNet-101, and MobileNetV2. These pre-trained CNN architectures are capable of handling input images with three RGB colour channels and a pixel size of 224 x 224. As a result, before being fed into the network, images need to be resized.

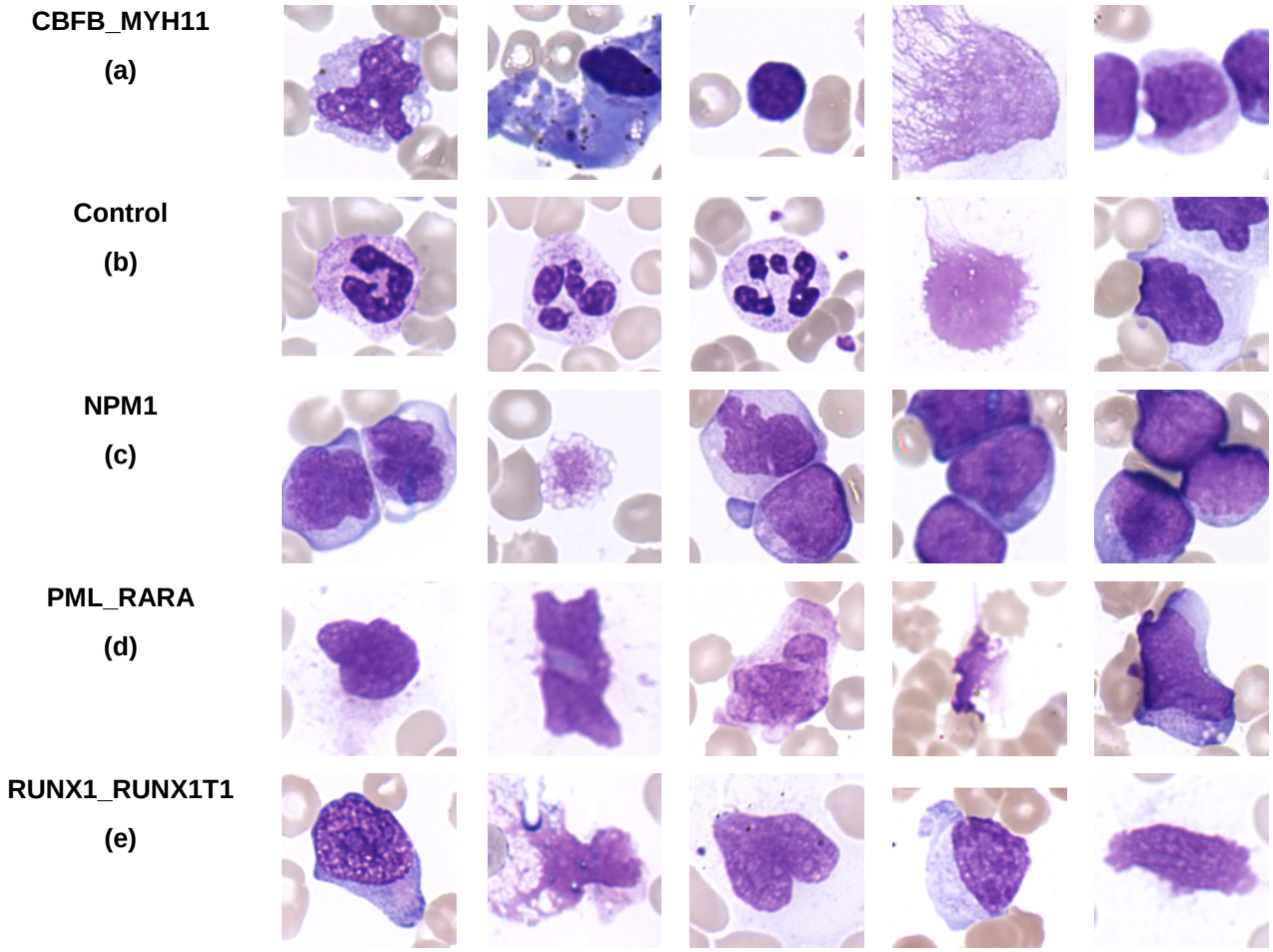


Figure 4. AML Dataset

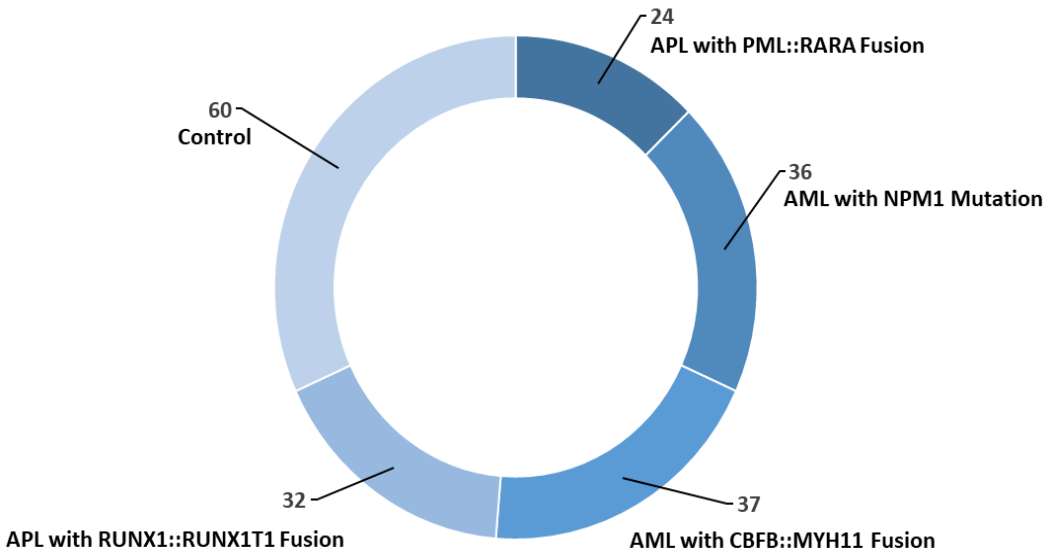


Figure 5. Division of 189 Blood Smears

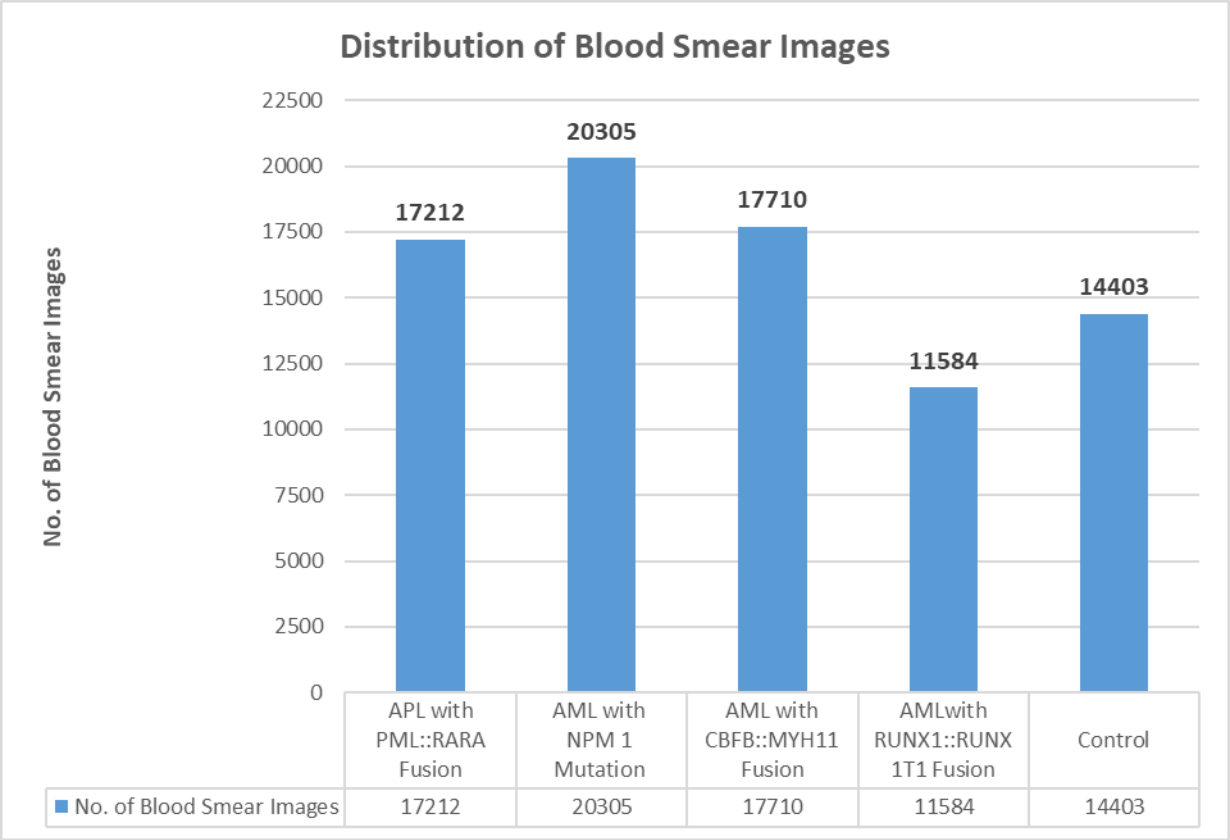


Figure 6. Division of 81214 Blood Smear Images

4.2.2 Augmentation

To improve the generalisation capacity of deep learning models, this study employed a data augmentation procedure as a significant preprocessing approach to artificially enhance the quantity and diversity of the training dataset.

The following transformations are used in the suggested model in order to elevate the no. of images

a ***Random Rotation:*** An arbitrary rotation between -10° and +10° was applied to the genetic abnormal single cell images. This kind of transformation makes the model more resilient to variations in cell appearance due to distinct orientations. A rotation of -10 to +10 was chosen for the proposed model.

b ***Random Reflection:*** The random application of horizontal and vertical flips replicated the unpredictability of the cells' possible placement on microscopic plates.

c ***Random Scaling:*** Images were sized between 0.9 and 1.1 times their original size along the X and Y axes using random scaling. Because blood smear images typically include cells of slightly different shapes and proportions, this helps the model learn to detect them.

4.2.3 Data Balancing

To remedy the class imbalance in the training dataset, oversampling is employed to equalise the number of samples in each class. In medical imaging tasks like leukemia classification, this is especially important because the classifier may be biased toward majority classes, because in our dataset, the PML-RARA class had fewer images, 8109 and the control class had the maximum no. of images, that is, 14214. To implement data balancing using an oversampling mechanism, the model first determined the number of images in each class, followed by the determination of the maximum number of images that could be included in a single class; this became the target size for the remaining classes.

The code's oversampling procedure operates as follows:

a ***Calculation of Class Distribution:*** First, the model determines the number of images in each class, followed by the determination of the maximum number of images that can be included in a single class, which becomes the target size for the remaining classes.

b ***Reason for Replication:*** For every class whose images are not the expected size, the algorithm calculates the number of additional images required. Two steps are necessary to achieve the objective:

I. ***Full Replications:*** A sufficient number of instances of each image in the class are simulated to approach the desired size.

II. ***Random sampling:*** It chooses images from the same class at random to account for the remaining difference in order to avoid excessive replication of any one image.

c ***Dataset Creation:*** In this process, all of the original and repeated images are combined to create new images. Training bias is reduced because the dataset now contains an equal number of images from each class.

d ***Integration of Augmentation:*** Eventually, three extractors —MobileNetV2, ResNet-101, and VGG-16 — extract features using the oversampled dataset and augmented images.

4.2.4 Segmentation

One of the crucial preprocessing steps is segmentation, which enhances the visibility of leukaemia cells and separates them from extraneous background data. To prevent downstream classification models from being misled by irrelevant regions, this stage ensures that they focus on the image's informative elements, such as nucleus or cytoplasm patterns.

The first step in the suggested paradigm is to define input and output directories for the original and segmented image organization, respectively. It loops through every image in a dataset that is arranged into subfolders, each representing a distinct class (for example, APL with PML-RARA fusion, AML with NPM1 mutation, control, AML with CBFB-MYH11 fusion (without NPM1 mutation), and AML with RUNX1-RUNX1T1 fusion). To ensure uniformity, the image is initially examined to check whether it is in RGB format or not. The image is then converted to the colour format known as HSV (Hue, Saturation, and Value). For biomedical imaging segmentation tasks, HSV is favoured over RGB because it simplifies the extraction of cells based on colour characteristics that distinguish them from the backdrop by separating colour information (colour and intensity) from brightness. Based on the thresholds, a binary mask emerged. The original RGB image was then subjected to this mask, and all pixels outside of it were zeroed out, or set to black, and the background noise was removed. The segmented image is then stored in an appropriate output segmented dataset folder for the same class, highlighting only the prospective leukemic areas. In this way, image segmentation contributes significantly to the classification accuracy of cells by preventing the model from being misled by irrelevant and unnecessary patterns or elements present in the images.

4.3 Feature Extractor and Classifier

In this research, we implemented the stacking strategy in conjunction with three feature extractors. Compared to employing a single CNN, this method may yield better accuracy and adaptability by combining the diversity and robustness achieved by three distinct CNNs. The employment of PCA followed feature extraction. PCA reduced the dimensionality of the feature vector by removing irrelevant features. Furthermore, we used the ensemble stacking strategy, which has attained popularity when combined with deep learning methodologies.

4.3.1 MobileNetV2

A potent and effective method of combining the representational strength of deep learning with conventional machine learning classifiers for tasks such as object recognition, disease classification, or medical image analysis is to utilise MobileNetV2 as a feature extractor.

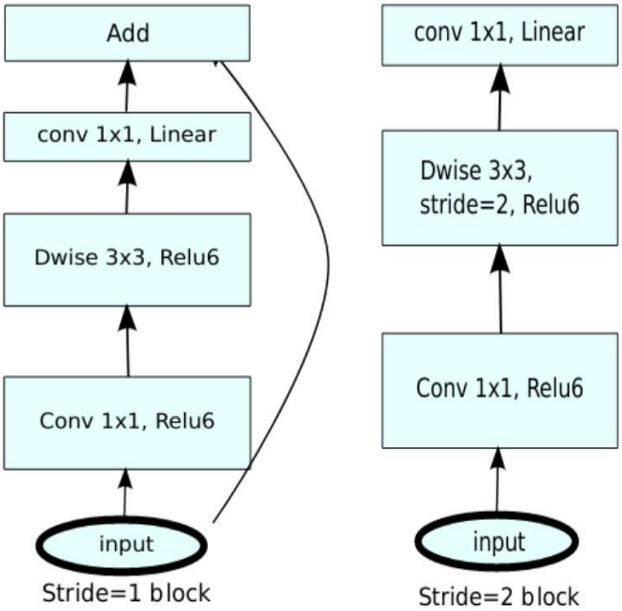
MobileNetV2 is mainly designed for mobile and embedded vision applications. It is compact yet effective. The upgraded version of MobileNetV1 is now known as MobileNetV2 [48, 49]. It is suitable for devices with limited resources, as it is designed to achieve high accuracy at a low computational cost. Two key enhancements that it suggested to expand on MobileNet's success are inverted residual blocks and linear bottleneck layers. After condensing features using a short bottleneck layer and conducting a lightweight depthwise convolution, the inverted residual block improves the feature dimension.

Depth-wise separable convolutions are used in the MobileNetV2 architecture in place of conventional convolutional layers, which are essential for computer vision applications but computationally costly. Three convolution layers and seventeen bottleneck layer blocks make up the MobileNetV2 structure, as shown in Table 4 [50]. MobileNetV2 block 1 (MVB1) and MobileNetV2 block 2 (MVB2) are two forms of conventional bottleneck layer blocks that are useful to increase system effectiveness and efficiency [51]. The MVB1 block features an upside-down residual bottleneck design with a stride of 1, as illustrated in the Figure. 5, while the MVB2 block has a bottleneck architecture with a stride 2. These bottleneck blocks utilise 3 × 3 kernels and ReLU6 activation functions to render the system resistant to low-precision computation, as shown in Figure. 7.

MobileNetV2 is a lightweight CNN architecture designed for embedded and mobile vision applications. The pre-trained MobileNetV2 model is used as the base model in the proposed approach. It retrieves the features of high-level and significant visual characteristics from a dataset without the need for its final classification layers, thereby implementing the concept of transfer learning [52]. When labelled data is scarce, this approach works especially well because it eliminates the need to train a deep network from scratch and saves a significant amount of time while preserving accuracy.

Table 4. MobileNetV2 Architecture

Layer Type	Input Size	Operator	Output Size	Kernel Size	Stride S
Initial Conv	224x224x3	conv	112x112x32	3x3	2
Inverted Residual Block	112x112x32	MVB1	112x112x16	3x3	1
Inverted Residual Block x2	112x112x16	MVB2	56x56x24	3x3	2
Inverted Residual Block x3	56x56x24	MVB2	28x28x32	3x3	2
Inverted Residual Block x4	28x28x32	MVB2	14x14x64	3x3	2
Inverted Residual Block x3	14x14x64	MVB1	14x14x96	3x3	1
Inverted Residual Block x3	14x14x96	MVB2	7x7x160	3x3	2
Inverted Residual Block x1	7x7x160	MVB1	7x7x320	3x3	1
Final Conv	7x7x320	conv 1*1	7x7x1280	1x1	1
Global Avg Pooling	7x7x1280	Avg. Pool	1x1x1280	-	-
Fully Connected	1x1x1280	conv 1*1	1x1x1000	-	-



(a) MVB1 **(b) MVB2**

Figure 7. MobileNetV2 Architecture

In this research, the MobileNetV2's last classification layers were eliminated, and the result of an earlier layer, the global average pooling layer, was used as a feature vector for single-cell images. These feature vectors are essentially condensed representations of the image's most significant patterns and visual attributes. After these features were collected, the final classification process was carried out by feeding them into ML classifiers, SVM and RF.

A lightweight CNN architecture designed for embedded and mobile vision applications is called MobileNetV2. The pre-trained MobileNetV2 model is used as the base model in the proposed approach. It retrieves the features of high-level and significant visual characteristics from a dataset without requiring its final classification layers, thereby implementing the concept of transfer learning [52]. When labelled data is scarce, this approach works especially well because it eliminates the need to train a deep network from scratch and saves a significant amount of time while preserving accuracy.

In this research, the MobileNetV2's last classification layers were eliminated and the result of an earlier layer, the global average pooling layer, as a feature vector for a single-cell images. Condensed representations of the image's most significant patterns and visual attributes are what these feature vectors are basically. After these features were collected, the final

classification process was carried out by feeding them into ML classifiers, SVM and RF.

Suppose, $I \in R^{H\times W\times C}$ is the input image which consists of height H, width W and channels C.

The convolutional layers of MobileNetV2 up to the global average pooling layer are represented by f_{MN} (I).

The extracted feature vector is denoted as $F_{MN} \in R^{1\times 1\times D}$, where D is the number of output channels.

The method of feature extraction can be expressed as follows in Eq. (1):

$$F_{MN} = GAP(f_{MN}(I)) \quad (1)$$

Here, f_{MN} (I) outputs the final convolutional feature map of size H′×W′×D

Task of Global Average Pooling (GAP) can be represented as:

$$GAP(f_{MN}(I)) = \left(\frac{1}{H'W'}\sum_{i=1}^{H'}\sum_{j=1}^{W'} f_{MN}(I)_{i,j,1}, \ldots, \frac{1}{H'W'}\sum_{i=1}^{H'}\sum_{j=1}^{W'} f_{MN}(I)_{i,j,D}\right) \quad (2)$$

This produces a 1D feature vector, $F_{MN}\in R^{D}$, which is fed into a classifier (such SVM or a fully connected layer) for the categorization of leukemia cells.

4.3.2 ResNet-101

The capacity of ResNet-101, a deep CNN with 101 layers, to learn rich hierarchical representations makes it a popular tool for extracting features in computer vision tasks. When employing ResNet-101 as a base learner to extract the features from the single cell images, the output from an intermediate layer that preserves spatial and semantic characteristics is used in place of or in addition to the final classification layers (such as the fully connected and softmax layers).

For the suggested model, the result of the global average pooling 'pool5' layer was utilized to generate a 2048-dimensional feature vector for each input image that is both compact and informative. Afterwards, this feature vector became the input for SVM and RF classifiers. By extracting high-level features from the 'pool5' layer, we converted unprocessed image into informative vectors that was used by SVM and RF to classify the cells. In addition to being computationally efficient, this hybrid strategy is also very accurate, scalable, and interpretable. This makes it very useful in clinical situations where accurate, timely, and repeatable leukemia identification is essential, providing pathologists and other medical professionals with a trustworthy decision-support tool. Figure. 8 shows the architecture of ResNet-101 [53].

Before using ResNet-101, images processed through the pre-processing stage.

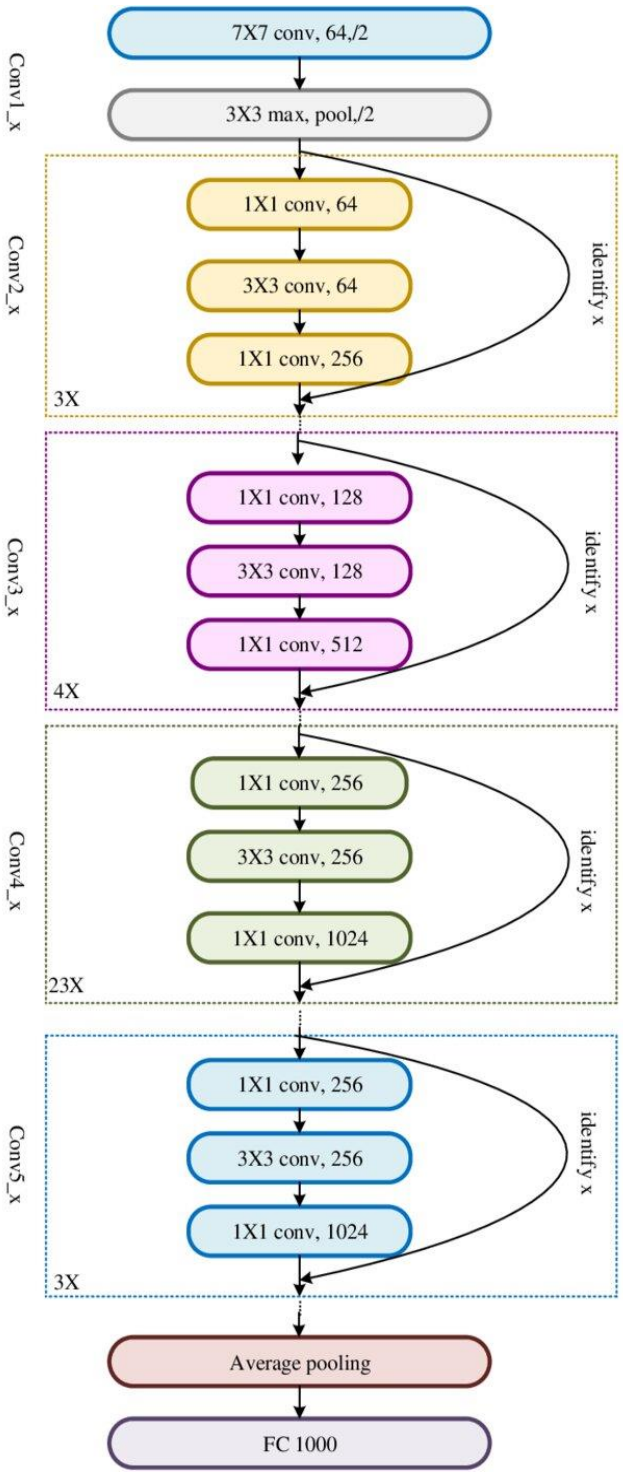


Figure 8. Architecture of ResNet-101 Convolutional Neural Network

The activations function was then used to feed the image through the pre-trained network, and chosen 'pool5' layer so that the outcome can be compatible with SVM and RF classifiers. In this manner, the suggested model processed each image. This method is very effective for applications like object recognition and medical image analysis because it makes use of the representational capacity of a deep CNN, such as ResNet-101, while allowing for flexible downstream processing with traditional SVM and RF approaches.

Suppose, $I \in R^{H\times W\times C}$ is the input image.

The convolutional layers of ResNet-101 up to the pool5 layer are represented by $f_{RN}(I)$.

The extracted feature vector is denoted as $F_{RN} \in R^{1\times1\times D}$, where D is the number of output channels.

The feature extraction process can be mathematically written as:

$$F_{RN} = AvgPool(f_{RN}(I)) \quad (3)$$

Here, the final convolutional feature map, typically 7 × 7× 2048, is $f_{RN}(I) \in R^{H'\times W'\times 2048}$.

AvgPool refers to global average pooling over spatial dimensions:

$$F_{RN} = \left(\frac{1}{H'W'}\sum_{i=1}^{H'}\sum_{j=1}^{W'} f_{RN}(I)_{i,j,1}, \ldots, \frac{1}{H'W'}\sum_{i=1}^{H'}\sum_{j=1}^{W'} f_{RN}(I)_{i,j,2048}\right) \quad (4)$$

This produces a 2048-dimensional feature vector, $F_{RN}\in R^{2048}$, which can then be passed to classifiers such as SVM and RF or fully connected network for leukemia classification.

4.3.3 VGG-16

A deep CNN architecture called VGG-16 was created by Oxford's Visual Geometry Group and is well known for its depth, generalization power, and ease of use. As the name indicates, it is a 16-layer network with fully linked and convolutional layers. A significant portion of the network's architecture is based on 3x3 convolutional filters as shown in Figure. 9 [54].

VGG-16 performs exceptionally well in extracting hierarchical characteristics from complex images. The distinctions between groups of leukemia classified according to genetic abnormalities—such as CBFB::MYH11, NPM1, RUNX1::RUNX1T1, and PML::RARA fusions—are frequently subtle and intricately linked to the minute morphological characteristics of leukemic cells. VGG-16 is very good at collecting such small patterns because it uses stacked 3×3 convolutional layers, which enable it to learn and keep fine spatial resolution features across layers. This is especially helpful for leukemia datasets, which frequently include class imbalance. Additionally, it is simpler to comprehend which characteristics are being learned because of VGG-16's comparatively shallow and interpretable structure when compared to more contemporary architectures. This is advantageous for biological insight and clinical validation. So, using VGG-16 to extract features helps the model be more accurate and easier to understand because it can tell apart visually similar leukemia subtypes linked to different genetic issues.

4.3.4 Support Vector Machine (SVM)

Support Vector Machine (SVM) is used as one of the meta-classifiers in the proposed stacking-based leukemic cells classification model. It is responsible for combining the information taken from three CNNs, MobileNetV2, ResNet-101, and VGG-1. Three pre-trained DL models, MobileNetV2, ResNet-101, and VGG-16, are renowned for their effectiveness and capacity for feature extraction. MobileNetV2 captures lightweight, fine-grained features due to its depthwise separable convolutions, which makes it ideal for mobile and real-time applications, while ResNet-101 extracts deep hierarchical features using its residual connections. The extracted features are particularly effective at learning complex patterns in medical images. The models separately generated feature vectors from each image: 1280-dimensional from MobileNetV2, 2048-dimensional from ResNet-101, and 4096-dimensional from VGG-16. These feature vectors were then compressed using the PCA method and followed by concatenation to create a single high-dimensional vector. This single high-dimensional vector served as the input to SVM and RF ensemble classifiers.

In order to determine the appropriate decision edge among the numerous leukemia classes, the combined feature vectors were used to train the SVM. Its ability to handle high-dimensional data, particularly with kernel functions like the radial basis function (RBF), makes it a great candidate for medical image classification, where feature interactions can be complex and nonlinear. In this study, SVM was extended to multiclass classification using the fitcecoc function through an Error-Correcting Output Codes (ECOC) framework.

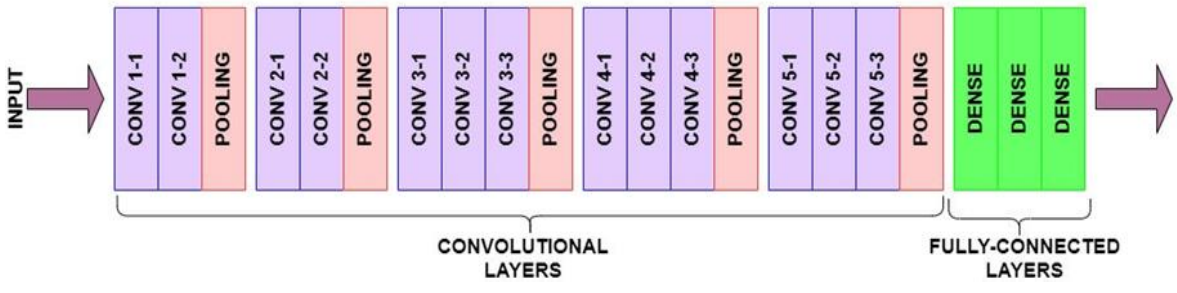


Figure 9. Architecture of VGG-16 Convolutional Neural Network

A distinct binary SVM classifier was trained for every class in order to differentiate it from every other class. We utilized the advantage of SVM's robust generalization capabilities and optimized the discriminative strength of the fused deep features from both networks by employing them as the stacking classifier. Through the integration of complementary information from two different CNNs, this architecture not only increases robustness but also improves classification accuracy.

4.3.5 Random Forest (RF)

With MATLAB's TreeBagger function, the Random Forest (RF) classifier [55], on the other hand, is trained as an ensemble of decision trees. When the feature space is broad, as is often the case with healthcare imaging data, RF is especially helpful because of its well-deserved reputation for managing intricate, nonlinear relationships in data. RF and SVM Meta learners worked on the merged and PCA-reduced deep features to differentiate between control samples and various leukemia subtypes. The majority vote or mean probability of individual trees were used to produce predictions, and bootstrap sampling was used to build numerous trees. A random collection of characteristics is chosen by RF at each node split, which enhances model generalization and lowers correlation between trees. For this research, 100 trees were used to provide forecast stability and variety.

4.4 Training Policy

In this research, hybrid architecture has been implemented. The input image with size 144 x 144 was taken from the database and resized to 224 x 244 x 3. The image size has been selected as per the convenience of the feature extractors MobilenetV2, ResNet-101, and VGG-16. Resizing the image was followed by data augmentation and oversampling techniques to deal with class imbalance issues. Subsequently, the segmentation of images was done to select the ROI. Single blood cell images then went through the next stage, feature extraction, which was carried out by MobilenetV2, ResNet-101, and VGG-16. Feature extraction was followed by the employment of Principle Component Analysis (PCA). PCA reduced the dimensionality of the feature vector by removing irrelevant features and finally, the ensemble classifier SVM and RF played the role of categorizing the images into five distinct groups.

The Adaptive Moment Estimation (Adam) optimizer, renowned for its adaptive learning skills and quick convergence, especially in complicated networks like CNNs, was used to set up the training process for the deep learning models in our system. Our mini-batch size was set at 32, which means that the network processes 32 images at a time before updating its weights. This size strikes a compromise between gradient estimate quality and memory use. To guarantee convergence, the models were trained for a maximum of 50 epochs, allowing for several runs over the complete training dataset. In order to avoid exceeding the loss minimum and to enable steady learning, the initial learning rate was set at 1e-4. The dataset was randomized at each epoch to prevent the model from remembering the data's order and to encourage generalization and prevent overfitting. To precisely mimic real-world variability, we employed a validation set of augmented images to validate the model. After every 30 mini-batches, the model's performance on this validation set was assessed, enabling regular evaluation and early identification of overfitting or under fitting during training.

Transfer learning is a potent technique that involves adapting a model that has been pre-trained on an extensive dataset to tackle a related but distinct task using usually less data and training time [56]. Typically, these networks' topmost fully connected classification layers were swapped out for new layers specifically designed for the leukemia classification task. Leukemia-specific features are learned by the model during training by modifying these new layers and perhaps parts of the subsequent convolutional layers, while keeping the useful general visual features acquired from ImageNet's millions of images.

During transfer learning, layer ‘global_average_pooling2d_1’ was used in MobileNetV2. The average of each feature map across its spatial dimensions was calculated by this layer, which takes the role of the conventional connected layers. For ResNet-101, we utilized ‘poo5’ layer and for VGG-16, ‘fc7’ has been utilized. As the last stages of feature reduction prior to categorization, these layers are crucial. These layers helped in the reduction of the feature maps' spatial dimensions produced by earlier convolutional layers to a one-dimensional vector. The existence and intensity of several learning features, including edges, textures, forms, and more intricate patterns, were efficiently encoded throughout the image by the vector, created by these layers. This was achieved by condensing intricate spatial information into a small, fixed-length feature vector that represents each input image in a reliable and discriminative manner. These derived characteristics are fed into new, task-specific classification layers on the leukemia dataset during transfer learning.

PCA is an essential in reducing dimensionality in our leukemia classification model, which makes use of three effective CNN-based feature extractors: MobileNetV2, ResNet-101, and VGG-16. High-dimensional feature vectors were produced by these three networks (e.g., 1280 for MobileNetV2, 2048 for ResNet-101, and 4096 for VGG-16), which when concatenated result in an incredibly huge and possibly

duplicate feature space. In order to overcome this difficulty, PCA projected the original high-dimensional data onto a lower-dimensional subspace while maintaining the majority of the variance, finding the directions (principal components) in which the data varies the most. By removing less informative features, this helped minimize overfitting and decreased training time and computing complexity for the ensuing SVM and RF classifiers. Through efficient feature space compression, PCA made guaranteed that only the most discriminative elements were kept, enhancing the ensemble classification framework's overall performance stability.

A strategy for ensemble learning called stacking, or stacked generalization, combines many base models to enhance predictive performance by utilizing their complementing advantages. In order to extract various feature representations from the input images, the proposed stacking technique separately trains three potent CNNs: MobileNetV2, ResNet-101, and VGG-16. The distinct architectures and depth of each network allowed it to learn different elements of the data, producing complementary predictions. These three extractors generated three distinct train feature sets and in similar way three test feature vectors, as shown in Eq. (5) and (6). Following the generation of all base models' outputs, stacking has employed an integrated classifier, which included SVM and RF, as a meta-learner, which received the inputs retrieved features from all three base models. SVM and RF jointly integrated these outputs in the best possible way to ensure the prominent accuracy of the model. Therefore, stacking allowed the proposed model to combine the benefits of MobileNetV2, ResNet-101, and VGG-16 by overcoming the shortcomings of these networks to achieve more reliable leukemia detection and classification.

$$X_{TrainCombined} = [\, f_{MobTrain,}\ f_{ResTrain,}\ f_{VggTrain}\,] \quad (5)$$

$$X_{TestCombined} = [\, f_{MobTest,}\ f_{ResTest,}\ f_{VggTest}\,] \quad (6)$$

Here, three feature vectors, namely, $f_{MobTrain,}$ $f_{ResTrain,}$ *and* $f_{VggTrain}$, were generated by MobileNetV2, ResNet-101 and VGG-16 respectively from augmented oversampled training dataset. Feature vectors also derived from testing dataset, which are $f_{MobTest,}$ $f_{ResTest, and}$ $f_{VggTest}$. These features have the important genetic information which are useful for effective classification of cells. $X_{TrainCombined}$ and $X_{TestCombined}$ processed by PCA and then served ad input to SVM and RF meta learners.

Furthermore, row-wise normalization has been employed using z-core method to the feature matrices produced by MobileNetV2, ResNet-101 and VGG-16. A feature vector that the corresponding neural network extracted from a single cell image is represented by each row of these matrices. Generally, z-score method performs column-wise normalization but when we use it with value 2, it normalizes the image on the row basis, as we have done in this research. The normalization process was carried out independently for each row. This procedure effectively projected feature vectors onto the unit hypersphere by guaranteeing that each one has the same magnitude. Z-score normalization can be presented as shown in Eq. (7):

$$N_i = \frac{i-\mu}{\sigma} \quad (7)$$

Here, i represents the image, μ is used to represent mean pixel value and σ symbolizes standard deviation.

Improving classification performance and lessening the impact of feature scale are the driving forces behind this normalization. Combining characteristics from three distinct deep networks can result in considerable differences in the dynamic range and amplitude of each network's features. In the absence of normalization, the SVM and RF might favour the network with attributes of higher magnitude. Normalization emphasizes the direction of features rather than their absolute scale, which increases the classifier's robustness and guarantees that all sets of features contribute fairly. This is especially significant in stacking technique, where balanced feature representation is essential for best results.

Utilising a powerful AI supercomputer, the "4 × NVIDIA DGX Station with four NVIDIA A100 Tensor Core GPUs and 40GB of RAM each," the suggested model has been run. MATLAB 2024a is the program used to implement the model. All of the methodologies are used in the same setting to ensure the effectiveness and fairness of the comparative study.

Overall, the extracted combined feature vectors processed by an ensemble Meta learner, which learns to predict the difference between healthy and cancerous cells and then classify them according to their genetic characteristics.

5. Results and Discussion

The proposed model differentiates healthy cell images from four malignant classes: CBFB::MYH11, NPM1, PML_RARA and RUNX1::RUNX1T1. The dataset of 81214 images was divided into three sets: training sets with 70% single cell images, validation with 15%, and 15% of testing dataset.

Model evaluation metrics for five leukemia subtypes - CBFB::MYH11, Control, NPM1, PML::RARA, and RUNX1::RUNX1T1 - are shown in the classification performance Table 5. All classes show consistently high accuracy, with values ranging from 97.78% to 98.88%, demonstrating the model's strong overall ability to produce accurate predictions. The model's accuracy, precision, and sensitivity for CBFB::MYH11 were 97.78%, 95.44%, and 94.04%, respectively, demonstrating its dependability in identifying true positive cases.

Table 5. Performance Evaluation Matrices of The Proposed Model

Category	Accuracy	Precision	Sensitivity	Specificity	F1-Score
CBFB::MYH11	97.78 %	95.44 %	94.04 %	98.79 %	94.73 %
Control	98.88 %	97.61 %	97.93 %	99.20 %	97.77 %
NPM1	98.24 %	96.75 %	95.14 %	99.11 %	95.94 %
PML::RARA	98.77 %	94.61 %	96.89 %	99.08 %	95.74 %
RUNX1::RUNX1T1	98.05 %	93.88 %	95.19 %	98.66 %	94.53 %
Overall	98.35 %	95.87 %	95.84 %	98.97 %	95.74 %

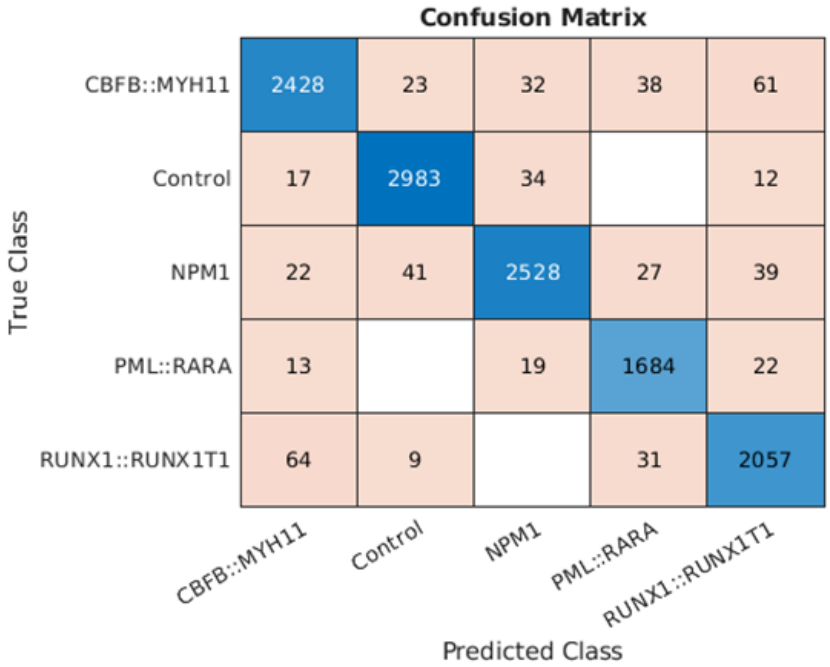


Figure 10. Confusion Matrix

Additionally, its 98.79% specificity shows that it hardly ever misclassifies other classes as CBFB::MYH11. An F1-score of 94.73% indicates a good balance between sensitivity and precision.

With the best performance of all the categories—98.88% accuracy, 97.61% precision, 97.93% sensitivity, and 99.20% specificity - the Control class showed excellent detection and few misclassifications. A 96.75% accuracy, 95.14% sensitivity, and 99.11% specificity were achieved. Despite doing marginally worse than Control, NPM1 still performs well, ensuring accurate classification. With a remarkable 96.89% sensitivity and 99.08% specificity, PML::RARA also showed excellent performance on all metrics, proving its ability to detect the majority of real cases while avoiding false positives.

Although having somewhat lower metrics than the others, RUNX1::RUNX1T1 maintained good accuracy at 98.05%, precision at 93.88%, and sensitivity at 95.19%. For every class, the model's average accuracy, sensitivity, specificity, and F1-score are 95.87%, 95.84%, 98.97%, and 98.35%, respectively. Together, these metrics show the model's strong and balanced classification performance, with low errors and dependable generalisation across a variety of leukemia subtypes.

The confusion matrix shown in Fig. 10 provides a comprehensive performance evaluation of a multi-class classification model used to distinguish between five leukemia subtypes: CBFB::MYH11, Control, NPM1, PML::RARA, and RUNX1::RUNX1T. Five classes—Control, NPM1, PML::RARA, RUNX1::RUNX1T1, CBFB::MYH11, and NPM1 are shown in this image. The genuine class is represented by each row in the confusion matrix, whereas the projected class is represented by each column. The diagonal elements for each class indicate the proportion of correctly identified examples, whereas the off-diagonal values indicate the proportion of incorrectly categorised occurrences. Out of a total of 2,582 images, 2,428 were correctly classified, and a total of 154 images were misclassified for the CBFB::MYH11 category. The highest misclassification occurred for RUNX1: RUNX1T1, with 61 cases. The proposed model achieved the highest accuracy for

healthy class images by correctly classifying 2983 images out of 3046. Collectively only 63 images were wrongly categorised for the rest of the categories. For the third category, NPM1 The model accurately classified 2,528 NPM1 single-cell images, but it had some misunderstandings in 22 images of the CBFB::MYH11 class, 41 images of the control class, 27 images of the PML_RARA class and 39 images of the RUNX1::RUNX1T1 class. Similarly, a total of 1684 images were classified correctly for the PML_RARA category, and a total of 54 images were wrongly detected for three categories: PML_RARA (13 images), NPM1 (19 images) and the highest for RUNX1::RUNX1T1 (22 images). For the last category, RUNX1::RUNX1T1, the model classified 2057 images correctly but got confused with the CBFB::MYH11 category, as it misclassified 64 images for this category.

Overall, the diagonal's high values indicate that the model does a good job of differentiating between the five groups. Misclassification across genetically or morphologically identical subtypes, particularly between CBFB::MYH11 and RUNX1::RUNX1T1. 64 images of the true RUNX1::RUNX1T1 were identified as CBFB::MYH11, whereas, 61 images of the category CBFB::MYH11 were detected as RUNX1::RUNX1T1. A shared pathogenic pathway results from the disruption of the core-binding factor complex caused by the RUNX1::RUNX1T1 and CBFB::MYH11 fusion proteins. This leads to gene expression profiles that overlap, which makes it hard for a classifier to tell them apart. Additionally, morphological connection makes it more difficult to distinguish between dysplasia and blast morphology. The NPM1-mutated AML group and the Control group also exhibit a pattern of confusion. NPM1 mutations cause cytoplasmic mislocalization of the NPM1 protein, leading to a wide spectrum of blast morphology. The "Control" group contains pre-leukemic states, causing diagnostic ambiguity, which results in misclassification. This failure suggests that feature selection or model optimisation needs to be improved. This matrix illustrates both the strengths and weaknesses of the suggested leukemia categorisation method.

In addition to this, the Figure. 11 shows a t-distributed Stochastic Neighbour Embedding (t-SNE) representation of extracted features utilized for identifying various leukemia subtypes. By projecting high-dimensional data, like the features taken from MobileNetV2, ResnNet-101, and VGG-16, into a two-dimensional space, t-SNE is a dimensionality reduction technique that allows one to visually evaluate the separability of various classes. Distinct clusters are observed for the four AML subtypes and the control group, demonstrating effective separation in the learned feature space. While some overlap is seen between subtypes with similar morphologies, the overall clustering provides evidence that the model captures biologically relevant distinctions. This visualization bridges computational predictions with clinical heterogeneity, supporting the reliability of the proposed approach.

This figure clearly shows five different clusters, each corresponding to a different subtype of leukemia or the control class. The orange cluster indicates control (healthy) samples, the blue cluster the CBFB::MYH11 fusion subtype, the yellow cluster the NPM1 mutation, the purple group the PML::RARA fusion subtype, and the green color group the RUNX1::RUNX1T1 fusion subtype. The development of distinct, non-overlapping clusters suggests that the feature extractor, which is probably a mix of CNN models like MobileNetV2, ResNet-101, and VGG-16, has learnt discriminative features that are capable of successfully separating the various classes.

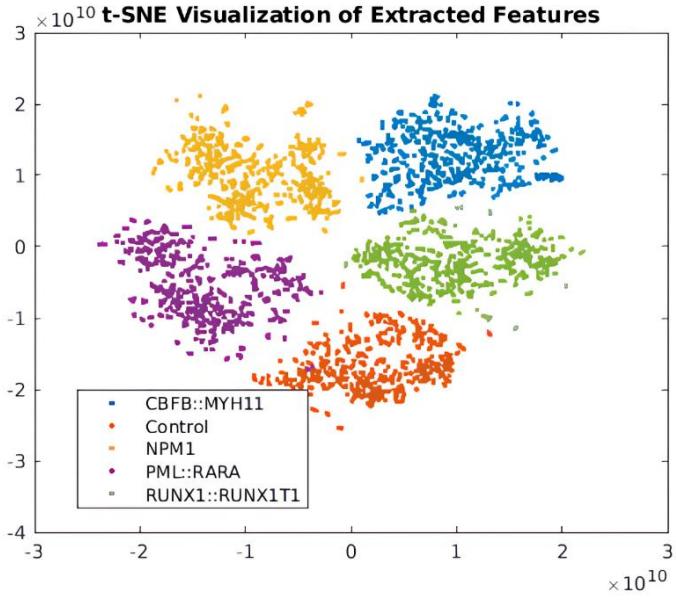


Figure 11. t-SNE Visualization

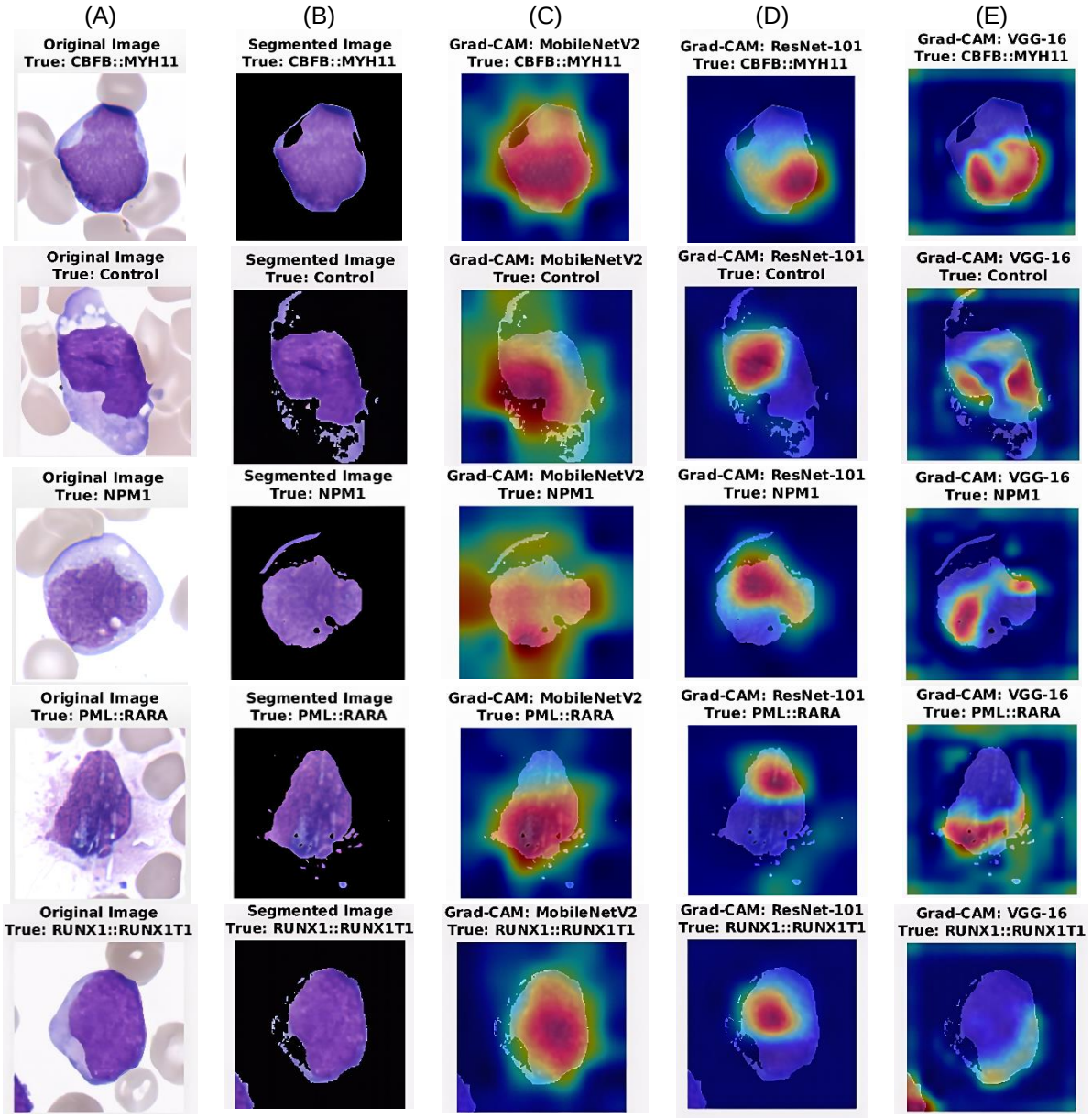


Figure 12. GRAD-CAM Images

When it comes to leukemia classification, where the visual resemblance between subtypes can make classification difficult, this is crucial in medical image analysis.

This also implies that because the input feature space is already very structured and class-specific, downstream classifiers like SVM or RF can get high accuracy. Overall, the model's capacity to distinguish between leukemic subgroups and healthy cells using learnt feature embedding is strongly qualitatively validated by this t-SNE plot.

In line with the aim of advancement, GRAD-CAM also utilized to enhance the interpretability of the model. Grad-CAM is a method for visualizing and comprehending how deep learning models, especially CNNs, make decisions. By emphasizing the areas of an input image that are crucial for a particular forecast, it successfully demonstrates the model's "reasoning" behind its results. To put it simply, it aids in determining what features the CNN is emphasizing while generating predictions. In this study, Grad-CAM helped in evaluating the extraction of pertinent features from the input images by the backbone networks, MobileNetV2, ResNet-101, and VGG-16 when applied on segmented images. The gradients of the predicted class are calculated in relation to the feature maps of the final convolutional layer in order for Grad-CAM to function. Using these gradients, each feature map's importance is weighted, creating a heatmap that shows the areas that had the biggest influence on the model's choice. Applying Grad-CAM to images that have been processed by MobileNetV2, ResNet-101 and VGG-16 allows us to see which areas of the image are being targeted by the feature extraction process and whether those characteristics correspond with areas that are important to leukemia, such as the cytoplasm or nucleus of leukemic cells. Grad-CAM heatmaps produced from

leukemic and control cell images that were accurately identified. The locations that the model used for classification are indicated by the highlighted areas mainly, yellow and red. Grad-CAM highlighted the cytoplasmic borders and nucleus in more than 85% of leukemic images, which is consistent with morphological indicators used by hematologists in manual diagnosis. This suggests that the model's decision-making is not just accurate but also therapeutically significant.

By using this visualization as a feedback mechanism, the model is able to assess the significance and relevance of the retrieved characteristics for categorization on a qualitative level. When biologically significant regions are continuously highlighted in the Grad-CAM heatmaps, it means that the feature extraction layers are working well. Grad-CAM was therefore used as a diagnostic tool to validate and improve the feature extraction procedure with the application on segmented images as illustrated in Figure. 12.

In the above given image, Figure 11 (A) represents original image, its corresponding segmented image is being displayed in (B). Figure 9 (C), (D), and (E) provide the GRAD-CAM visualization for employed networks: MobileNetV2, ResNet-101, and VGG-16 respectively.

The medical significance of our concept is further supported by the interpretability analysis. Since nuclei and cytoplasmic boundaries are recognised morphological markers of leukemic transformation, they are regularly highlighted in the Grad-CAM maps. The system's clinical usefulness is more confidently affirmed by this agreement between the model emphasis and expert visual criteria. In addition to achieving high classification performance, our framework's visual explanations provide transparent insights—two crucial components for clinical adoption of AI-driven diagnostic solutions.

Furthermore, a thorough assessment of the diagnostic performance of classification models is offered by the ROC (Receiver Operating Characteristic) curve. This study compared the classification competencies of five models by generating and analysing their ROC curves: the proposed hybrid model, the individual Vision Transformer, ResNet-101, MobileNetV2, and VGG-16 models. In an ROC curve, the y-axis represents sensitivity (true positive rate), and the x-axis represents 1-Specificity (false positive rate). This curve enables the comparison of several models in addition to aiding in the assessment of the trade-offs between sensitivity and specificity. As a result, depending on their unique needs and circumstances, practitioners can choose models with knowledge. The suggested model demonstrated exceptional discriminative abilities, with the highest Area under the Receiver Operating Characteristic (AROC) value of 0.97, as shown in Figure. 13.

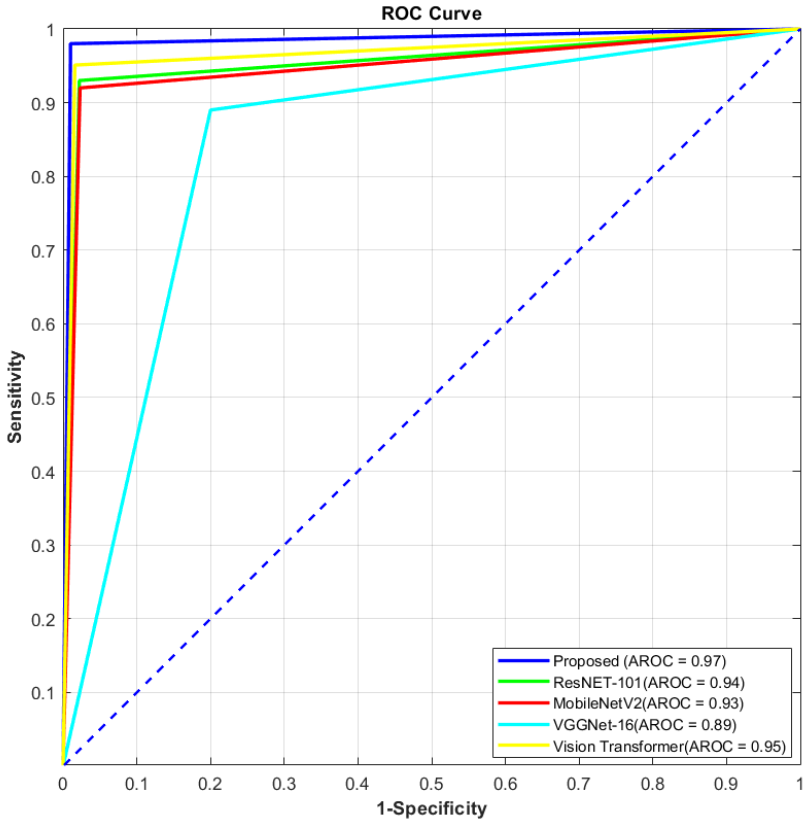


Figure 13. Comparison of ROC Curve among Individual Models and Proposed Model

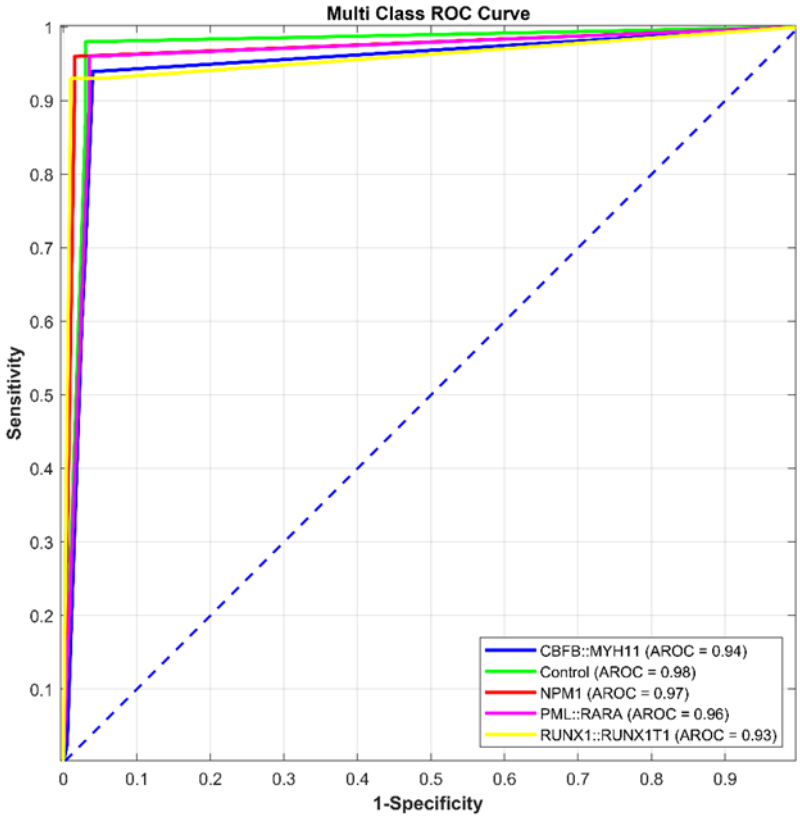


Figure 14. Multi Class ROC Curve for Proposed Model

Table 6. Performance Evaluation Matrices Of The Ablation Study

Experiments	Accuracy	Precision	Sensitivity	Specificity	F1-Score
Baseline Model	57.74 %	88.42 %	74.42 %	92.82 %	90.57 %
+ Segmentation	61.05 %	90.22 %	82.69%	86.01 %	88.06 %
+ Segmentation + Augmentation	63.9 %	89.06 %	81.58 %	87.24 %	88.14 %
+ Segmentation + Augmentation + Oversampling	83.10%	93.80%	94.02%	95.25%	94.52%
+ Segmentation + Augmentation + Oversampling + PCA (Proposed Model)	98.35 %	95.87 %	95.84 %	98.97 %	95.74 %

Vision Transformer came in second, with an AROC of 0.95, then ResNet-101 came in third with an AROC of 0.93, while MobileNetV2 and VGG-16 performed marginally worse, with AROCs of 0.92 and 0.89, respectively. With a curve closest to the top left, the proposed ensemble stack model outperformed the others in classification, demonstrating its improved capacity to reduce false positives while retaining high sensitivity.

Additionally, ROC curve extended to five classes. Figure. 14 shows the ROC curve for the five different classes of the proposed model. All categories have high AROC values, with Control having the best discrimination performance (AROC = 0.98), followed by NPM1 (0.97), PML::RARA (0.96), CBFB::MYH11 (0.94), and RUNX1::RUNX1T1 (0.93). All of these AROC values above 0.90 show outstanding classification accuracy, proving that the model is quite trustworthy in differentiating between control and multiple leukemia subtypes.

In continuation, McNemar's test has been carried out in between the suggested hybrid model with vision transformer on the test dataset of 12184 blood smear images. McNemar's test is used to statistically determine whether the performance difference between the two models is significant or no

Table 7. Performance Evaluation Matrices of the Proposed Model

Category	No. of Images	Accuracy	Precision	Sensitivity	Specificity	F1-Score
CBFB::MYH11	504	98.19 %	95.95 %	93.42 %	99.18 %	94.67 %
Control	689	98.41 %	95.33 %	98.08 %	98.52 %	96.68 %
NPM1	600	98.42 %	96.63 %	95.56 %	99.15 %	96.09 %
PML::RARA	489	99.09 %	96.05 %	98.65 %	99.18 %	97.33 %
RUNX1::RUNX1T1	643	98.19 %	96.84 %	94.85 %	99.13 %	95.83 %
Overall	2925	98.46 %	96.16 %	96.11 %	99.03 %	96.12 %

In our model, the proposed model and vision transformer architecture properly classified 11028 images, whereas, both frameworks misclassified around 487 blood smear images. However, differences in the performance emerged in the remaining cases: when suggested model correctly classified 383 images that were misclassified by vision transformer model, while vision transformer framework accurately categorized 286 images that were misclassified by suggested model. This classification leads to the result $\chi 2 = 14.06$. After that, this value is contrasted with the critical value derived from the chi-square distribution for a certain degree of freedom and significance level (often $\alpha = 0.05$). Since there is just one degree of freedom in McNemar's test (df =1), the appropriate critical value at df=0.05 is 3.841 and the critical value for $\alpha=0.001$ is 10.828.

Since, $\chi 2 = 14.06 > 10.828$, the p-value (0.000176) is less than 0.001. This implies that the performance difference between the suggested model and vision transformer is statistically significant, meaning the suggested model and vision transformer model perform differently. This demonstrates that the proposed model likely provides a genuine performance improvement over vision transformer rather than differences arising due to chance.

Furthermore, we conducted an ablation study of the offered model through a series of experiments, adding and removing components to evaluate the contribution of different techniques employed in the model's development. First, the experimental framework starts with a baseline model that includes a group of feature extractors (MobileNetV2, ResNet-101, and VGG-16) and classifiers (SVM, RF). The impact of sophisticated preprocessing techniques, such as data augmentation, image segmentation, and oversampling, will then be thoroughly assessed by gradually integrating them into this baseline. Eventually, the inclusion of PCA has been shown. The outcomes of the various experiments with different configurations are depicted in Table 6.

The importance of each element to the overall effectiveness of the suggested leukemia classification system is highlighted in the ablation research shown in Table 6. The outcomes illustrate how various preprocessing and augmentation methods affect the leukemia classification system's functionality. The accuracy of the baseline model was 57.74%. The model displayed unbalanced behaviour with high specificity (92.82%) but low sensitivity (74.42%), although having a comparatively high precision (88.42%) and F1-score (90.57%). This suggests that while the model did a good job of identifying negative samples, it had trouble accurately classifying positive ones. With the introduction of segmentation, overall performance increased, improving sensitivity to 82.69% and accuracy to 61.05%. This demonstrates how segmentation aids the model in concentrating on the pertinent areas of the images, improving the identification of true positives. With the addition of data augmentation, accuracy increased to 63.9% while sensitivity (81.58%) and specificity (87.24%) remained balanced. This demonstrates that by decreasing overfitting, augmentation improved the model's robustness. Performance significantly improved with the addition of oversampling, with accuracy reaching 83.10%, precision 93.80%, sensitivity 94.02%, specificity 95.25%, and F1-score 94.52%. This enhancement shows how oversampling can effectively fix the dataset's class imbalance. With 98.35% accuracy, 95.87% precision, 95.84% sensitivity, 98.97% specificity, and 95.74% F1-score, the suggested model was the best outcome of using PCA on top of segmentation, augmentation, and oversampling. These findings demonstrate that the discriminative power and generalisation of the classification system are significantly improved by the combination of segmentation, augmentation, oversampling, and PCA.

Additionally, the validation of the proposed model has been conducted using a subset of the Human Leukemia Cytomorphology Collection dataset. A few images were randomly selected from each category. The number of images per category and the obtained result is shown in Table 7 that confirms that the suggested model performed well on the subset as well.

The outcomes demonstrate how the suggested model reliably separates healthy (control) and leukemic

blood smear images. The model can also distinguish single-cell images into four subtypes of leukemic cells: CBFB::MYH11, NPM1, PML::RARA, and RUNX1::RUNX1T1. This proposed model can be highly beneficial to pathologists and medical experts in the detection and classification of leukemia.

6. Conclusion and Future Work

In this research, the effectiveness of our leukemia detection and classification algorithm is demonstrated, showing that it can reliably identify and classify AML subtypes based on underlying genetic alterations. In conjunction with SVM and RF classifiers, the model utilises sophisticated deep learning-based feature extractors, including MobileNetV2, ResNet-101, and VGG-16, to efficiently identify subtle patterns and morphological variations in blood cell images that are indicative of particular genetic disorders. Normalisation, data augmentation, image segmentation, and oversampling are crucial preprocessing steps that significantly enhance the model's generalisation across a diverse range of patient samples. The model's capacity to distinguish between leukemia subtypes associated with certain genetic fusions or mutations, such as PML::RARA, NPM1, CBFB::MYH11, and RUNX1::RUNX1T1, is the basis for the categorisation. This genetic-level separation via image-based analysis and comparative study of various individual CNN models and their outcomes in the distinction of the various leukemic and benign classes indicates the proposed hybrid, ensemble stacking model's potential as a useful diagnostic tool in individualized leukemia treatment by assisting hematologists in accurate and significant predictions with fewer efforts and also saving time.

We describe the following potential directions to turn this intriguing research from a proof-of-concept into a clinically useful tool:

- Multi-Center External Validation: Using data from distinct organisations, a multi-center validation research is the most important next step. This will thoroughly evaluate the model's clinical performance, generalisability, and resilience to real-world unpredictability.
- Integration with genetic testing Pipelines: We intend to investigate how this image-based model can be integrated with other diagnostic modalities, specifically pipelines for genetic testing (e.g., FISH, PCR, NGS). A more thorough and conclusive diagnosis might be obtained by creating a fused diagnostic framework that integrates morphological, genomic, and clinical data. This would open the door for really integrated, AI-powered precision medicine in hematology.
- Edge-Device Deployment for Clinical Labs: We will concentrate on optimising the model for deployment on edge devices in clinical labs and point-of-care settings in order to optimise its practical utility. To enable real-time analysis on affordable devices, this entails model compression, quantisation, and hardware-software co-design. This ensures quick results without depending on online connectivity and addresses data privacy concerns.
- Enhanced Validation via k-Fold Cross-Validation: : Before moving forward with extensive external trials, we will use k-fold cross-validation on larger datasets as part of our immediate experimental roadmap to enhance the stability of the model, lower variability, and further solidify the performance estimates.

By implementing these directives, we aim to transform our study into a deployable, proven system that significantly advances and standardises leukemia diagnoses and classification.

References

[1] E. Aby, S. Salaji, K. K. Anilkumar, T. Rajan, A review on leukemia detection and classification using Artificial Intelligence-based techniques. Computers and Electrical Engineering, 118(B), (2024) 109446. https://doi.org/10.1016/j.compeleceng.2024.109446

[2] K. Venkatesh, S. Pasupathy, S.P. Raja, Acute myeloid leukemia multi-classification using enhanced few-shot learning technique. Scalable Computing: Practice and Experience, 23(4), (2022) 377–388. https://doi.org/10.12694/scpe.v23i4.2048

[3] M.A. Alsalem, A.A. Zaidan, B.B. Zaidan, M. Hashim, H.T. Madhloom, N.D. Azeez, S. Alsyisuf, A review of the automated detection and classification of acute leukaemia: Coherent taxonomy, datasets, validation and performance measurements, motivation, open challenges and recommendations. Computer Methods and Programs in Biomedicine, 158, (2018) 93–112. https://doi.org/10.1016/j.cmpb.2018.02.005

[4] X. Parisi, J.R. Bledsoe, Diagnosis of pediatric acute myeloid leukemia and mixed-phenotype acute leukemia. Diagnostic Histopathology, 31(10),(2025) 531-563. https://doi.org/10.1016/j.mpdhp.2025.07.001

[5] National Cancer Institute. (2017) Cancer stat facts: Acute myeloid leukemia (AML), SEER, [Online]. Available: https://seer.cancer.gov/statfacts/html/amyl.html.

[6] L. Faivdullah, F. Azahar, Z.Z. Htike, W.N. Naing, Leukemia detection from blood smears. Journal of Medical and Bioengineering, 4(6), (2015) 488–491.

[7] H.S. Newton, M.A. Dobrovolskaia, Immunophenotyping: Analytical approaches and role in preclinical development of nanomedicines. Advanced Drug Delivery Reviews, 185, (2022) 114281. https://doi.org/10.1016/j.addr.2022.114281

[8] A. Kumar, V. Tilak, A. Ali. Cytogenetic analysis of chronic myeloid leukemia in an eastern Indian population. International Journal of Clinical and Medical Education Research, 2(2), (2023) 37–42.

[9] A. Sahu, P.K. Das, I. Paul, S. Meher, A hybrid deep learning framework for automatic detection of brain tumours using different modalities. IEEE Transactions on Emerging Topics in Computational Intelligence, 9(2), (2025) 1216–1225. https://doi.org/10.1109/TETCI.2024.3442889

[10] S. Dodia, A.B., P.A. Mahesh, Recent advancements in deep learning based lung cancer detection: A systematic review. Engineering Applications of Artificial Intelligence, 116, (2022) 105490. https://doi.org/10.1016/j.engappai.2022.105490

[11] E.H. Aktekin, M.B. Çöteli, A. Erbay, N. Yazici, A prospective study for the examination of peripheral blood smear samples in pediatric population using artificial intelligence. Turkish Journal of Medical Sciences, 55(2), (2025) 386–397. https://doi.org/10.55730/1300-0144.5982

[12] C. Matek, S. Schwarz, K. Spiekermann, and C. Marr, Human-level recognition of blast cells in acute myeloid leukaemia with convolutional neural networks. Nature Machine Intelligence, 1, (2019) 538–544. https://doi.org/10.1038/s42256-019-0101-9

[13] C. Matek, S. Krappe, C. Münzenmayer, T. Haferlach, C. Marr, Highly accurate differentiation of bone marrow cell morphologies using deep neural networks on a large image dataset, Blood. The Journal of the American Society of Hematology, 138(20), (2021) 1917-1927. https://doi.org/10.1182/blood.2020010568

[14] J.N. Kather, J. Krisam, P. Charoentong, T. Luedde, E. Herpel, C.A. Weis, T. Gaiser, A. Marx, N.A. Valous, D. Ferber, L. Jansen, Predicting survival from colorectal cancer histology slides using deep learning: A retrospective multicenter study. PLoS Medicine, 16(1), (2019) e1002730. https://doi.org/10.1371/journal.pmed.1002730

[15] E. Arvaniti, K.S. Fricker, M. Moret, N. Rupp, T. Hermanns, C. Fankhauser, N. Wey, P.J. Wild, J.H. Rueschoff, M. Claassen, Automated Gleason grading of prostate cancer tissue microarrays via deep learning. Scientific reports, 8, (2018) 12054. https://doi.org/10.1038/s41598-018-30535-1

[16] G. Campanella, M.G. Hanna, L. Geneslaw, A. Miraflor, V. Werneck Krauss Silva, K.J. Busam, E. Brogi, V.E. Reuter, D.S. Klimstra, T.J. Fuchs, Clinical grade computational pathology using weakly supervised deep learning on whole slide images. Nature medicine, 25(8), (2019) 1301-1309. https://doi.org/10.1038/s41591-019-0508-1

[17] M.F. Pittenger, C. Kerr. (2014) Stem Cells, Tissue Engineering (Second Edition), 23–65. https://doi.org/10.1016/B978-0-12-420145-3.00002-X

[18] Q. Cai, H. Lan, Y. Deng, B. Xian, Z. Luo, J. Li, Z. Liao, Flow cytometry in acute myeloid leukemia and detection of minimal residual disease. Clinica Chimica Acta, 564, (2025) 119945. https://doi.org/10.1016/j.cca.2024.119945

[19] College of American Pathologists, What's new in AML classification: WHO 2022 vs. International Consensus Classification. [Online]. Available: https://www.cap.org/member-resources/articles/whats-new-in-aml-classification-who-2022-vs-international-consensus-classification

[20] X. Zhang, J. Sun, W. Yu, J. Jin, Current views on the genetic landscape and management of variant acute promyelocytic leukemia. Biomarker Research, 9(1), (2021) 33. https://doi.org/10.1186/s40364-021-00284-x

[21] G. Cavallaro, S. Salmoiraghi, R. Cavagna, C. Pavoni, C. Belotti, M. Milani, E. Oldani, M. Frigeni, T. Intermesoli, A. Grassi, R. Bassan, Clinical impact of tyrosine kinases related mutations in NPM1 mutated acute myeloid leukaemia: A report from the Northern Italy Leukaemia Group (NILG) multicentre randomized trial 02/06, Leukemia Research, 153, (2025) 107702. https://doi.org/10.1016/j.leukres.2025.107702

[22] R.A.H. Abo Elwafa, M.M. ElBordiny, A. Deghedy, A. Elghandour, E.T. Ammar, The fusion genes and their relation with genetic variants in Egyptian AML patients. Asian Pacific Journal of Cancer Prevention: APJCP, 26 (3), (2025) 1069–1078. https://doi.org/10.31557/APJCP.2025.26.3.1069

[23] M. Gerritsen, R. Knops, B.L. Mandos, M. Decker, T. Ripperger, B.A. van der Reijden, J.H. Martens, J.H. Jansen, Acute myeloid leukemia associated RUNX1 variants induce aberrant expression of transcription factor TCF4. Leukemia, 39(2), (2025) 520–523. https://doi.org/10.1038/s41375-024-02470-w

[24] F. Lagunas-Rangel, V. Chávez-Valencia, M. Gómez-Guijosa, C. Cortés-Penagos, Acute myeloid leukemia—Genetic alterations and their clinical prognosis. International journal of hematology-oncology and stem cell research, 11(4), (2017) 328–329.

[25] V. Acharya, V. Ravi, T.D. Pham, C. Chakraborty, Peripheral blood smear analysis using automated computer-aided diagnosis system to identify acute myeloid leukemia. IEEE Transactions on

Engineering Management, 70(8), (2023) 2760–2773. https://doi.org/10.1109/TEM.2021.3103549

[26] J. Rawat, A. Singh, J. Virmani, J.S. Devgun, Computer assisted classification framework for prediction of acute lymphoblastic and acute myeloblastic leukemia. Biocybernetics and Biomedical Engineering, 37(4), (2017) 637–654. https://doi.org/10.1016/j.bbe.2017.07.003

[27] T. Jiang, J.L. Gradus, A.J. Rosellini, Supervised machine learning: A brief primer. Behavior Therapy, 51(5), (2020) 675–687. https://doi.org/10.1016/j.beth.2020.05.002

[28] K.L. Narayanan, R.S. Krishnan, Y.H. Robinson, S. Vimal, T.A. Rashid, C. Kaushal, M.M. Hassan, Enhancing acute leukemia classification through hybrid fuzzy C means and random forest methods. Measurement: Sensors, 39, (2025) 101876. https://doi.org/10.1016/j.measen.2025.101876

[29] M. Shaheen, R. Khan, R.R. Biswal, M. Ullah, A. Khan, M.I. Uddin, M. Zareei, A. Waheed, Acute myeloid leukemia (AML) detection using AlexNet model, Complexity, 2021, (2021) 1–8. https://doi.org/10.1155/2021/6658192

[30] J. Zhao, M. Zhang, Z. Zhou, J. Chu, F. Cao, Automatic detection and classification of leukocytes using convolutional neural networks. Medical & biological engineering & computing, 55(8), (2017) 1287–1301. https://doi.org/10.1007/s11517-016-1590-x

[31] K. Liu, J. Hu, Classification of acute myeloid leukemia M1 and M2 subtypes using machine learning. Computers in Biology and Medicine, 147, (2022) 105741. https://doi.org/10.1016/j.compbiomed.2022.105741

[32] W. Yu, J. Chang, C. Yang, L. Zhang, H. Shen, Y. Xia, J. Sha, Automatic classification of leukocytes using deep neural network. In 2017 IEEE 12th international conference on ASIC (ASICON), IEEE, Guiyang, China, (2017) 1041–1044. https://doi.org/10.1109/ASICON.2017.8252657

[33] B. Subrahmanyeswara Rao, Accurate leukocoria predictor based on deep VGG-net CNN technique. IET Image Processing, 14(10), (2020) 2241–2248. https://doi.org/10.1049/iet-ipr.2018.6656

[34] N. Veeraiah, Y. Alotaibi, A.F. Subahi, MayGAN: Mayfly optimization with generative adversarial network-based deep learning method to classify leukemia from blood smear images. Computer Systems Science and Engineering, 46(2), (2023) 2039–2058. https://doi.org/10.32604/csse.2023.036985

[35] K. Venkatesh, S. Pasupathy, S.P. Raja, Acute myeloid leukemia multi-classification using enhanced few-shot learning technique. Scalable Computing: Practice and Experience, 23(4), (2022) 377–388. https://doi.org/10.12694/scpe.v23i4.2048

[36] L. Boldú, A. Merino, A. Acevedo, A. Molina, J. Rodellar, A deep learning model (ALNet) for the diagnosis of acute leukaemia lineage using peripheral blood cell images. Computer Methods and Programs in Biomedicine, 202, (2021) 105999. https://doi.org/10.1016/j.cmpb.2021.105999

[37] R. Agrawal, S. Satapathy, G. Bagla, K. Rajakumar. (2019) Detection of white blood cell cancer using image processing. In 2019 International Conference on Vision Towards Emerging Trends in Communication and Networking (ViTECoN), Vellore, India, 1–6. https://doi.org/10.1109/ViTECoN.2019.8899602

[38] F. Qin, N. Gao, Y. Peng, Z. Wu, S. Shen, A. Grudtsin, Fine-grained leukocyte classification with deep residual learning for microscopic images. Computer methods and programs in biomedicine, 162, (2018) 243–252. https://doi.org/10.1016/j.cmpb.2018.05.024

[39] N. Rohaziat, M.R.M. Tomari, W.N.W. Zakaria, (2022) White blood cells type detection using YOLOv5. In 2022 IEEE 5th international symposium in robotics and manufacturing automation (ROMA), IEEE, 1–6. https://doi.org/10.1109/ROMA55875.2022.9915690

[40] J.N. Eckardt, J.M. Middeke, S. Riechert, T. Schmittmann, A.S. Sulaiman, M. Kramer, K. Sockel, F. Kroschinsky, U. Schuler, J. Schetelig, C. Röllig, Deep learning detects acute myeloid leukemia and predicts NPM1 mutation status from bone marrow smears. Leukemia, 36(1), (2022) 111–118. https://doi.org/10.1038/s41375-021-01408-w

[41] L. Zare, M. Rahmani, N. Khaleghi, S. Sheykhivand, S. Danishvar, Automatic detection of acute leukemia (ALL and AML) utilizing customized deep graph convolutional neural networks. Bioengineering, 11(7), (2024) 644. https://doi.org/10.3390/bioengineering11070644

[42] N.K. Ramesh, V. Kadappa, R.D.D. Veerabhaskar, D.R. Sirigiri, P.T. Sreedhar, Deep learning architectures for detection of acute myeloid leukemia. International Journal of Computational Science and Engineering (IJCSE), 28(3), (2025) 278–291.

[43] V.J. Ramya, S. Lakshmi, Acute myelogenous leukemia detection using optimal neural network based on fractional black-widow model. Signal, Image and Video Processing, 16(1), (2022) 229–238. https://doi.org/10.1007/s11760-021-01976-5

[44] A.E. Aby, S. Salaji, K.K. Anilkumar, T. Rajan, Classification of acute myeloid leukemia by pre-trained deep neural networks: A comparison with

different activation functions. Medical Engineering & Physics, 135, (2025) 104277. https://doi.org/10.1016/j.medengphy.2024.104277

[45] H. Patel, H. Shah, G. Patel, A. Patel, Hematologic cancer diagnosis and classification using machine and deep learning: State-of-the-art techniques and emerging research directives. Artificial Intelligence in Medicine, 152, (2024) 102883. https://doi.org/10.1016/j.artmed.2024.102883

[46] P.K. Das, A. Sahu, S. Meher, R. Panda, A. Abraham, Enhanced detection of acute leukemia: A hybrid machine learning framework with adaptive weight-optimized level set evolution. Engineering Applications of Artificial Intelligence, 161, Part C, (2025) 112244. https://doi.org/10.1016/j.engappai.2025.112244

[47] M. Hehr, A. Sadafi, C. Matek, P. Lienemann, C. Pohlkamp, T. Haferlach, K. Spiekermann, C. Marr, Explainable AI identifies diagnostic cells of genetic AML subtypes. PLOS Digit. Health, 2(3), (2023) e0000187. https://doi.org/10.1371/journal.pdig.0000187

[48] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, L.C. Chen, (2018) MobilenetV2: Inverted residuals and linear bottlenecks, In Proceedings of the IEEE conference on computer vision and pattern recognition, Salt Lake City, UT, USA, 4510–4520. https://doi.org/10.1109/CVPR.2018.00474

[49] M.S A.P.U, Comparative Analysis of Suitability of Deep Learning Models in Quality Assurance of Fabrics. International Research Journal of Multidisciplinary Technovation, 7(4), (2025) 41–57. https://doi.org/10.54392/irjmt2544

[50] Q. Zhu, H. Zhuang, M. Zhao, S. Xu, R. Meng, A study on expression recognition based on improved MobileNetV2 network. Scientific Reports, 14(1), (2024) 8121.

[51] R. Shah, J. Shastri, M.H. Bohara, B.Y. Panchal, P. Goel. (2022) Detection of different types of blood cells: A comparative analysis. In 2022 IEEE International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE),IEEE, Ballari, India, 1–6. https://doi.org/10.1109/ICDCECE53908.2022.9793132

[52] S.T. Krishna, H.K. Kalluri, Deep learning and transfer learning approaches for image classification. International Journal of Recent Technology and Engineering (IJRTE), 7(5S4), (2019) 427–432.

[53] R. Kumari, J. Wasim, A deep learning approach for human facial expression recognition using residual network–101. Journal of Current Science and Technology. 13(3), (2023) 517–532. 2023. https://doi.org/10.59796/jcst.V13N3.2023.2152

[54] LearnDataSci, (2023) Hands-on transfer learning with Keras, [Online]. https://www.learndatasci.com/tutorials/hands-on-transfer-learning-keras/

[55] N. Syam, R. Kaul, Random forest, bagging, and boosting of decision trees, in Machine Learning and Artificial Intelligence in Marketing and Sales: Essential Reference for Practitioners and Data Scientists. Emerald Publishing Ltd., (2021) 139–182.

[56] M.A. Shames, M.Y. Kamil, Early Diagnosis of Lung Infection via Deep Learning Approach. International Research Journal of Multidisciplinary Technovation, 6(3), (2024) 216–224. https://doi.org/10.54392/irjmt24316

Authors Contribution Statement

Hema Patel: Conceptualization, Methodology, Formal Analysis, Software, Investigation, Writing- Original Draft. Himal Shah: Validation. Gayatri Patel: Supervision, Writing - Review and Editing. Atul Patel: Resources, Validation, Supervision. All the authors have read and agreed to the published version of the manuscript.

Funding

The authors declare that no funds, grants or any other support were received during the preparation of this manuscript.

Competing Interests

The authors declare that there are no conflicts of interest regarding the publication of this manuscript.

Data Availability

The data supporting the findings of this study can be obtained from the corresponding author upon reasonable request.

Has this article screened for similarity?

Yes

About the License