IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 11 March 2024, accepted 8 April 2024, date of publication 15 April 2024, date of current version 23 April 2024.

*Digital Object Identifier 10.1109/ACCESS.2024.3388715*

RESEARCH ARTICLE

# Feature Fusion Based Ensemble of Deep Networks for Acute Leukemia Diagnosis Using Microscopic Smear Images

**MD HASIB AL MUZDADID HAQUE HIMEL[1], MD. AL MEHEDI HASAN[1], (Member, IEEE), TARO SUZUKI[2], AND JUNGPIL SHIN[2], (Senior Member, IEEE)**

[1]Department of Computer Science & Engineering, Rajshahi University of Engineering & Technology (RUET), Rajshahi 6204, Bangladesh
[2]School of Computer Science and Engineering, The University of Aizu, Aizuwakamatsu 965-8580, Japan

Corresponding author: Jungpil Shin (jpshin@u-aizu.ac.jp)

This work was supported by the Competitive Research Fund of The University of Aizu, Japan.

**ABSTRACT** A type of blood malignancies known as leukemia leads to elevated quantities of abnormally formed blood cells and often originates in the bone marrow. An abrupt rise in the quantity of immature blood cells is one of the distinctive characteristics of acute leukemia, that predominantly affects both children and adults. The possibility exists to drastically lower the death rate associated with acute leukemia through early detection and diagnosis. A significant and time-consuming task for the early detection of acute leukemia is microscopic examination of blood cells. In this paper, a two-staged deep learning-based computer-aided diagnosis system regarding microscopic blood smear images is proposed to assist hematologists and improve the diagnosis accuracy of acute leukemia in which for the feature fusion-based ensemble, two deep neural network branches adopting pretrained, fine-tuned EfficientNetB7 and MobileNetV3Large architectures were employed, and feature maps generated from those branches were fused and fed to the second stage of the architecture to achieve the final result. Additional dropout layers and ReLu activation were employed in the architecture to speed up the network, and compound scaling, bottleneck, and fusion architectures enhanced the overall performance. The ALLIDB1, ALLIDB2, and ASH databases were incorporated to evaluate the performances of the proposed method. The experimental findings demonstrated that the proposed approach detected acute leukemia (ALL, AML, Healthy) with an accuracy of 99.3%, F1-score of 99.3%, and AUC score of 0.997. Whereas in detecting acute lymphocytic leukemia (ALL, Healthy) and acute myeloid leukemia (AML, Healthy), the accuracies reached 100.0% and 99.8%, respectively. Thus, we believe that clinics can adopt our proposed architecture for quick and automated diagnosis.

**INDEX TERMS** Acute leukemia detection, deep learning, transfer learning, feature fusion-based ensembling, microscopic smear image.

## I. INTRODUCTION

Leukemia is a type of blood malignancy that often develops in the bone marrow and affects the replication process of white blood cells (WBC), resulting in a large number of abnormal white blood cells and loss of the human immune system, regardless of age. It has been anticipated that roughly 62,770 new leukemia cases will be recorded, which will cause the death of 23,670 individuals in 2024 according to the American Cancer Society [1]. Acute leukemia develops rapidly, and crowding caused by the rapid increase in the number of immature blood cells prevents the bone marrow from producing healthy blood cells, resulting in low hemoglobin and platelets, which then spill into the

The associate editor coordinating the review of this manuscript and approving it for publication was Chao Zuo.

bloodstream and spread to other organs. Acute Leukemia is further classified as Acute Lymphocytic Leukemia (ALL) and Acute Myeloid Leukemia (AML) in accordance with the French-American-British (FAB) classification model [2]. ALL is the most prevalent kind of leukemia in young kids, whereas AML affects adults significantly more than kids, and men far more than women. Early diagnosis and appropriate treatment may help to save lives. However, manual diagnosis of acute leukemia is a time-consuming and error-prone process that requires highly qualified hematologists with extensive expertise to make an accurate early diagnosis. Furthermore, this process is quite challenging due to the complex structure of blood cells, the existence of noise, blurriness, intensity inhomogeneity, weakly defined edges, and cell overlapping [3], [4], [5], [6].

Many studies have recently been published on the use of machine learning and deep learning to classify microscopic blood smear images and develop effective computer-aided diagnosis (CAD) systems for the early identification and diagnosis of acute leukemia [3], [7], [8], [9], [10], [11], [12], [13]. To categorize and identify leukemia using microscopic smear images, a number of machine learning methods have been proposed earlier. The most often used approach is Support Vector Machine (SVM), which is followed by K-Nearest Neighbor (KNN), Random Forest algorithm, and other classification methods [3]. Many researchers have proposed preprocessing and segmentation strategies before performing classification [3], [9], [14], [15], [16], [17]. These approaches have so far yielded a significant false-positive rate. To overcome the constraints, numerous researchers have proposed deep learning methodologies since deep learning architectures can extract complicated features directly from raw images. Because of the capability to extract deep as well as multiscaled characteristics and combine them to aid hematologists in making the final prediction, deep learning techniques have shown to be more effective than conventional approaches for object recognition and classification in several applications. The most frequently employed deep learning architecture for dealing with medical image classification problems is the convolutional neural network (CNN) [18]. Additionally, a significant number of cutting-edge architectures have been developed and are offering remarkable results on medical images. Despite the fact that these deep learning approaches have solved multiple constraints, training deep learning models on inadequate data frequently results in overfitting. Multiple strategies have been proposed to address this problem. Because of their high degree of efficiency, deep learning algorithms are rapidly adopting data augmentation strategies [19]. The most popular classical augmentation techniques, such as rotating, flipping, scaling, etc., provide new images with the same semantic information as the original ones. Additionally, Generative Adversarial Networks (GANs) have been employed in the medical field for augmentation by generating artificial images [20], [21].

We can observe from the relevant studies on classifying blood microscopic smear images that prior to localizing the white blood cell based on the position of the segmented nucleus, in [22], the authors suggested color space conversion and k-means algorithm-based nucleus segmentation to separate them from the blood smear image. In order to classify the localized blood cell image, a modified convolutional neural network (CNN) architecture employing the concept of feature fusion of the first and final convolutional layers was proposed. In [5], the authors extracted the texture features using a two-dimensional Discrete Orthonormal S-Transform (DOST), and its dimensionality is reduced using both Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA). The reduced features are then fed to the proposed AdaBoost combined with Random Forest (ADBRF) classifier, that uses Random Forest as the base classifier. Prior to applying the Gray Level Run Length Matrix (GLRLM) for feature extraction, in [23], the authors employed the Otsu Thresholding approach to identify leukocytes, followed by Marker-based Watershed to refine the line of irregularly shaped leukocytes. The extracted features are then fed to the Support Vector Machine (SVM) classifier. In [9], the authors applied Contrast Limited Adaptive Histogram Equalization (CLAHE) to enhance the image quality prior to lymphocyte segmentation using a color-based K-means clustering technique, followed by feature extraction using a Gray Level Co-Occurrence Matrix (GLCM) and Gray Level Run Length Matrix (GLRLM) methods. Then, dimensionality is reduced using Principal Component Analysis (PCA), and the reduced features are fed to the Support Vector Machine (SVM) classifier. In [24], the authors proposed a hybrid transfer learning-based method that employs MobilenetV2 and ResNet18, preserving the advantages of both architectures and also introduced a novel weight factor. In [25], the authors presented transfer learning-based feature extraction utilizing MobileNetV2 followed by classification employing Support Vector Machine (SVM) in the classification layer of MobileNetV2. Prior to classification utilizing the VGG16 architecture, in [26], the authors suggested an adaptive unsharpening method for image enhancement, which is made up of normalizing the radius of the cells followed by estimating the focus quality and enhancing the sharpness of the images adaptively using VAR-PCANet. In [7], the authors utilized UNet to segment the nucleus of cells, followed by deep feature extraction and fusion using SqueezeNet, AlexNet, and GoogleNet. Then, by utilizing mutual information (MI), recursive feature elimination (RFE), and minimum recursive maximal relevance (mRmR), features were selected. Furthermore, the intersected features were fused before performing statistical analysis using ANOVA, followed by classification utilizing Support Vector Machine (SVM). In [27], the authors transformed the images from RGB to CMYK color space and applied histogram equalization, followed by segmentation utilizing Zack's algorithm. After that, by using roundness ratio, the grouped and ungrouped lymphocytes were recognized and separated. Later, color and shape features were extracted from these lymphocytes, which were then fed to Support Vector

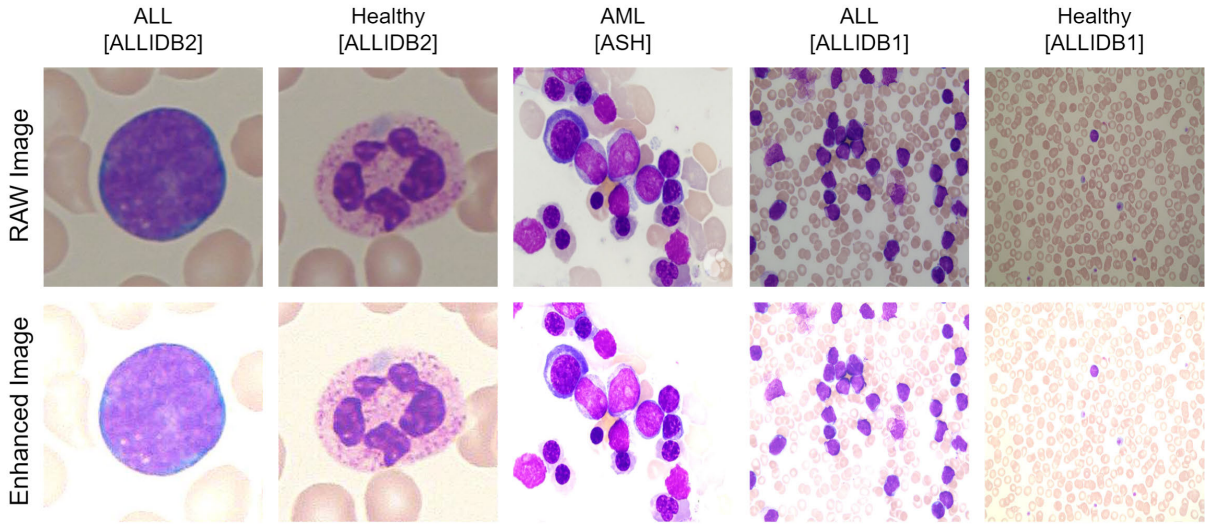



**FIGURE 1.** Raw microscopic smear images and their corresponding enhanced images after applying image enhancement techniques.

Machine (SVM) classifier. In [28], the authors extracted the green channel of the RGB image and enhanced the image using median filtering and histogram equalization. After that, the white blood cells (WBC) were segmented using the thresholding-based segmentation technique. Later, from the segmented cells, significant features like geometrical, statistical, shape, and discrete cosine transform (DCT)-based features were extracted and fed to Support Vector Machine (SVM) classifier. In [29], the authors converted the color space from RGB to CMYK, followed by image enhancement using histogram equalization. Segmentation was performed using thresholding estimation by Zack's algorithm, and features such as shape, color, texture along with hybrid features were extracted from the segmented cells. Social Spider Optimization Algorithm (SSOA) was employed to select the appropriate features which were fed to K-Nearest Neighbors (KNN) algorithm. Because of providing outstanding performance on small datasets, transfer learning has become a prominent medical image analysis technique. In [10], the authors proposed a computationally efficient and lightweight transfer learning-based approach employing ShuffleNet, which achieved good results by preserving the advantages of depthwise separable convolution, group convolution, and channel shuffling.

Although it is debatable whether standard machine learning methods might be more efficient in cases of data scarcity, data augmentation approaches can solve this problem. It inspired us to design a new automated acute leukemia diagnosis system employing ensemble deep learning, in which the segmentation process is not required, unlike other state-of-the-art approaches. In this research work, we present a CAD architecture regarding microscopic blood smear images based on ensemble deep learning for the purpose of diagnosing acute leukemia. The proposed architecture was evaluated using the ALLIDB1 [30], ALLIDB2 [30], and American Society of Hematology (ASH) [31] datasets. Due to the limited availability of microscopic blood smear images to use in training our proposed architecture, we adopted two strategies:

- To increase the number of samples, traditional data augmentation techniques were adopted to construct transformed versions of microscopic smear images (such as flipping and rotation).
- Instead of training the employed deep learning models from scratch, we finetuned the pretrained version of these models trained on the ImageNet dataset.

These two strategies assisted in training the proposed architecture with the available microscopic smear images and helped it to perform significantly on the test dataset of 1248 images. We have included the Receiver Operating Characteristic (ROC) curve and the area under the curve (AUC) score to present a summary of the performance of the proposed architecture and other pretrained, fine-tuned base models.

This paper's main contributions are as follows:

- An acute leukemia diagnosis system has been proposed and evaluated on three individual publicly available acute leukemia microscopic image datasets. The architecture was tested against other state-of-the-art methods and other adopted base models.
- We created an augmented dataset of microscopic smear images with multiclass labels (''ALL'', ''AML'', ''Healthy'') using traditional augmentation techniques for detecting acute leukemia, which can serve as a standard for the research community.
- We adjusted the pixel values of the images and applied an edge-detecting Laplacian filter over the images for enhancing the quality of the microscopic smear images.
- We proposed a two-staged feature fusion-based stacked ensembling integrated CAD architecture employing pretrained, fine-tuned EfficientNetB7 and

MobileNetV3Large architectures for the detection of acute leukemia.

The structure of the rest of this paper is as follows:
Sections II and III present a detailed description of the employed materials and the proposed architecture, respectively. The experimental analysis and discussion are presented in Section IV. Finally, Section V brings the paper to a conclusion.

## II. MATERIALS

### A. DATASET DESCRIPTION

In this research, the publicly accessible standard ALLIDB1 [30], ALLIDB2 [30], and American Society of Hematology (ASH) [31] datasets were employed to evaluate the efficacy of our suggested approach. All images in the ALLIDB1 and ALLIDB2 datasets were captured in JPG format with a 24-bit RGB color space using an optical microscope (magnification ranging from 300 to 500) with a Canon PowerShot G5 camera. ALLIDB1 is made up of numerous leukocytes per image, while ALLIDB2 is made up of blood smears with one leukocyte per image. ALLIDB1 consists of 108 microscopic images (59 healthy and 49 ALL) with a resolution of 2592×1944. ALLIDB2 consists of 260 microscopic images (130 healthy and 130 ALL) with a resolution of 257×257. In order to identify AML, 130 microscopic images of AML were collected from the American Society of Hematology (ASH) database.

## III. PROPOSED ARCHITECTURE

This section introduces the architecture of the ensemble deep learning-based CAD system that we have proposed for the diagnosis of acute leukemia using blood microscopic smear images. The whole approach is divided into two stages: (i) Data Preprocessing and (ii) Detection. Each stage is described in detail below:

### A. DATA PREPROCESSING

A substantial quantity of labeled data is necessary for training deep learning models. In terms of medical applications, most acquired datasets show a scarcity of instances that are often distributed unevenly, acting as a significant challenge when attempting to train deep learning architectures [32]. Due to the nature of our dataset, we only used rotation (0°, 90°, 180°, 270°) and flip as the traditional augmentation techniques. The images were subsequently improved using image enhancement methods. The microscope light is adjusted to produce images of the blood samples when they are studied under a microscope. As a result, the brightness of the microscope fluctuates with time, as do the reflections, all of which contribute to the worsening of deep learning model performance. As a result, image enhancement methods are beneficial. So, the brightness of the images was adjusted by adding 30 to each pixel value and the contrast was adjusted by scaling the pixel values by 1.4. Later, an edge-detecting Laplacian filter [[0, −1, 0], [−1, 5, −1], [0, −1, 0]] was applied. This filter recognizes areas that are changing, such as blast cells. Equations (1), (2), and (3) explain the operation of the Laplacian filter. $\nabla^2 f$ represents a second order derivative that ends up with a 1-pixel shift.

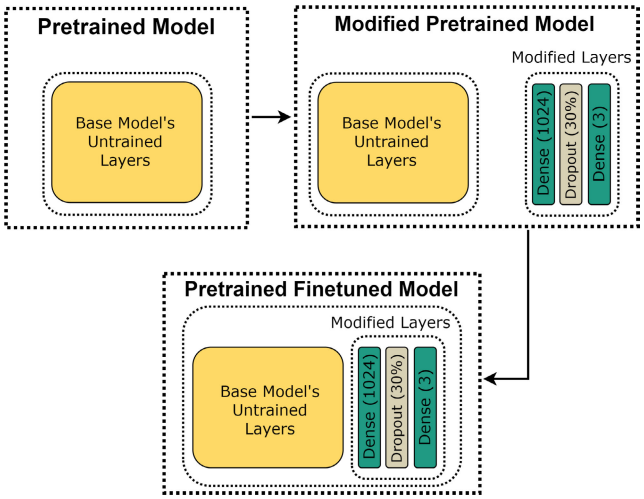



**FIGURE 2.** Architecture of the pretrained, fine-tuned base model.

$$\nabla^2 f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} \tag{1}$$

$$\frac{\partial^2 f}{\partial x^2} = f(x+1, y) + f(x-1, y) - 2f(x, y) \tag{2}$$

$$\frac{\partial^2 f}{\partial y^2} = f(x, y+1) + f(x, y-1) - 2f(x, y) \tag{3}$$

The enhanced images were then acquired. Figure 1 displays both the original images and the corresponding enhanced versions. Following that, we created a three-class dataset by adding images from the ALLIDB2 dataset's Healthy and ALL classes along with images from the ASH dataset's AML class.

### B. DETECTION

The detection stage is divided into two stages. In the first stage, pretrained, fine-tuned base models are employed and in the second stage, the generated feature maps are fused and fed to the fully connected meta-classifier to get the final detection result.

#### 1) MODIFIED PRETRAINED FINE-TUNED EFFICIENTNETB7

EfficientNetB7 [33] architecture has introduced a compound scaling method that uniformly scales the dimensions of depth ($d$), width ($w$), and resolution ($r$) of the base model architecture, EfficientNetB0. Multi-objective Neural Architecture Search (NAS) is introduced to develop the base model architecture that incorporates mobile inverted bottleneck convolution (MBConv). Then, the compound scaling method is utilized using Equation (4) to scale up the base model architecture for developing EfficientNetB7.

$$d = \alpha^{\phi}; \quad where \quad \alpha \geq 1$$
$$w = \beta^{\phi}; \quad where \quad \beta \geq 1$$

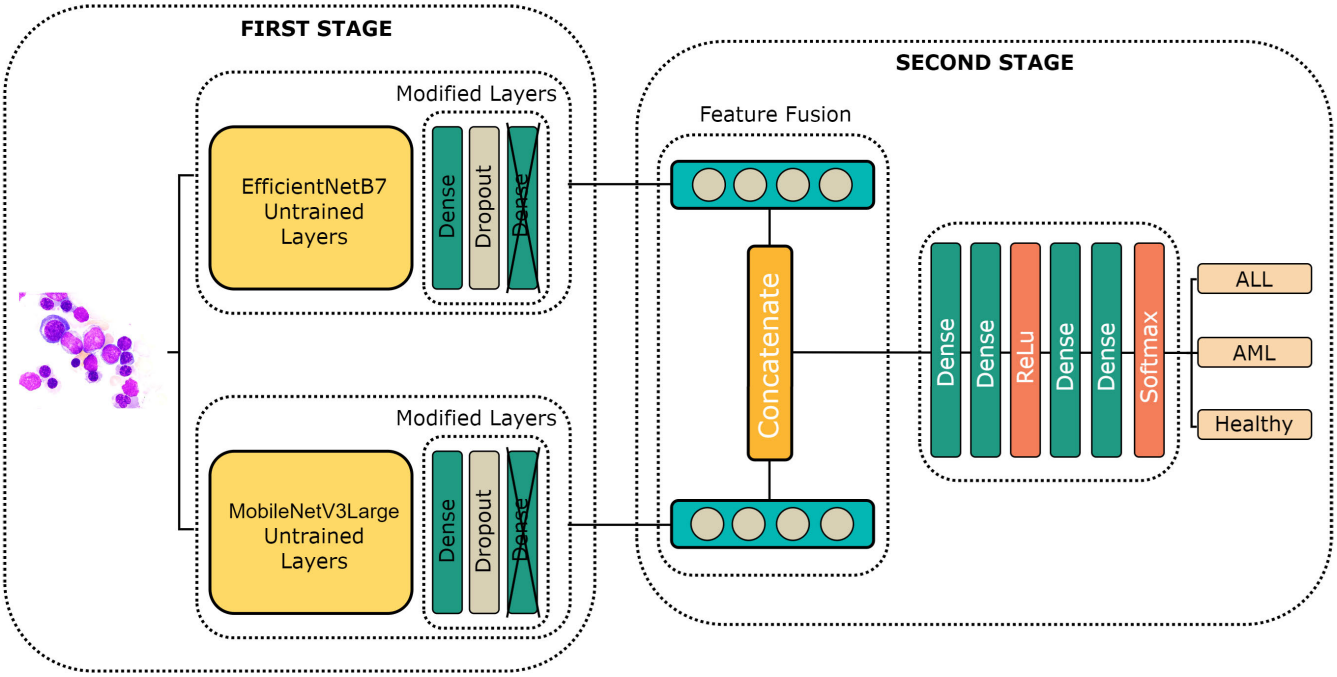



**FIGURE 3.** The proposed two-staged feature fusion-based stacked ensemble architecture incorporated with pretrained, fine-tuned base models and ensemble of feature maps in order to classify acute leukemia.

$$r = \gamma^{\phi}; \qquad where \quad \gamma \geq 1 \tag{4}$$

The values of $\alpha$, $\beta$, and $\gamma$ are determined by a small grid search, and the compound coefficient $\phi$ denotes the available computational resources. The ImageNet dataset [34] was used primarily to train the architecture to classify 1000 different classes. A modified pretrained, fine-tuned EfficientNetB7 architecture is then created by employing a dense layer with 1024 nodes and a "ReLu" activation function, followed by a dropout layer with a drop probability of 0.3 and another dense layer with 3 nodes and a "Softmax" activation function. Later, the prepared multiclass leukemia dataset is used to train the whole architecture to classify microscopic smear images.

### 2) MODIFIED PRETRAINED FINE-TUNED MOBILENETV3LARGE

MobileNetV3Large [35] architecture has improvised the linear bottleneck and inverted residual structure of the previous version by introducing squeeze and excitation in the bottleneck structure. Also, hard-swish nonlinearity is introduced to the layers. The inclusion of squeeze and excitation aids in creating the output feature map by assigning unequal weights to distinct channels from the input. A platform-aware Neural Architecture Search (NAS) is introduced to find the overall network architecture by optimizing each network block alongside the NetAdapt algorithm that is used to determine the optimal amount of filters for each layer. Similar to the Modified Pretrained Fine-tuned EfficientNetB7 architecture, intially, the ImageNet dataset [34] was used to train the initial architecture to classify 1000 different classes. We then make a modified fine-tuned MobileNetV3Large architecture by employing a Dense layer with 1024 nodes and a "ReLu" activation function, followed by a Dropout layer with a drop probability of 0.3, and another Dense layer with 3 nodes and a "Softmax" activation function. Finally the whole architecture is trained on the prepared multiclass leukemia dataset to classify microscopic smear images.

### 3) FEATURE FUSION BASED ENSEMBLE

In this stage, a feature fusion-based ensembling has been proposed to recognize acute leukemia.

In the first stage, two modified pretrained deep neural networks were introduced which belong to EfficientNet and MobileNet architectures. Pretrained models are used to retrain the models on a different dataset using previously learned weights. It is possible to reduce the time required to train by utilizing this method without requiring a large amount of data [36], [37], [38], [39], [40], [41]. As a consequence, it is ideal for image-based medical applications that deal with small training datasets having just a few images per class. Our proposed methodology employs two pretrained deep CNN architectures. Figure 2 depicts how the modified pretrained, fine-tuned base models were adapted in the first stage of the proposed architecture. To create pretrained, fine-tuned base models, the layers of the base models were made frozen. Here, the number of features produced by the probabilistic confidence map created by Softmax activation is equivalent to the number of classes in the dataset. The entire set of pretrained, fine-tuned base models was retrained after that. The weights of the pretrained, fine-tuned base models were preserved. Later, to train the whole two-staged feature fusion-based ensembling architecture, these saved weights were loaded and kept frozen. Due to the high

**TABLE 1. Hyperparameters of the proposed architecture.**

| Hyperparameter | Value |
|---|---|
| Batch Size | 32 |
| Number Of Epochs | 100 |
| Optimizer | Adam |
| Hidden Layer Activation Function | ReLu, Sigmoid |
| Output Layer Activation Function | Softmax |
| Early Stopping Patience | 50 |
| Train - Valid - Test Split | 70% - 10% - 20% |
| Loss Function | Categorical Cross Entropy |

memory requirements involved in parallelly training two pretrained, fine-tuned base models, this two-staged approach was adapted. Additionally, our proposed architecture allowed the pretrained, fine-tuned base models to preserve their underlying behavioral traits of extracting impactful features by keeping the weights intact. As a result, during the feature fusion-based ensemble, they were not affected by each other.

As feature maps are ensembled in the feature fusion-based ensembling method, the two feature maps generated by the pretrained, fine-tuned base models belong to two different deep neural network architecture families (EfficientNet and MobileNet) were concatenated in the second stage. Equation (5) explains how the feature maps were concatenated. As we are concatenating the feature maps generated from the Dense layer of the pretrained, fine-tuned base models, $N_k$ denotes the number of nodes, and the value is the same for each model. $k$ denotes the number of feature maps, and $F_{MAP}$ denotes the final concatenated feature map.

$$F_{MAP} = \sum_{k=1}^{n} N_k \tag{5}$$

Prior to this process, the last Dense layer of each pretrained, fine-tuned base model was discarded, and after the concatenation of two feature maps, a Dense layer with 1000 units and sigmoid activation, followed by a Dense layer with 100 units and ReLu activation, and a dense layer with 10 units and sigmoid activation, were added to create a two-layer metaclassifier model. Finally, the output Dense layer with Softmax activation was added. Following the individual training (modified pretrained, fine-tuned EfficientNetB7 and modified pretrained, fine-tuned MobileNetV3Large) using the corresponding dataset in the first stage, Figure 3 depicts how the feature maps of each pretrained, fine-tuned base model were generated and fed into the proposed feature fusion-based stacked ensemble architecture as inputs for the final prediction in the second stage. For training the model, we employed 70% of the augmented dataset along with ImageDataGenerator, 10% for validation, and the remaining 20% for testing. As mentioned earlier, the training process of the proposed architecture had two stages since the pretrained, fine-tuned base models were retrained separately while training the whole proposed architecture. This approach not only decreases training time and memory consumption but also retains the unique architectural traits of each pretrained, fine-tuned base model to choose the optimal collection of features among all the extracted features of the pretrained, fine-tuned base models.

### C. HYPERPARAMETERS FOR PROPOSED ARCHITECTURE

The loss function, activation function, batch size, optimizer, number of epochs, etc. are all important factors to consider. All of these factors have been considered as hyperparameters for training the proposed architecture. In the hidden layers of the modified layers of the first stage and in the meta-classifier of the second stage, the ''sigmoid'' and the ''ReLu'' activation functions are employed. Additionally, ''Softmax'' is used for multiclass classification in the output layer. Table 1 displays the hyperparameter values that were employed.

### D. PERFORMANCE METRICS

Performance metrics are used to evaluate each machine learning architecture's performance. The effectiveness of classification may be evaluated in many different ways. In this study, we employed the accuracy, precision, recall, F1-score, area under the curve (AUC) score, and confusion matrix for the performance evaluation of acute leukemia detection. As indicated in Equation (6), *Accuracy* is the proportion of correctly predicted data points among all the data points, where $T_P$ stands for the number of correct positive predictions, $T_N$ for the number of correct negative predictions, $F_P$ for the number of incorrect positive predictions, and $F_N$ for the number of incorrect negative predictions. The accuracy of the positive predictions is measured by *Precision* whereas the completeness of the positive predictions is measured by *Recall*. The harmonic mean of the recall values and precision values is represented by the $F1 - score$. The ''macro'' average is utilized to calculate the AUC score. Furthermore, a confusion matrix is created from these measures to demonstrate the trade-off between the predicted acute leukemia class labels and the actual acute leukemia class labels.

$$Accuracy = \frac{T_P + T_N}{T_P + F_P + T_N + F_N} \tag{6}$$

$$Precision = \frac{T_P}{T_P + F_P} \tag{7}$$

$$Recall = \frac{T_P}{T_P + F_N} \tag{8}$$

$$F1 - score = \frac{2 \times T_P}{2 \times T_P + F_P + F_N} \tag{9}$$

## IV. EXPERIMENTATION AND PERFORMANCE ANALYSIS

### A. EXPERIMENTAL ENVIRONMENT

The hardware utilized for conducting the experiments regarding this study was a Windows 11 PC equipped with a 2.20 GHz Intel(R) Xeon(R) processor, 16 GB of RAM, and an NVIDIA Tesla P100 GPU. All experiments using our proposed two-staged feature fusion-based stacked ensemble architecture to detect acute leukemia were carried out using Python 3.10.10 and TensorFlow 2.9.0.

**TABLE 2.** Acute leukemia detection performance on the test data.

| Method | Accuracy | Precision | Recall | F1-score | AUC score |
|---|---|---|---|---|---|
| Xception | 90.5% | 90.5% | 90.5% | 90.5% | 0.98 |
| NASNetLarge | 91.6% | 91.7% | 91.6% | 91.6% | 0.98 |
| InceptionV3 | 92.1% | 92.1% | 92.1% | 92.1% | 0.98 |
| ResNet50V2 | 92.1% | 92.1% | 92.1% | 92.1% | 0.98 |
| ResNet101V2 | 93.0% | 93.1% | 93.0% | 93.0% | 0.99 |
| EfficientNetB6 | 94.0% | 94.0% | 94.0% | 94.0% | 0.99 |
| EfficientNetB7 | 96.4% | 96.4% | 96.4% | 96.4% | 0.99 |
| MobileNetV3Large | 98.4% | 98.4% | 98.4% | 98.4% | 0.99 |
| **Proposed Method** | **99.3%** | **99.3%** | **99.3%** | **99.3%** | **0.99** |

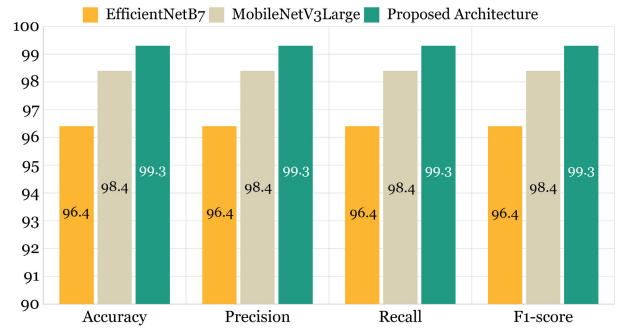



**FIGURE 4.** Performance comparison of the proposed method and the modified pretrained, fine-tuned base models which are employed in the proposed method regarding acute leukemia detection.

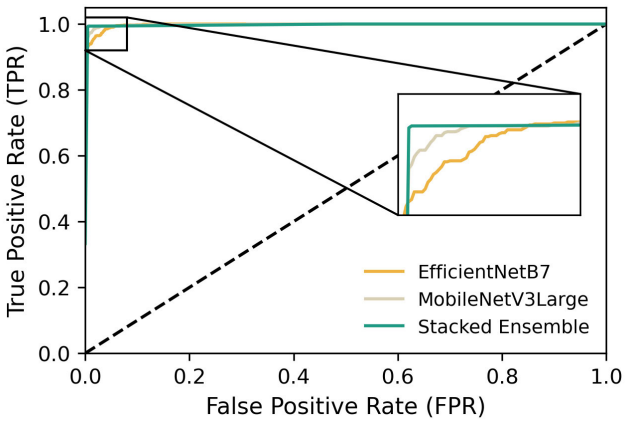



**FIGURE 5.** ROC curves of the proposed method and the employed modified pretrained, fine-tuned base models demonstrating the detection performance.

### B. PERFORMANCE EVALUATION OF DETECTION

The results and performance comparison of the proposed two-staged feature fusion-based acute leukemia detection architecture with other pretrained, fine-tuned base models to identify acute leukemia are shown in Table 2. The modified pretrained, fine-tuned EfficientNetB7 model and the modified pretrained, fine-tuned MobileNetV3Large model outperformed the other base models and were employed in the two branches of the proposed two-staged feature fusion-based stacked ensemble architecture. The parameter description regarding the proposed architecture, the employed modified pretrained, fine-tuned base models, and the other modified pretrained, fine-tuned testing models are shown in Table 3. The required time for testing is also included, where the number of steps for testing was 39. As demonstrated in

**TABLE 3.** Parameter description for acute leukemia detection.

| Method | Parameter | | Test Time (ms/step) |
|---|---|---|---|
| | Total | Trainable | |
| Xception | 23,938,579 | 1,028,099 | 112 |
| NASNetLarge | 89,977,917 | 1,028,099 | 262 |
| InceptionV3 | 24,879,883 | 1,028,099 | 106 |
| ResNet50V2 | 26,641,899 | 1,028,099 | 115 |
| ResNet101V2 | 45,703,659 | 1,028,099 | 121 |
| EfficientNetB6 | 44,293,242 | 1,028,099 | 167 |
| EfficientNetB7 | 67,686,786 | 1,028,099 | 209 |
| MobileNetV3Large | 6,535,531 | 1,028,099 | 94 |
| **Proposed [Branch 1]** | **67,683,711** | **1,025,024** | |
| **Proposed [Branch 2]** | **6,532,456** | **1,025,024** | **228** |
| **Proposed [Meta-classifier]** | **2,150,143** | **2,150,143** | |

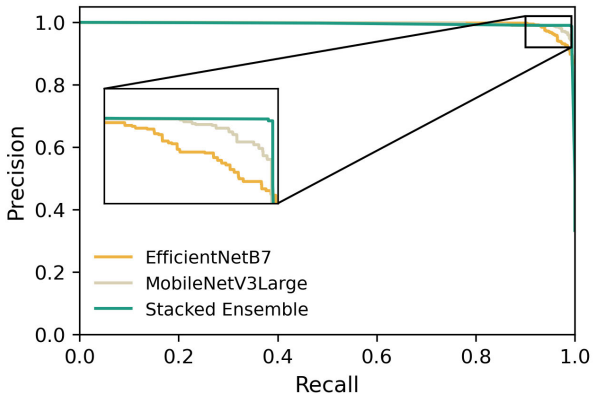



**FIGURE 6.** Precision/recall curves of the proposed method and the employed modified pretrained, fine-tuned base models demonstrating the detection performance.

Figure 4, our proposed two-staged feature fusion-based stacked ensemble technique outperformed the other pretrained, fine-tuned base models along with the employed modified pretrained, fine-tuned base models with varied numbers of deep layers in terms of Accuracy, Precision, Recall, and F1-score. The feature fusion-based ensemble learning strategy was emphasized in the findings to improve performance, and we can observe improvements in the Accuracy, Precision, Recall, F1-score, and AUC score. The Accuracy, precision, Recall, and F1-score of our proposed two-staged feature fusion-based stacked ensemble approach were all 99.3%. Additionally, the AUC score was used to compare the acute leukemia detection results.

#### 1) CONFUSION MATRIX

As confusion matrix is taken into consideration for demonstrating the performance, the confusion matrix in Figure 7

**TABLE 4.** Overall performance of the proposed two-staged feature fusion-based stacked ensemble approach in detecting acute leukemia, acute lymphocytic leukemia and acute myeloid leukemia.

| | Dataset (Class) | Accuracy | Precision | Recall | F1-score | AUC score |
|---|---|---|---|---|---|---|
| **Acute Leukemia** | Combined Dataset (ALL, AML, Healthy) | 99.3% | 99.3% | 99.3% | 99.3% | 0.99 |
| **Acute Lymphocytic Leukemia** | ALLIDB1 Dataset (ALL, Healthy) | 100.0% | 100.0% | 100.0% | 100.0% | 1.00 |
| **Acute Lymphocytic Leukemia** | ALLIDB2 Dataset (ALL, Healthy) | 99.4% | 99.4% | 99.4% | 99.4% | 0.99 |
| **Acute Myeloid Leukemia** | ASH Dataset (AML, Healthy) | 99.8% | 99.8% | 99.8% | 99.8% | 0.99 |

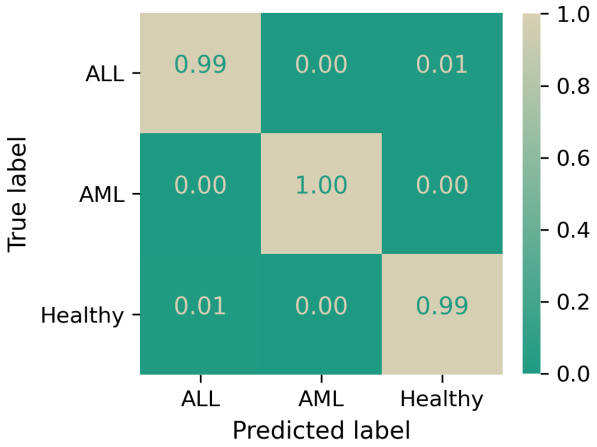



**FIGURE 7.** Confusion matrix of the proposed method highlighting significant classification trade-offs among classes regarding acute leukemia detection.

reflects the performance of the proposed architecture in determining acute leukemia. The normalized confusion matrix highlights significant classification trade-offs across classes (ALL, AML, Healthy). It is noticed that among the images of the test set, all the images of the AML class were classified perfectly and only 1% test images of both the ALL class and the Healthy class were misclassified by the proposed architecture, whereas 1% test images of the AML class and 5% test images of both the ALL class and the Healthy class were misclassified by the modified pretrained, fine-tuned EfficientNetB7 base model and the modified pretrained, fine-tuned MobileNetV3Large base model misclassified 2% and 3% test images of the ALL class and the Healthy class respectively and predicted all the test images of the AML class perfectly, clearly indicating that the proposed architecture can detect acute leukemia more accurately than the employed base models.

### 2) ROC AND PRECISION/RECALL CURVES

It is noticeable in Table 2 that the AUC score regarding the proposed architecture and the employed two modified pretrained, fine-tuned base models are the same which is 0.99 and to interpret it graphically, ROC curves were taken into consideration. Figure 5 shows the plots of the ROC curves of the True Positive Rate against the False Positive Rate regarding the proposed architecture and the employed base models. The proposed two-staged feature fusion-based stacked ensemble architecture has a larger area under the ROC curve than the employed two base models.

**TABLE 5.** Class-wise performance of acute leukemia detection on the test data.

| Class | Method | Precision | Recall | F1-score |
|---|---|---|---|---|
| ALL | EfficientNetB7 | 95.4% | 95.0% | 95.2% |
| | MobileNetV3Large | 96.9% | 98.3% | 97.6% |
| | **Proposed Method** | **98.8%** | **99.0%** | **98.9%** |
| AML | EfficientNetB7 | 99.5% | 99.3% | 99.4% |
| | MobileNetV3Large | 100.0% | 99.8% | 99.9% |
| | **Proposed Method** | **100.0%** | **100.0%** | **100.0%** |
| Healthy | EfficientNetB7 | 94.3% | 95.0% | 94.6% |
| | MobileNetV3Large | 98.3% | 97.1% | 97.7% |
| | **Proposed Method** | **99.0%** | **98.8%** | **98.9%** |

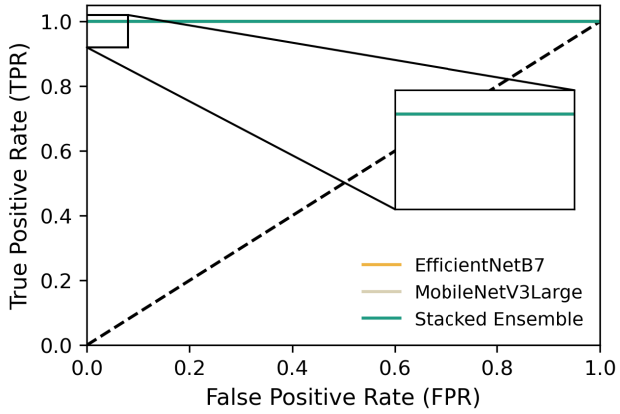
To prove the robustness of the proposed architecture, Precision-vs-Recall curves were also taken into consideration. For various thresholds, the precision-vs-recall curve illustrates the trade-off between precision and recall. A strong recall is correlated with a low false negative rate, whereas a low false positive rate is indicative of excellent precision. Both high recall and precision are denoted by a substantial area under the curve. Figure 6 shows the plots of the precision-vs-recall curves of the Precision values against the Recall values regarding the proposed architecture and the employed base models. The proposed two-staged feature fusion-based stacked ensemble architecture has a larger area under the curve than the employed two base models. So the evidence of the enhanced performance of the proposed two-staged feature fusion-based stacked ensemble architecture is clearly visible from these comparisons.

### 3) CLASSWISE PERFORMANCE

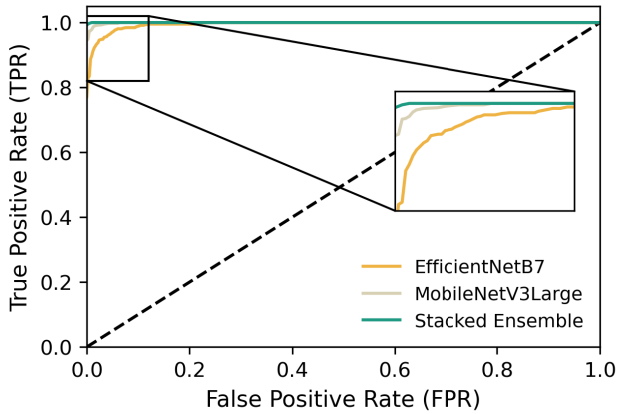
Apart from the confusion matrix, Precision, Recall, and F1-score were also taken into consideration to show more detailed class-wise performance in detecting acute leukemia. Table 5 shows the class-wise performance in detecting acute leukemia. In detecting the test images of the ALL class, the proposed two-staged feature fusion-based stacked ensemble architecture achieved highest precision, recall, and F1-score of 98.8%, 99.0%, and 98.9%, respectively whereas in detecting the test images of the AML class the proposed architecture achieved the perfect score of 100.0% and in detecting the test images of the Healthy class the proposed architecture achieved a precision of 99.0%, recall of 98.8%, and F1-score of 98.9% which are significant improvement from the scores of the employed two modified pretrained, fine-tuned base models' scores.

### 4) COMPARATIVE ANALYSIS

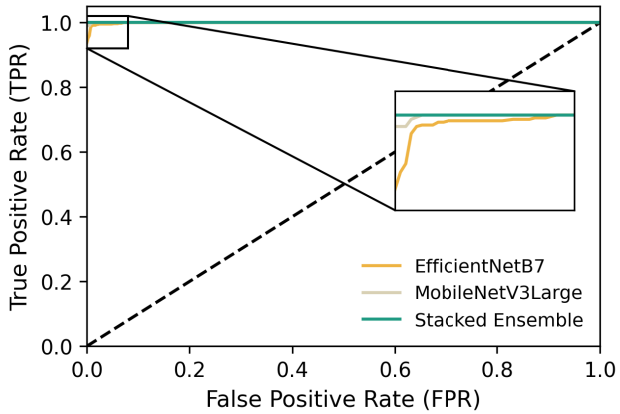
The preceding comparison shows that the proposed strategy yields the best results in detecting acute leukemia. Therefore,




(a) ROC curves regarding ALLIDB1 dataset.

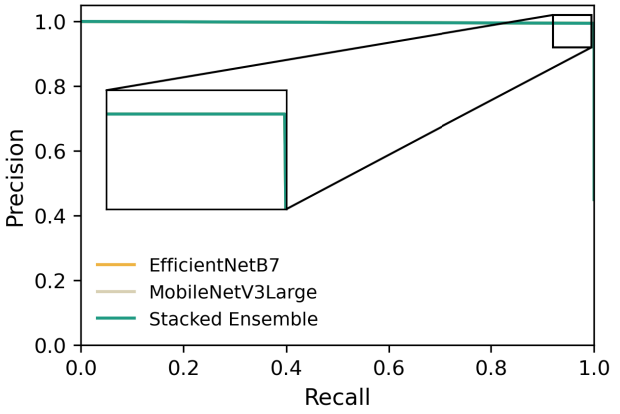



(b) ROC curves regarding ALLIDB2 dataset.

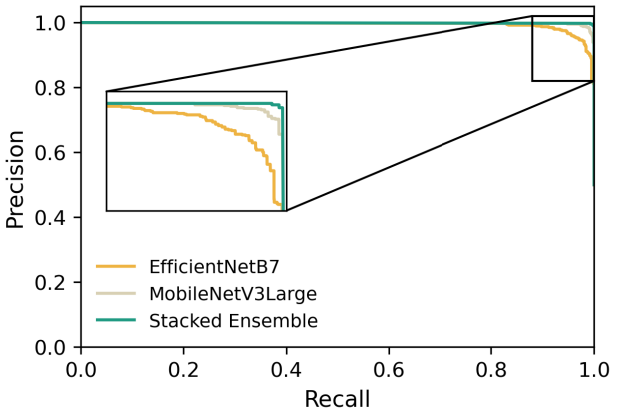



(c) ROC curves regarding ASH dataset.

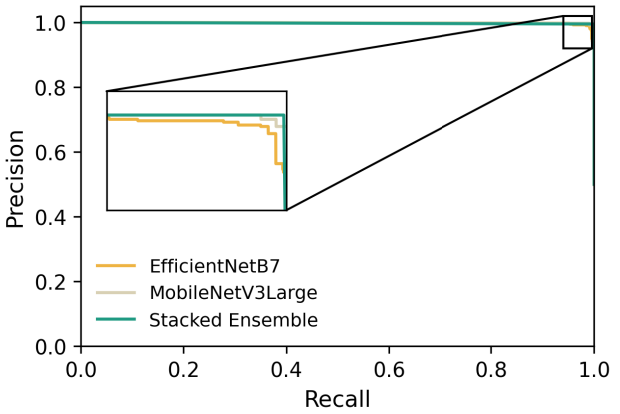
**FIGURE 8.** ROC curves of the proposed two-staged feature fusion-based stacked ensembling integrated architecture and the employed pretrained, fine-tuned base models demonstrating acute lymphocytic leukemia detection performances on both the ALLIDB1 and ALLIDB2 datasets, and acute myeloid leukemia detection performance on the ASH dataset.




(a) Precision/recall curves regarding ALLIDB1 dataset.




(b) Precision/recall curves regarding ALLIDB2 dataset.




(c) Precision/recall curves regarding ASH dataset.

**FIGURE 9.** Precision/recall curves of the proposed two-staged feature fusion-based stacked ensembling integrated architecture and the employed pretrained, fine-tuned base models demonstrating acute lymphocytic leukemia detection performances on both the ALLIDB1 and ALLIDB2 datasets, and acute myeloid leukemia detection performance on the ASH dataset.

we applied the proposed two-staged feature fusion-based stacked ensemble approach on both the ALLIDB1 and the ALLIDB2 datasets for detecting acute lymphocytic leukemia and on the ASH dataset for detecting acute myeloid leukemia. Table 4 describes the overall performance of our proposed two-staged feature fusion-based stacked ensemble approach in detecting acute leukemia, acute lymphocytic leukemia, and acute myeloid leukemia, which denotes that the proposed approach achieved an accuracy of 99.3% in detecting acute leukemia, 100.0% and 99.4% in detecting acute lymphocytic leukemia on the ALLIDB1 and the ALLIDB2 datasets,

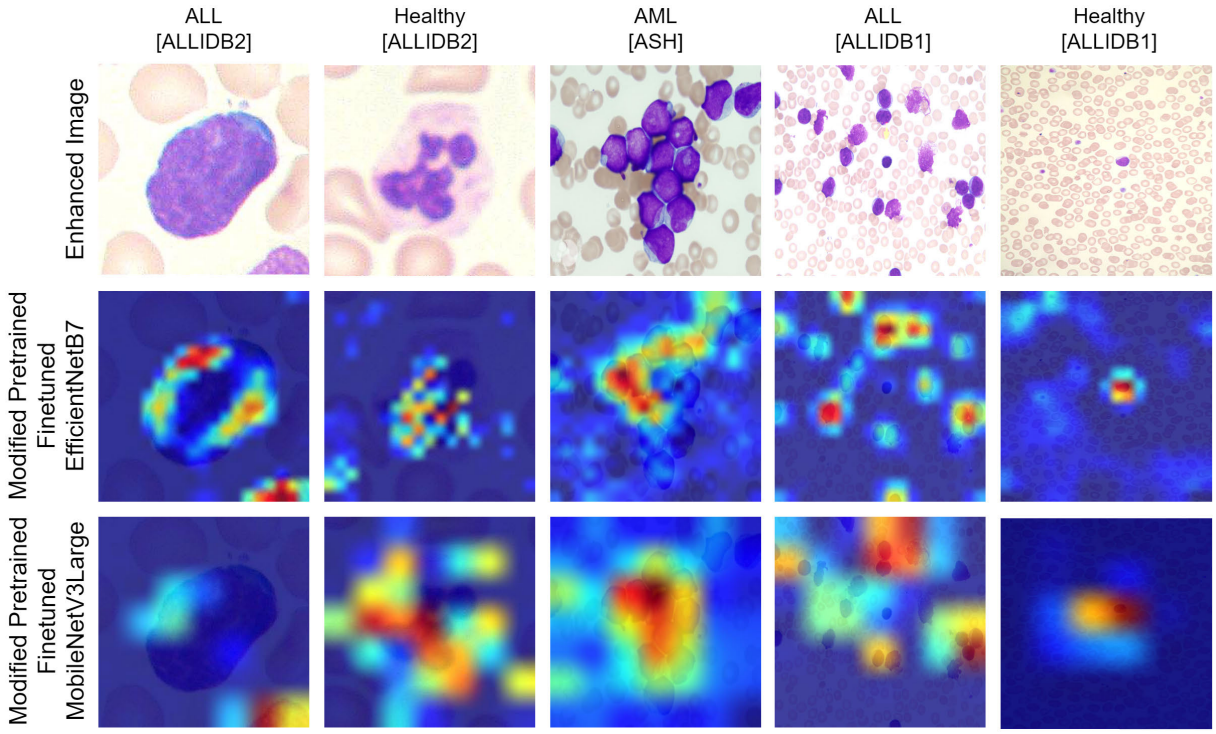



**FIGURE 10.** Input images and the corresponding superimposed images generated by the Grad-CAM regarding the base models of the two branches of the first stage of the proposed two-staged feature fusion-based stacked ensemble architecture.

respectively, and 99.8% in detecting acute myeloid leukemia on the ASH dataset.

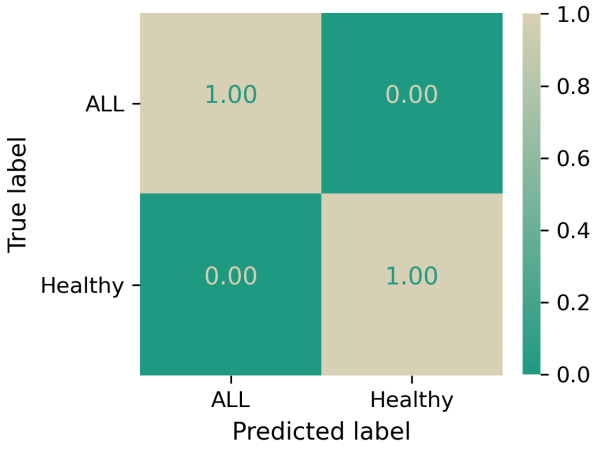
By employing the test data from the ALLIDB1 and the ALLIDB2 datasets, our proposed approach was compared with the state-of-the-art methods for the detection of acute lymphocytic leukemia. As shown in Table 6, we compared the results of our proposed two-staged feature fusion-based stacked ensemble approach to the results of existing state-of-the-art methods for the detection of acute lymphocytic leukemia and found that it performed significantly on the ALLIDB1 dataset by achieving 100% Accuracy, Precision, Recall, and F1-score. On top of that, the proposed approach had the highest AUC score (1.0).

Similarly, as shown in Table 7, we also compared the results of our proposed two-staged feature fusion-based stacked ensemble approach to the results of existing state-of-the-art methods for the detection of acute lymphocytic leukemia and found that it performed better on the ALLIDB2 dataset as well. In terms of Accuracy, our proposed approach outperformed the methods used in [7], [10], [11], [24], [25], [26], [29], and [47] by margins of 1.2%, 2.7%, 1.2%, 2.2%, 1.2%, 2.6%, 3.7%, and 0.7%, respectively. When compared to the Precision, the methods used in [7], [10], [11], [24], [25], and [26] were outperformed by 0.2%, 2.4%, 0.9%, 0.9%, 1.8%, and 3.2%, respectively. When compared to the Recall, the methods used in [7], [10], [11], [24], [25], [26], and [47] were outperformed by 1.3%, 2.9%, 1.4%, 3.5%, 0.4%, 1.9%, and 1.4%, respectively.

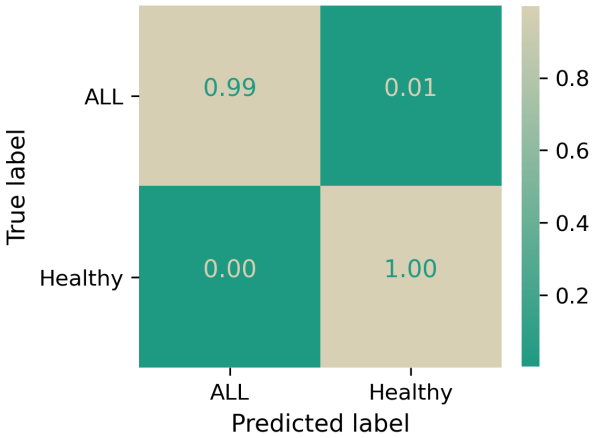
When compared to the F1-score, the methods used in [7], [10], [11], [24], [25], and [26] were outperformed by 3.1%, 2.7%, 1.2%, 2.2%, 1.1%, and 2.5%, respectively. Furthermore, the proposed approach achieved an outstanding AUC score of 0.99. Additionally, since confusion matrix, ROC curve, and Precision-vs-Recall curve were also employed to examine the predictions, Figure 8 represents the ROC curves, and Figure 9 represents the Precision-vs-Recall curves demonstrating the capability of distinguishing between classes regarding the ALLIDB1, ALLIDB2, and ASH datasets, whereas Figure 11 represents normalized confusion matrices that reveal substantial classification trade-offs across classes for all three datasets.

### C. GRAD-CAM EVALUATION

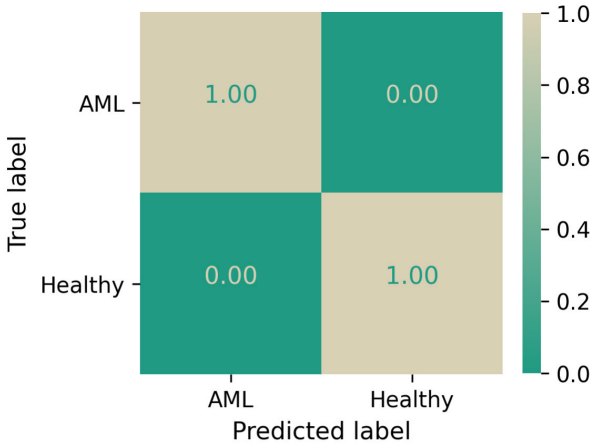
The Gradient Weighted Class Activation Mapping (Grad-CAM) offers a comprehensible representation of deep learning architectures. Grad-CAM provides a clear explanation for every interconnected neural network architecture, helping to get insight into the architecture during the detection task. Grad-CAM was taken into consideration to evaluate whether the normal and the blast cell sections present in the input images had significant impacts on detecting acute leukemia and to observe the evidence of detection visually. The Grad-CAM determines the gradient of the target class on each feature map and then averages them to figure out the relevance of each map. Grad-CAM is an efficient approach that maintains the execution speed without any




(a) Confusion matrix regarding ALLIDB1 dataset.




(b) Confusion matrix regarding ALLIDB2 dataset.




(c) Confusion matrix regarding ASH dataset.

**FIGURE 11.** Confusion matrices of the proposed feature fusion-based stacked ensembling integrated architecture highlighting significant classification trade-offs among classes regarding acute lymphocytic leukemia detection on both the ALLIDB1 and ALLIDB2 datasets, and acute myeloid leukemia detection on the ASH dataset.

**TABLE 6.** Performance comparison with the state-of-the-art methods on ALLIDB1 dataset utilizing our proposed method.

| | Year | Method | Accuracy | Precision | Recall | F1-score |
|---|---|---|---|---|---|---|
| Das and Meher [10] | 2021 | Transfer Learning | 97.0% | 95.6% | 98.0% | 96.8% |
| Das and Meher [24] | 2021 | Transfer Learning | 99.4% | 99.3% | 99.6% | 99.4% |
| Khandekar *et al.* [42] | 2021 | Deep Learning | - | 95.6% | 92.0% | 92.0% |
| Das *et al.* [25] | 2022 | Hybrid | 99.4% | 98.6% | **100.0%** | 99.3% |
| Maaliw *et al.* [43] | 2022 | Transfer Learning | 99.6% | 99.6% | 98.8% | 99.1% |
| Abed and Mohammed [44] | 2022 | Deep Learning | 99.3% | - | 97.9% | - |
| Kumar *et al.* [45] | 2022 | Deep Learning | 98.1% | 98.3% | 98.3% | 98.3% |
| Al-khuzaie *et al.* [46] | 2023 | Transfer Learning | 99.5% | - | **100.0%** | - |
| Das *et al.* [11] | 2023 | Transfer Learning | 99.4% | **100.0%** | 98.7% | 99.3% |
| **Proposed Architecture** | 2024 | **Feature Fusion - Deep Learning** | **100.0%** | **100.0%** | **100.0%** | **100.0%** |

requirement for additional specifically built components. A single test image regarding each class was taken as input by the Grad-CAM and applied the proposed architecture, where the last convolutional layer was utilized to obtain the visualizations. Grad-CAM leveraged the gradients of the last convolutional layer and developed a heatmap that emphasized the most significant sections of an input image. When the heatmap was placed on the original input image, the specific areas based on which classification was done were determined. Figure 10 shows the class activation maps of the two modified pretrained, fine-tuned base models employed in the two branches of the first stage of the proposed architecture. We can see how significantly these two base models considered the normal and the blast cells of the input images to detect acute leukemia.

### D. ABLATION STUDY

Apart from the preceding comparisons, we performed more experiments regarding the proposed architecture to obtain a more precise ablation analysis. We performed the detection

**TABLE 7.** Performance comparison with the state-of-the-art methods on ALLIDB2 dataset utilizing our proposed method.

| | Year | Method | Accuracy | Precision | Recall | F1-score |
|---|---|---|---|---|---|---|
| Sahlol *et al.* [29] | 2019 | Machine Learning | 95.7% | - | - | - |
| Jha and Dutta [47] | 2019 | Deep Learning | 98.7% | - | 98.0% | - |
| Das and Meher [10] | 2021 | Transfer Learning | 96.7% | 97.0% | 96.5% | 96.7% |
| Alagu *et al.* [7] | 2021 | Hybrid | 98.2% | 99.2% | 98.1% | 96.3% |
| Das and Meher [24] | 2021 | Transfer Learning | 97.2% | 98.5% | 95.9% | 97.2% |
| Genovese *et al.* [26] | 2021 | Deep Learning | 96.8% | 96.2% | 97.5% | 96.9% |
| Das *et al.* [25] | 2022 | Hybrid | 98.2% | 97.6% | 99.0% | 98.3% |
| Das *et al.* [11] | 2023 | Transfer Learning | 98.2% | 98.5% | 98.0% | 98.2% |
| **Proposed Architecture** | 2024 | **Feature Fusion - Deep Learning** | **99.4%** | **99.4%** | **99.4%** | **99.4%** |

process only employing a single branch in the first stage of the proposed architecture to observe the effect of feature fusion of two different feature maps belonging to two completely different base model architectures. From Table 8 we can see that when the modified pretrained, fine-tuned EfficientNetB7 was only employed in the single branch of the first stage, it achieved an Accuracy, Recall, and F1-score of 96.4% whereas 96.5% as Precision which is 0.1% greater when the modified pretrained, fine-tuned EfficientNetB7 was simply applied to detect acute leukemia, as shown in Table 2. When the modified pretrained, fine-tuned MobileNetV3Large was only employed in the single branch of the first stage, it achieved an Accuracy, Precision, Recall, and F1-score of 98.8% which is 0.4% greater when the modified pretrained, fine-tuned MobileNetV3Large was simply applied to detect acute leukemia, as shown in Table 2. When both the base models were employed in the two branches in the first stage of the proposed architecture, we can see the Accuracy, Precision, Recall, and F1-score reached to 99.3% which is clearly superior than the single branch operation. And it clearly signifies the effect of the feature fusion technique on two different feature maps belonging to two completely different base model architectures.

### E. DISCUSSION

To detect acute leukemia using microscopic smear images, our proposed two-staged feature fusion-based stacked ensemble integrated CAD system explores the difficulties and addresses the problems with significant performance. Furthermore, it has shown superior performance in detecting acute lymphocytic leukemia and acute myeloid leukemia as well. UNet, Zack's algorithm, threshold-based techniques, Random Forest, ADBRF, SVM, KNN, ResNet18, VGG16, MobileNetV2, ShuffleNet, and other state-of-the-art segmentation and classification algorithms, as well as hybrid algorithms combining both machine learning and deep learning algorithms, have been proposed to address the difficulties regarding this. When the performance of our proposed feature fusion-based stacked ensemble integrated CAD system is compared with the performance of the state-of-the-art approaches (as shown in Table 6 and Table 7), it is obvious that our proposed method can greatly exceed the prior efforts.

Three key elements were responsible for the improved performance. To begin, rather than immediately feeding the raw images into our proposed two-staged feature fusion-based stacked ensemble integrated CAD architecture, the raw images were converted to improved images by adjusting the intensity of brightness and contrast. The Laplacian filter also assisted in defining the poorly defined edges and overcoming the blurriness problem.

**TABLE 8.** Ablation study of acute leukemia detection on the test data.

| Approach | Accuracy | Precision | Recall | F1-score |
|---|---|---|---|---|
| Single Branch (EfficientNetB7) | 96.4% | 96.5% | 96.4% | 96.4% |
| Single Branch (MobileNetV3Large) | 98.8% | 98.8% | 98.8% | 98.8% |
| **Dual Branch (Proposed)** | **99.3%** | **99.3%** | **99.3%** | **99.3%** |

Secondly, the underlying pretrained structure of our proposed approach provided robust baseline feature map representations, enabling the method to leverage freshly acquired features as well as previously discovered universal features and gradients during training. This transfer learning strategy enhanced the convergence and overall performance. Also, it required very little time to retrain the pretrained, fine-tuned base models and instances to identify key patterns for improved performance.

Finally, the feature fusion-based stacked ensembling approach utilized the key features from the fusion of feature maps generated by two pretrained, fine-tuned base models belong to two different deep neural network architecture families (EfficientNet and MobileNet), preserving unique architectural traits, which boosted the overall performance significantly. In addition, the compound scaling architecture and the squeeze and excitation integrated bottleneck architecture of the pretrained, fine-tuned base models greatly influenced the overall performance, whereas the two-staged training approach greatly reduced the amount of time and memory needed for training the whole architecture.

## V. CONCLUSION

We evaluated the effectiveness of feature fusion-based stacked ensembling incorporating deep learning to diagnose acute leukemia using microscopic smear images in this study. Furthermore, an automated CAD System for diagnosing acute leukemia is proposed, to which other cutting-edge approaches, such as transfer learning and ensemble techniques are adapted.

For acute leukemia detection, a two-staged feature fusion-based stacked ensembling integrated CAD system employing pretrained, fine-tuned EfficientNetB7 and MobileNetV3Large models is presented. Initially, enhanced microscopic smear images were generated from raw smear images, and the proposed architecture performed significantly well in detecting acute leukemia. This led us to employ our proposed approach in detecting acute lymphocytic leukemia and acute myeloid leukemia, and it showed superior performance on the ALLIDB1, ALLIDB2, and ASH datasets, obtaining 100.0%, 99.4%, and 99.8% accuracies, respectively.

Our proposed automated CAD system aims to provide hematologists with a clinical tool that can be used as a second recommendation for automated acute leukemia diagnosis utilizing microscopic smear images by demonstrating exceptional performance. Despite the fact that it has showed excellent results, there is still opportunity for improvement. We will be working on this in the future to add more perfection to our proposed architecture.

## REFERENCES

[1] R. L. Siegel, A. N. Giaquinto, and A. Jemal, "Cancer statistics, 2024," *CA, A Cancer J. Clinicians*, vol. 74, no. 1, pp. 12–49, Jan. 2024.

[2] J. M. Bennett, D. Catovsky, M.-T. Daniel, G. Flandrin, D. A. Galton, H. R. Gralnick, and C. Sultan, "Proposals for the classification of the acute leukaemias French–American–British (FAB) co-operative group," *Brit. J. Haematol.*, vol. 33, no. 4, pp. 451–458, 1976.

[3] P. K. Das, S. Meher, R. Panda, and A. Abraham, "A systematic review on recent advancements in deep and machine learning based detection and classification of acute lymphoblastic leukemia," *IEEE Access*, vol. 10, pp. 81741–81763, 2022.

[4] K. Al-Dulaimi, J. Banks, K. Nugyen, A. Al-Sabaawi, I. Tomeo-Reyes, and V. Chandran, "Segmentation of white blood cell, nucleus and cytoplasm in digital haematology microscope images: A review-challenges, current and future potential techniques," *IEEE Rev. Biomed. Eng.*, vol. 14, pp. 290–306, 2021.

[5] S. Mishra, B. Majhi, and P. K. Sa, "Texture feature based classification on microscopic blood smear for acute lymphoblastic leukemia detection," *Biomed. Signal Process. Control*, vol. 47, pp. 303–311, Jan. 2019.

[6] L. H. S. Vogado, R. M. S. Veras, F. H. D. Araujo, R. R. V. Silva, and K. R. T. Aires, "Leukemia diagnosis in blood slides using transfer learning in CNNs and SVM for classification," *Eng. Appl. Artif. Intell.*, vol. 72, pp. 415–422, Jun. 2018.

[7] S. Alagu, N. P. Ahana, and B. K. Bhoopathy, "Automatic detection of acute lymphoblastic leukemia using UNET based segmentation and statistical analysis of fused deep features," *Appl. Artif. Intell.*, vol. 35, no. 15, pp. 1952–1969, Dec. 2021.

[8] L. Boldu, A. Merino, S. Alférez, A. Molina, A. Acevedo, and J. Rodellar, "Automatic recognition of different types of acute leukaemia in peripheral blood by image analysis," *J. Clin. Pathol.*, vol. 72, no. 11, pp. 755–761, Nov. 2019.

[9] P. K. Das, P. Jadoun, and S. Meher, "Detection and classification of acute lymphocytic leukemia," in *Proc. IEEE-HYDCON*, Jul. 2020, pp. 1–5.

[10] P. K. Das and S. Meher, "Transfer learning-based automatic detection of acute lymphocytic leukemia," in *Proc. Nat. Conf. Commun. (NCC)*, Jul. 2021, pp. 1–6.

[11] P. K. Das, B. Sahoo, and S. Meher, "An efficient detection and classification of acute leukemia using transfer learning and orthogonal softmax layer-based model," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 20, no. 3, pp. 1817–1828, May/Jun. 2023.

[12] S. Paswan and Y. K. Rathore, "Detection and classification of blood cancer from microscopic cell images using SVM KNN and NN classifier," *Int. J. Adv. Res. Ideas Innov. Technol.*, vol. 3, pp. 315–324, Jan. 2017.

[13] N. Patel and A. Mishra, "Automated leukaemia detection using microscopic images," *Proc. Comput. Sci.*, vol. 58, pp. 635–642, Jan. 2015.

[14] E. Abdulhay, M. A. Mohammed, D. A. Ibrahim, N. Arunkumar, and V. Venkatraman, "Computer aided solution for automatic segmenting and measurements of blood leucocytes using static microscope images," *J. Med. Syst.*, vol. 42, no. 4, pp. 1–12, Apr. 2018.

[15] P. K. Das, S. Meher, R. Panda, and A. Abraham, "An efficient blood-cell segmentation for the detection of hematological disorders," *IEEE Trans. Cybern.*, vol. 52, no. 10, pp. 10615–10626, Oct. 2022.

[16] L. B. Dorini, R. Minetto, and N. J. Leite, "Semiautomatic white blood cell segmentation based on multiscale analysis," *IEEE J. Biomed. Health Informat.*, vol. 17, no. 1, pp. 250–256, Jan. 2013.

[17] J. Zhao, M. Zhang, Z. Zhou, J. Chu, and F. Cao, "Automatic detection and classification of leukocytes using convolutional neural networks," *Med. Biol. Eng. Comput.*, vol. 55, no. 8, pp. 1287–1301, Aug. 2017.

[18] D. R. Sarvamangala and R. V. Kulkarni, "Convolutional neural networks in medical image understanding: A survey," *Evol. Intell.*, vol. 15, no. 1, pp. 1–22, Mar. 2022.

[19] M. L. Claro, R. D. M. S. Veras, A. M. Santana, L. H. S. Vogado, G. B. Junior, F. N. S. D. Medeiros, and J. M. R. S. Tavares, "Assessing the impact of data augmentation and a combination of CNNs on leukemia classification," *Inf. Sci.*, vol. 609, pp. 1010–1029, Sep. 2022.

[20] O. Bailo, D. Ham, and Y. M. Shin, "Red blood cell image generation for data augmentation using conditional generative adversarial networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2019, pp. 1039–1048.

[21] Y. Chen, X.-H. Yang, Z. Wei, A. A. Heidari, N. Zheng, Z. Li, H. Chen, H. Hu, Q. Zhou, and Q. Guan, "Generative adversarial networks in medical image augmentation: A review," *Comput. Biol. Med.*, vol. 144, May 2022, Art. no. 105382.

[22] P. P. Banik, R. Saha, and K.-D. Kim, "An automatic nucleus segmentation and CNN model based classification method of white blood cell," *Exp. Syst. Appl.*, vol. 149, Jul. 2020, Art. no. 113211.

[23] S. Mishra, B. Majhi, and P. K. Sa, "GLRLM-based feature extraction for acute lymphoblastic leukemia (ALL) detection," in *Recent Findings in Intelligent Computing Techniques*, vol. 2. Berlin, Germany: Springer, 2018, pp. 399–407.

[24] P. K. Das and S. Meher, "An efficient deep convolutional neural network based detection and classification of acute lymphoblastic leukemia," *Exp. Syst. Appl.*, vol. 183, Nov. 2021, Art. no. 115311.

[25] P. K. Das, B. Nayak, and S. Meher, "A lightweight deep learning system for automatic detection of blood cancer," *Measurement*, vol. 191, Mar. 2022, Art. no. 110762.

[26] A. Genovese, M. S. Hosseini, V. Piuri, K. N. Plataniotis, and F. Scotti, "Acute lymphoblastic leukemia detection based on adaptive unsharpening and deep learning," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Jun. 2021, pp. 1205–1209.

[27] S. Shafique, S. Tehsin, S. Anas, and F. Masud, "Computer-assisted acute lymphoblastic leukemia detection and diagnosis," in *Proc. 2nd Int. Conf. Commun., Comput. Digit. Syst. (C-CODE)*, Mar. 2019, pp. 184–189.

[28] Z. F. Mohammed and A. A. Abdulla, "An efficient CAD system for ALL cell identification from microscopic blood images," *Multimedia Tools Appl.*, vol. 80, no. 4, pp. 6355–6368, Feb. 2021.

[29] A. T. Sahlol, A. M. Abdeldaim, and A. E. Hassanien, "Automatic acute lymphoblastic leukemia classification model using social spider optimization algorithm," *Soft Comput.*, vol. 23, no. 15, pp. 6345–6360, Aug. 2019.

[30] R. D. Labati, V. Piuri, and F. Scotti, "All-IDB: The acute lymphoblastic leukemia image database for image processing," in *Proc. 18th IEEE Int. Conf. Image Process.*, Sep. 2011, pp. 2045–2048.

[31] *American Society of Hematology—Hematology.org*. Accessed: Jun. 6, 2023. [Online]. Available: https://www.hematology.org/

[32] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. Van Der Laak, B. Van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Med. Image Anal.*, vol. 42, pp. 60–88, Dec. 2017.

[33] M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 6105–6114.

[34] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 248–255.

[35] A. Howard, M. Sandler, B. Chen, W. Wang, L.-C. Chen, M. Tan, G. Chu, V. Vasudevan, Y. Zhu, R. Pang, H. Adam, and Q. Le, "Searching for MobileNetV3," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 1314–1324.

[36] Y. Bengio, F. Bastien, A. Bergeron, N. Boulanger-Lewandowski, T. Breuel, Y. Chherawala, M. Cisse, M. Cote, D. Erhan, and J. Eustache, "Deep learners benefit more from out-of-distribution examples," in *Proc. 14th Int. Conf. Artif. Intell. Statist.*, 2011, pp. 164–172.

[37] Y. Bengio, "Deep learning of representations for unsupervised and transfer learning," in *Proc. ICML Workshop Unsupervised Transf. Learn.*, 2012, pp. 17–36.

[38] H. E. Kim, A. Cosa-Linan, N. Santhanam, M. Jannesari, M. E. Maros, and T. Ganslandt, "Transfer learning for medical image classification: A literature review," *BMC Med. Imag.*, vol. 22, no. 1, p. 69, Dec. 2022.

[39] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Jan. 2009.

[40] R. Ribani and M. Marengoni, "A survey of transfer learning for convolutional neural networks," in *Proc. 32nd SIBGRAPI Conf. Graph., Patterns Images Tuts. (SIBGRAPI-T)*, Oct. 2019, pp. 47–57.

[41] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?" in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 27, 2014, pp. 1–9.

[42] R. Khandekar, P. Shastry, S. Jaishankar, O. Faust, and N. Sampathila, "Automated blast cell detection for acute lymphoblastic leukemia diagnosis," *Biomed. Signal Process. Control*, vol. 68, Jul. 2021, Art. no. 102690.

[43] R. R. Maaliw, A. S. Alon, A. C. Lagman, M. B. Garcia, J. A. B. Susa, R. C. Reyes, M. C. Fernando-Raguro, and A. A. Hernandez, "A multistage transfer learning approach for acute lymphoblastic leukemia classification," in *Proc. IEEE 13th Annu. Ubiquitous Comput., Electron. Mobile Commun. Conf. (UEMCON)*, Oct. 2022, pp. 0488–0495.

[44] H. Abed and A. Mohammed, "Proposing an efficient CNN model for detection of acute lymphoblastic leukemia (ALL) using transfer learning," *Int. J. Advance Comput. Eng. Netw.*, vol. 10, no. 8, pp. 53–56, 2022.

[45] A. Kumar, J. Rawat, I. Kumar, M. Rashid, K. U. Singh, Y. D. Al-Otaibi, and U. Tariq, "Computer-aided deep learning model for identification of lymphoblast cell using microscopic leukocyte images," *Exp. Syst.*, vol. 39, no. 4, May 2022, Art. no. e12894.

[46] M. Y. Al-khuzaie, S. A. Zearah, and N. J. Mohammed, "Developing an efficient VGG19-based model and transfer learning for detecting acute lymphoblastic leukemia (ALL)," in *Proc. 5th Int. Congr. Human-Comput. Interact., Optim. Robotic Appl. (HORA)*, 2023, pp. 1–5.

[47] K. K. Jha and H. S. Dutta, "Mutual information based hybrid model and deep learning for acute lymphocytic leukemia detection in single cell blood smear images," *Comput. Methods Programs Biomed.*, vol. 179, Oct. 2019, Art. no. 104987.

**MD HASIB AL MUZDADID HAQUE HIMEL** received the B.Sc. degree in computer science & engineering (CSE) from the Department of Computer Science & Engineering, Rajshahi University of Engineering & Technology (RUET), Rajshahi, Bangladesh. His research interests include artificial intelligence, machine learning, deep learning, computer vision, pattern recognition, image processing, medical image processing, and federated learning. He was awarded scholarships for excellent academic achievements, in 2008, 2011, 2013, and 2016. He was also one of the winners of the 2018 National Game Development Idea Contest. In 2022, he has been upgraded to the Grandmaster Tier on Kaggle.

**MD. AL MEHEDI HASAN** (Member, IEEE) received the B.Sc., M.Sc., and Ph.D. degrees in computer science & engineering from the Department of Computer Science & Engineering, University of Rajshahi, Rajshahi, Bangladesh. He was a Lecturer, an Assistant Professor, an Associate Professor, and a Professor with the Department of Computer Science & Engineering, Rajshahi University of Engineering & Technology (RUET), Rajshahi. Recently, he completed his postdoctoral research with the School of Computer Science and Engineering, The University of Aizu, Aizuwakamatsu, Japan. He has coauthored more than 130 publications and has published in widely cited journals and conferences. His research interests include machine learning, deep learning, bioinformatics, health informatics, computer vision, probabilistic and statistical inference, medical image processing, sensor-based data analysis, human–computer interaction, operating systems, computer networks, and security.

**TARO SUZUKI** received the B.S. degree from Tokyo University of Science, in 1987, the M.E. degree from the University of Tsukuba, in 1993, and the D.S. degree from The University of Tokyo, in 1998. He was an Engineer with Mitsubishi Electric Toubu Computer Company, from 1987 to 1991. He was a Research Associate with the Institute of Electronics and Information Science and the TARA Center, University of Tsukuba, from 1995 to 1998; the School of Information Science, JAIST, until 2000; and the Research Institute of Electric Communication, Tohoku University, until 2001. He was an Assistant Professor with the School of Computer Science and Engineering, The University of Aizu, until 2006, where he was an Associate Professor, until 2012, and has been a Senior Associate Professor, since 2013. His research interests include language theory with automata, functional programming, formal methods, and foundation of deep learning and reinforcement learning. He is a member of ACM.

**JUNGPIL SHIN** (Senior Member, IEEE) received the B.Sc. degree in computer science and statistics and the M.Sc. degree in computer science from Pusan National University, South Korea, in 1990 and 1994, respectively, and the Ph.D. degree in computer science and communication engineering from Kyushu University, Japan, in 1999, under a scholarship from the Japanese Government (MEXT). He was an Associate Professor, a Senior Associate Professor, and a Full Professor with the School of Computer Science and Engineering, The University of Aizu, Japan in 1999, 2004, and 2019, respectively. He has coauthored more than 400 published papers for widely cited journals and conferences. His research interests include pattern recognition, image processing, computer vision, machine learning, human–computer interaction, non-touch interfaces, human gesture recognition, automatic control, Parkinson's disease diagnosis, ADHD diagnosis, user authentication, machine intelligence, bioinformatics, and handwriting analysis, recognition, and synthesis. He is a member of ACM, IEICE, IPSJ, KISS, and KIPS. He served as the program chair and as a program committee member for numerous international conferences. He serves as an Editor for IEEE journals, Springer, Sage, Taylor & Francis, *Sensors* (MDPI) and *Electronics*, and Tech Science. He serves as an Editorial Board Member of Scientific Reports. He serves as a reviewer for several major IEEE and SCI journals.