

HAQUE, R., AL SAKIB, A., HOSSAIN, M.F., ISLAM, F., AZIZ, F.I., AHMED, M.R., KANNAN, S., ROHAN, A. and HASAN, M.J.
2024. Advancing early leukemia diagnostics: a comprehensive study incorporating image processing and transfer learning. *BioMedInformatics* [online], 4(2), pages 966-991. Available from:
<https://doi.org/10.3390/biomedinformatics4020054>

Advancing early leukemia diagnostics: a comprehensive study incorporating image processing and transfer learning.

HAQUE, R., AL SAKIB, A., HOSSAIN, M.F., ISLAM, F., AZIZ, F.I., AHMED, M.R., KANNAN, S., ROHAN, A. and HASAN, M.J.

2024



Article

Advancing Early Leukemia Diagnostics: A Comprehensive Study Incorporating Image Processing and Transfer Learning

Rezaul Haque¹, Abdullah Al Sakib², Md Forhad Hossain³, Fahadul Islam¹, Ferdous Ibne Aziz⁴, Md Redwan Ahmed¹, Somasundar Kannan⁵, Ali Rohan⁶ and Md Junayed Hasan^{6,*}

- ¹ Department of Computer Science and Engineering, East West University, Dhaka 1212, Bangladesh; rezaulh603@gmail.com (R.H.); fahadulislam.user@outlook.com (F.I.); redrik278@gmail.com (M.R.A.)
² Department of Information Technology, Westcliff University, Irvine, CA 92614, USA; sakibabdulla685@gmail.com
³ Department of Information Technology, St. Francis College, New York, NY 11201-9902, USA; forhad6951@gmail.com
⁴ Department of Artificial Intelligence, Woosong University, Daejeon 34606, Republic of Korea; shifatshaheen92@gmail.com
⁵ School of Engineering, Robert Gordon University, Aberdeen AB10 7AQ, UK; s.kannan1@rgu.ac.uk
⁶ National Subsea Centre, Robert Gordon University, Aberdeen AB10 7AQ, UK; a.rohan@rgu.ac.uk
* Correspondence: j.hasan@rgu.ac.uk



Citation: Haque, R.; Al Sakib, A.; Hossain, M.F.; Islam, F.; Ibne Aziz, F.; Ahmed, M.R.; Kannan, S.; Rohan, A.; Hasan, M.J. Advancing Early Leukemia Diagnostics: A Comprehensive Study Incorporating Image Processing and Transfer Learning. *BioMedInformatics* **2024**, *4*, 966–991. <https://doi.org/10.3390/biomedinformatics4020054>

Academic Editors: Federico Mastroleo, Angela Ammirabile, Giulia Marvaso and Hans Binder

Received: 15 January 2024

Revised: 7 March 2024

Accepted: 25 March 2024

Published: 1 April 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: Disease recognition has been revolutionized by autonomous systems in the rapidly developing field of medical technology. A crucial aspect of diagnosis involves the visual assessment and enumeration of white blood cells in microscopic peripheral blood smears. This practice yields invaluable insights into a patient's health, enabling the identification of conditions of blood malignancies such as leukemia. Early identification of leukemia subtypes is paramount for tailoring appropriate therapeutic interventions and enhancing patient survival rates. However, traditional diagnostic techniques, which depend on visual assessment, are arbitrary, laborious, and prone to errors. The advent of ML technologies offers a promising avenue for more accurate and efficient leukemia classification. In this study, we introduced a novel approach to leukemia classification by integrating advanced image processing, diverse dataset utilization, and sophisticated feature extraction techniques, coupled with the development of TL models. Focused on improving accuracy of previous studies, our approach utilized Kaggle datasets for binary and multiclass classifications. Extensive image processing involved a novel LoGMH method, complemented by diverse augmentation techniques. Feature extraction employed DCNN, with subsequent utilization of extracted features to train various ML and TL models. Rigorous evaluation using traditional metrics revealed Inception-ResNet's superior performance, surpassing other models with F1 scores of 96.07% and 95.89% for binary and multiclass classification, respectively. Our results notably surpass previous research, particularly in cases involving a higher number of classes. These findings promise to influence clinical decision support systems, guide future research, and potentially revolutionize cancer diagnostics beyond leukemia, impacting broader medical imaging and oncology domains.

Keywords: acute and chronic leukemia; early detection; image processing; machine learning; transfer learning; medical imaging; clinical decision support

1. Introduction

Leukemia affects the generation of white blood cells essential for fighting infections and diseases [1]. The development is influenced by a complex combination of genetic, environmental, and lifestyle elements [2,3]. Despite connections to genetic mutations, chemical exposure, radiation, and viral infections, pinpointing its exact origins can be challenging. Dealing with leukemia comes with many challenges such as persistent tiredness, bruising, and weight reduction [4,5] Moreover, conducting detailed examinations and bone marrow

aspiration can increase physical strain [6]. Furthermore, mental health can be severely affected by the illness's unpredictability and the hard therapies, such as chemotherapy.

Leukemia appears in two main forms: acute and chronic [7]. Acute leukemia is characterized by the rapid buildup of immature blood cells, which is further classified into Acute Lymphoid Leukemia (ALL) and Acute Myeloid Leukemia (AML) [8]. Acute lymphoblastic leukemia mainly impacts developing lymphoid cells and is more common in children, whereas acute myeloid leukemia affects myeloid cells in individuals of all ages, characterized by various genetic mutations. Chronic leukemia advances gradually, marked by excessive production of abnormal, mature blood cells [9]. Chronic Lymphocytic Leukemia (CLL) is commonly found in older individuals, Chronic Myeloid Leukemia (CML) is frequently associated with the Philadelphia chromosome and responds well to targeted treatments, and Hairy cell leukemia (H) is a rare variant of CLL with a positive outlook when treated with purine analogs [10–12]. Treatment of these leukemia subtypes include interferon therapy or splenectomy.

Regular follow-up is crucial for monitoring and managing the disease. When diagnosing leukemia, it's crucial to carefully analyze white blood cells under a microscope, looking at their size, shape, and the presence of Auer rods. Additional tests such as bone marrow aspiration, flow cytometry, and complete blood counts are also essential. These discoveries help categorize the illness and guide treatment choices. Nevertheless, differing viewpoints among pathologists frequently result in varying diagnoses [13]. Evaluating blood images manually can be difficult due to issues like noise, blur, and obscured cells, especially in complex situations [14]. Identifying leukemia subtypes early is essential for predicting outcomes and tailoring treatment plans to improve patient well-being and reduce side effects [15,16]. However, obstacles continue to arise because of the labor-intensive process of manual identification, the growing amount of diagnostic data, and the changing genetic complexities.

Several studies [17–19] have investigated incorporating Machine Learning (ML) into automated pathological diagnosis, especially with the increase in digitized microscopic images. Machine learning algorithms use characteristics such as morphology and size to recognize cell types and abnormalities, improving accurate classification and allowing pathologists to concentrate on intricate aspects of diagnosis. Nevertheless, there are challenges such as the requirement for large datasets, complexities in interpretation, and the risk of overfitting [20,21]. Addressing these challenges involves leveraging pre-trained Transfer Learning (TL) models to overcome data constraints, minimize training time and expenses, while sustaining top performance [22,23]. These models, with the ability to continuously learn, could potentially transform pathology diagnosis and improve patient care by enhancing diagnostic speed. Yet, optimizing TL models for leukemia classification encounters difficulties because of the complex features found in medical images. Developing a TL model that can effectively classify leukemia is crucial. It should handle challenges like class imbalance, clinical relevance, generalization, and interpretability.

This research aims to enhance leukemia classification accuracy and reliability by integrating advanced image processing techniques and TL models. Figure 1 depicts the research workflow. Two datasets from Kaggle were collected for binary (ALL vs. Normal) and multiclass (ALL, AML, CML, CLL, H) classification. Binary classification is crucial for understanding the underlying pathology, aiding in early detection and management of specific subtypes. Examining multiclass classification enables a detailed analysis of leukemia subtypes, offering insights into genetic and morphological traits. Through the utilization of binary and multiclass classification techniques in our study, we aim to offer a holistic perspective on the reliability of leukemia classification in diverse diagnostic situations. Preprocessing involved resizing, normalization, and a novel Laplacian of Gaussian-based modified high-boosting (LoGMH) technique. Augmentation techniques, including brightness adjustments, padding, flips, translation, rotations, zooming, and cropping, were applied to enhance dataset variability. Feature extraction employed a Deep Convolutional Neural Network (DCNN) to capture informative features from preprocessed

and augmented images. ML models, such as K-Nearest Neighbors (KNN), Multilayer Perceptron (MLP), Random Forest (RF), Support Vector Machine (SVM), and Stochastic Gradient Descent (SGD), were trained on the extracted features for binary and multiclass classification. Additionally, TL was implemented using models like AlexNet, RetinaNet, XceptionNet, Inception-ResNet, and CenterNet. Evaluation will be conducted through 5-fold cross-validation and traditional metrics to ensure robust assessments of accuracy and generalization.

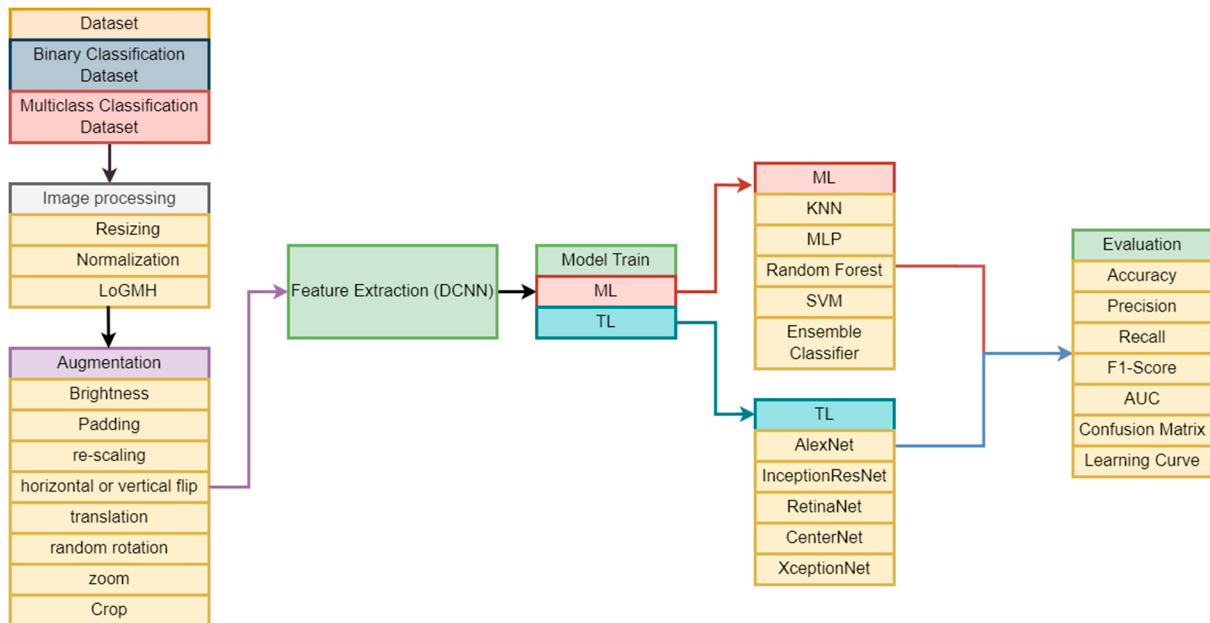


Figure 1. Proposed research methodology.

Our approach stands out for its innovative fusion of advanced TL techniques with comprehensive image processing methods tailored for leukemia classification. By combining these cutting-edge methodologies, we depart significantly from conventional approaches, offering a more resilient and effective solution for leukemia classification. The research aspires to contribute valuable insights influencing clinical decision support systems, guiding future research, and potentially revolutionizing cancer diagnostics, extending its impact beyond leukemia into broader medical imaging and oncology domains. Our contributions are as follows:

- **Development of Diverse ML and TL Models:** We contribute a range of ML and TL models tailored for accurate and reliable binary and multiclass classification of leukemia cells. This diverse set of models provides flexibility and options for effective classification in different scenarios.
- **Case Study with Two Datasets:** Our research conducts a comprehensive case study by considering two datasets, offering a detailed analysis of the effectiveness of the experimental models. This approach ensures a robust evaluation, addressing the reliability and generalization aspects of the proposed models.
- **Empirical Analysis for Model Evaluation:** We present a detailed empirical analysis, evaluating the effectiveness and efficiency of each proposed model for accurate binary and multiclass classification of leukemia cells. This thorough evaluation contributes insights into the performance of each model, aiding in the selection of appropriate models for specific applications.
- **Comparative Performance Analysis:** To benchmark our models, we conduct a comprehensive analysis comparing the performance and efficiency of the TL model against ML classifiers. This comparative study highlights the strengths and advancements of our proposed models in the context of leukemia cell classification.

The following sections are organized as follows: Section 2 presents a comprehensive overview of the latest research on leukemia classification. Section 3 provides an overview of the proposed method. Sections 4 and 5 cover the experimental evaluation and discussions, respectively. In conclusion, Section 6 wraps up the work and delves into potential areas for future research.

2. Related Works

The classification of leukemia subtypes has been a focal point in recent studies, leveraging ML techniques to enhance diagnostic accuracy and treatment outcomes. Hamidah et al. [24] targeted ALL, emphasizing the role of fusion genes resulting from translocations for subtype classification. Employing Multiclass Support Vector Machine Recursive Feature Elimination (MSVM-RFE), they identified MSVM Polynomial-Kernel with $d = 4$ as the optimal method, achieving impressive metrics, including 94% accuracy, 96% precision, 95% recall, and a rapid running time of 0.66 seconds. Fauzi et al. [25] delved into the significance of early cancer detection, particularly in the context of microarray cancer data complexity. They introduced Principal Component Analysis (PCA) to streamline features, enhancing classification accuracy. The Fuzzy Support Vectors Machines (FSVM) method was employed, resulting in an accuracy of 87.69% without PCA and 96.92% with PCA, showcasing the pivotal role of feature reduction in cancer classification. Dasariraju et al. [26] addressed challenges in AML diagnosis by introducing a ML model for efficient detection and classification of immature leukocytes. Leveraging image processing techniques and proposing new nucleus color features, the model achieved 92.99% accuracy for detection and 93.45% for classification into four types, surpassing previous methods. More et al. [27] explored autonomous systems for visualizing white blood cells in the context of acute lymphoblastic leukemia. Employing morphological techniques and SVM, their model demonstrated effectiveness in identifying leukemia-associated cells, with EMC-SVM overcoming the limitations of a single classifier and successfully categorizing white blood cells in sample blood smear images. Kashef et al. [28] shifted the focus to treatment outcomes of pediatric ALL patients, emphasizing the integration of clinical and medical data for ML-based classification. Analyzing data from 241 patients, XGBoost outperformed other classifiers with 88.5% accuracy in the first scenario, while SVM showed superiority with 94.90% accuracy in the second scenario.

However, previous studies using ML include challenges in handling complex microarray cancer data, reliance on manual examination methods, and the need for effective feature selection. Additionally, they faced issues related to the diversity and size of datasets, and some struggled with the interpretability of results. To overcome these limitations, researchers have turned to Convolutional Neural Network (CNN) models. CNNs excel in handling image data, making them suitable for tasks involving microscopic images of blood cells or tissue samples [29]. Focusing on acute lymphoblastic leukemia classification, Rahman et al. [30] developed a novel approach combining ML and Deep Learning. The research utilized a pipeline involving dataset construction, feature extraction using pre-trained CNN architectures, and classification with conventional classifiers. Achieving a maximum accuracy of 99.84%, the proposed model integrated ResNet50 CNN, SVC feature selector, and LR classifiers after incorporating Particle Swarm Optimization and Cat Swarm Optimization. Ahmed et al. [31] proposed a novel approach for diagnosing all leukemia subtypes using CNN and investigated the impact of data augmentation. Utilizing the ALL-IDB and ASH Image Bank datasets, the CNN model achieved 88.25% accuracy for leukemia versus healthy classification and 81.74% for multi-class classification of all subtypes. In a recent study, Saeed et al. 2023 [32] aimed to automate the diagnosis of ALL using a CNN model. Simulations on the Acute Lymphoblastic Leukemia-IDB 1 and Leukemia-IDB 2 datasets, along with data augmentation techniques, addressed overfitting. The proposed model achieved a remarkable 99.61% accuracy in ALL diagnosis, showcasing its effectiveness compared to existing methods in the field. By leveraging CNN models,

these studies aimed to enhance the accuracy and generalization capabilities of their models, providing a more sophisticated and automated approach to leukemia classification.

However, CNNs typically require a substantial amount of labeled data for training, which can be challenging to acquire, especially for rare subtypes of leukemia. TL addresses these limitations by leveraging pre-trained models on unrelated datasets, enabling the network to capture generic features. This approach allows CNNs to be effective even with limited labeled data for leukemia, as fine-tuning on specific datasets refines the model for accurate classification. Loey et al. [33] focused on leukemia detection through two automated classification models based on blood microscopic images, emphasizing early detection for improved remission rates. Utilizing TL, the first model preprocessed images and extracted features using the pre-trained AlexNet, followed by classification with well-known classifiers. The second model fine-tuned AlexNet for both feature extraction and classification. Experimentation on a dataset with 2820 images confirmed the superior performance of the second model, achieving 100% classification accuracy. But specific details about the dataset's diversity, potential biases, or generalizability to varied clinical scenarios are not explicitly mentioned. Ghongade et al. [34] utilized deep learning and TL for predicting ALL in lymphocytes. A pre-trained CNN extracted features from microscopic blood cell images, and a TL-based classification model accurately categorized cells into leukemia and non-leukemia classes. Abir et al. [35] focused on automated detection of ALL, emphasizing the challenges in manual diagnosis and the potential of computer-aided models. Using deep learning and TL, the proposed strategy achieved 98.38% accuracy in classifying ALL, with InceptionV3 performing best among the models. The introduction of local interpretable model-agnostic explanations (LIME) addressed explainability concerns in deep learning, ensuring the reliability of classifications. Experimental results validated the method's suitability for identifying ALL, offering valuable assistance to medical examiners. Yet, the study lacks detailed insights into potential false-positive or false-negative rates, crucial for assessing the model's clinical utility. Additionally, the generalizability of these findings to diverse patient populations and the robustness of the models in real-world clinical settings remain areas for further investigation.

Initial studies utilized ML techniques to classify ALL subtypes based on fusion genes, achieving impressive accuracy metrics. Nevertheless, the difficulties in managing intricate microarray cancer data and comprehending the outcomes led to a transition towards CNN models. Nevertheless, CNNs present difficulties in obtaining labeled data, which can be overcome through the use of TL. Although these studies showed notable progress, certain limitations, such as the diversity of the dataset and its generalizability, were noted. Additional research is necessary to fill these gaps and guarantee the practicality and reliability of the proposed models in real-life situations. In addition, there are areas in research that have not been thoroughly explored, such as acute and chronic subtypes. There is also a need for better interpretability and explainability in the findings. Additionally, there are challenges in accessing and standardizing data, and it is important to have a more comprehensive understanding of how to integrate models into clinical workflows. In order to ensure clinical applicability, promote successful implementation in practical healthcare environments, and advance the discipline, it is vital to address these gaps.

Our paper aims to fill the gaps in existing research on leukemia classification. Initially, we utilize two datasets, one for binary classification and another for multiclass classification, to thoroughly investigate different leukemia subtypes. Additionally, our image processing technique, which involves the use of LoGMH, greatly enhances preprocessing and feature enhancement. Additionally, the augmentation techniques utilized, including brightness adjustments, padding, and various transformations, contribute to the model's resilience and effectively tackle issues associated with dataset diversity. Furthermore, our utilization of DCNNs for feature extraction demonstrates our commitment to employing cutting-edge techniques for precise representation. At last, our extensive model training, which includes traditional ML classifiers and TL models, enables a comprehensive evaluation and validation of various methodologies. The evaluation metrics, such as the confusion

matrix and learning curve, offer a comprehensive analysis of the model's performance. In our study, we strive to address research gaps by utilizing a combination of diverse datasets, advanced preprocessing techniques, sophisticated feature extraction methods, and a comprehensive evaluation approach for leukemia classification.

3. Materials and Methods

3.1. Data Description

For the purpose of binary classification, we employed the C-NMC dataset [36], which comprises 15,114 lymphocyte images obtained from 118 subjects. This dataset offers a diverse range of lymphocyte samples, including malignant and healthy cells, making it suitable for training and evaluating ML models for early leukemia diagnostics. These images were organized into three folders, each with its own name: "CNMC test data" contained 1867 cells, 1219 malignant cells from 13 subjects, and 648 healthy cells from 15 subjects; and "C-NMC training data" comprised 10,661 cells, 7272 malignant cells from 47 subjects, and 3389 healthy cells from 26 subjects. The dataset's organization into test and training sets, along with its substantial size and expert annotation by seasoned oncologists, enhances its suitability for research purposes. There are a total of 2133 photos in the test set and 8528 in the training set; 20% of the total images are from the test set. These files contain single cell images of lymphocytes, both benign and cancerous, that have been identified by seasoned oncologists. A representative subset of the dataset is illustrated in Figure 2.

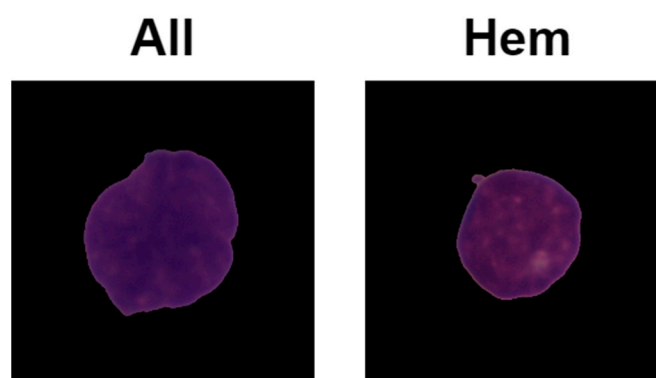


Figure 2. Sample of C-NMC dataset.

The Leukemia Dataset 0.2 [37] was utilized for multiclass classification, consisting of 5 classes divided into 2 folders: train and test. These include ALL, AML, CLL, CML, and H. With a large collection of blood smear images representing different leukemia types, this dataset facilitates the exploration of machine learning approaches for classifying diverse leukemia subtypes. The inclusion of multiple classes allows for a more nuanced analysis of the performance of the developed algorithms. We have a total of 20,000 images in the dataset. Every class contains a substantial number of images, totaling 4000. Every sample image of the five classes is displayed in

When conducting research involving medical datasets, especially those with sensitive patient information, ethical considerations and data privacy aspects are of utmost importance. In this study, measures are rigorously enforced to maintain patient confidentiality and privacy rights. Through a focus on ethical research practices and responsible data handling, the study aims to safeguard patient confidentiality and push forward early leukemia detection. Figure 3.

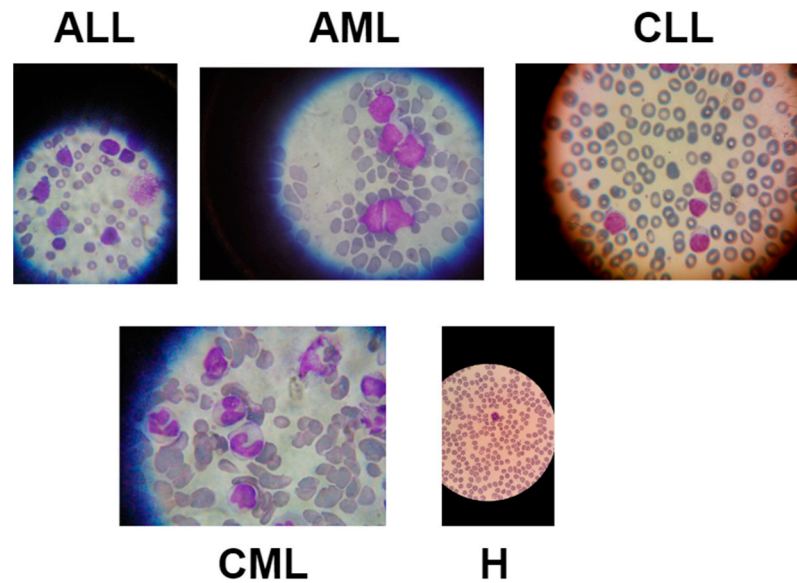


Figure 3. Sample image of each class of Leukemia Dataset 0.2.

3.2. Image Preprocessing

In the preprocessing phase, we employed a series of techniques to enhance the quality and variability of the dataset, optimizing it for subsequent leukemia classification. The images were resized to a standardized dimension of (227×227) pixels, facilitating consistent input dimensions for the subsequent stages. Normalization was applied to standardize pixel values across all images, ensuring uniformity in data representation [38]. A distinctive technique, LoGMH, was incorporated to accentuate relevant features within the images. The process begins with the application of Gaussian blurring to the input image, shown in Equation (1). This involves convolving the image with a Gaussian kernel $G(x, y)$, smoothing the pixel values and reducing noise. Here, (σ) is the standard deviation, and (x, y) are the spatial coordinates. The Laplacian operator (∇^2) is then applied to the blurred image, shown in Equation (2). This operator calculates the second spatial derivative of the image intensity, with respect to both (x) and (y) . Then, high-boost filter (Equation (3)) enhances high-frequency components in the image. Parameters (α) and (β) control the amplification of the original image and the Laplacian, respectively. Finally, the output is normalized using Equation (4) to ensure that pixel values fall within a standardized range, typically $[0, 1]$. Figures 4 and 5 illustrates each step of employed preprocessing technique on our experimental binary and multiclass dataset.

$$G(x, y) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/(2\sigma^2)} \quad (1)$$

$$\nabla^2 I(x, y) = \frac{\partial^2 I}{\partial x^2} + \frac{\partial^2 I}{\partial y^2} \quad (2)$$

$$H(x, y) = \alpha \cdot I(x, y) - \beta \cdot \nabla^2 I(x, y) \quad (3)$$

$$\text{Normalized_LoGMH}(x, y) = \frac{H(x, y) - \min(H)}{\max(H) - \min(H)} \quad (4)$$

LoGMH is a method that emphasizes important characteristics in images by detecting local changes in gradient magnitude [39]. It highlights slight variations in pixel intensity and gradient direction, which can reveal significant image features like edges, textures, and structural details [40]. Figures 4 and 5 demonstrate the transformative effects of integrating LoGMH into the preprocessing pipeline. The processed images appear enhanced compared to the original image. It exhibits sharper edges, clearer textures, and more pronounced structural details. This enhancement is particularly valuable in medical imaging applications,

where accurately identifying and distinguishing between different cell types is crucial for diagnosis [41]. This processed image serves as an improved representation of the original, with enhanced visual cues that facilitate more accurate and robust classification by the sub-sequent stages of the classification pipeline. This process is valuable in improving image features, making them more discernible. In addition, by integrating LoGMH into the pre-processing pipeline, the dataset goes through a transformation that boosts its diversity of features [39]. As a result, the classification model's generalization capacity improved and the dataset is better able to capture the variety of features seen in leukemia cell ages.

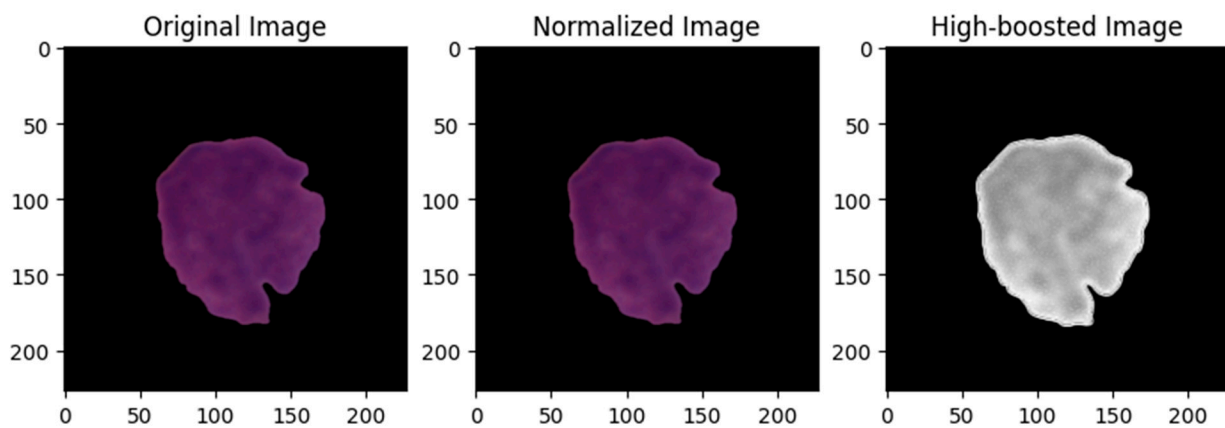


Figure 4. Image preprocessing techniques on C-NMC dataset.

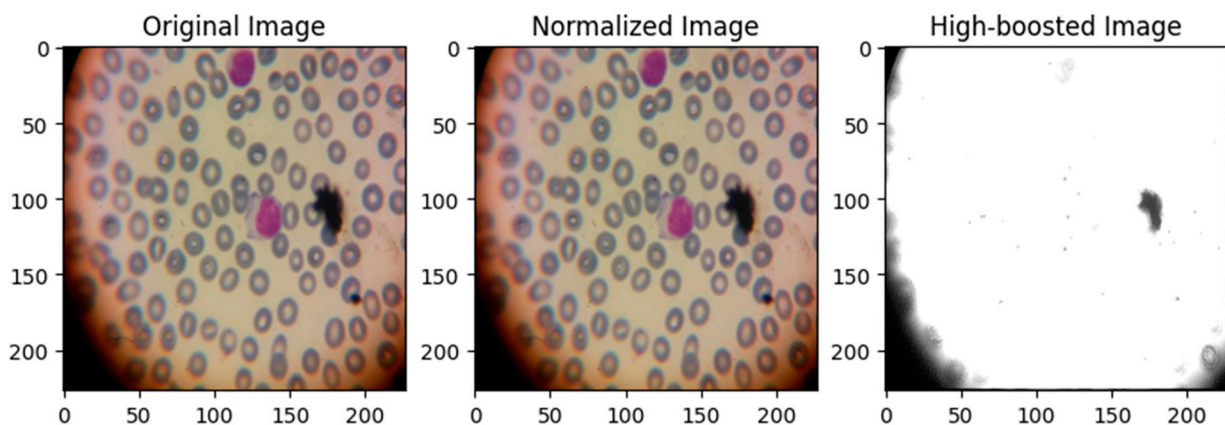


Figure 5. Image preprocessing techniques on multiclass dataset.

3.3. Image Augmentation

The datasets were divided into three parts: 75% for training, 5% for validation, and 20% for testing. This division was done strategically to ensure a balanced distribution for effective learning and assessment. Furthermore, a wide range of image augmentation techniques were applied to the dataset, which greatly improved the model's ability to handle variations in real-world medical images. Adjustments to brightness were made by manipulating pixel intensity, using parameters to determine the extent of the adjustment. The process of padding includes adding pixels to the boundaries of an image, adjusting the dimensions of the modified image, and using parameters to determine the scaling factor. Flipping an image either horizontally or vertically creates mirrored variations, while translation moves images both horizontally and vertically, with parameters that control the amount of translation. The random rotation feature allows for the rotation of images at various angles, while the zooming function can adjust the scale of the image. Additionally, the crop tool enables the selective extraction of specific portions of an image. The various

techniques used in this process helped create a dataset that was more adaptable, leading to better performance of the model during both training and testing.

3.4. Feature Extraction Using DCNN

Following meticulous image preprocessing, feature extraction is conducted through the utilization of a DCNN. The architecture used is derived from the CNN designed for ImageNet object detection. It consists of eight learned layers, including five convolutional layers and three fully connected layers. This study focuses on extracting features from the initial five layers of this architecture. These layers consist of a series of convolution operations, ReLU, Local Response Normalization (LRN), and pooling operations, as shown in Figure 6. The features extracted after the pooling operation in the fifth layer are considered appropriate for integration into classifiers to identify different types of leukemia. The first layer of this DCNN architecture consists of a convolutional layer that uses 96 filters of size $(11 \times 11 \times 3)$. Its main purpose is to extract low-level edge features. Following that, a ReLU layer improves the network’s non-linear properties by applying the rectifying function to every input value. The subsequent pooling layer conducts subsampling on small rectangular blocks from the previous layer, employing a max-pooling technique with (3×3) rectangular blocks and a 2-pixel interval. As a result, an LRN layer is applied to normalize brightness, which improves the visibility of important features and reduces the impact of irrelevant ones. The extracted features, especially those from the 5th pooling layer, provide crucial discriminative information for the subsequent classification of leukemia subtypes.

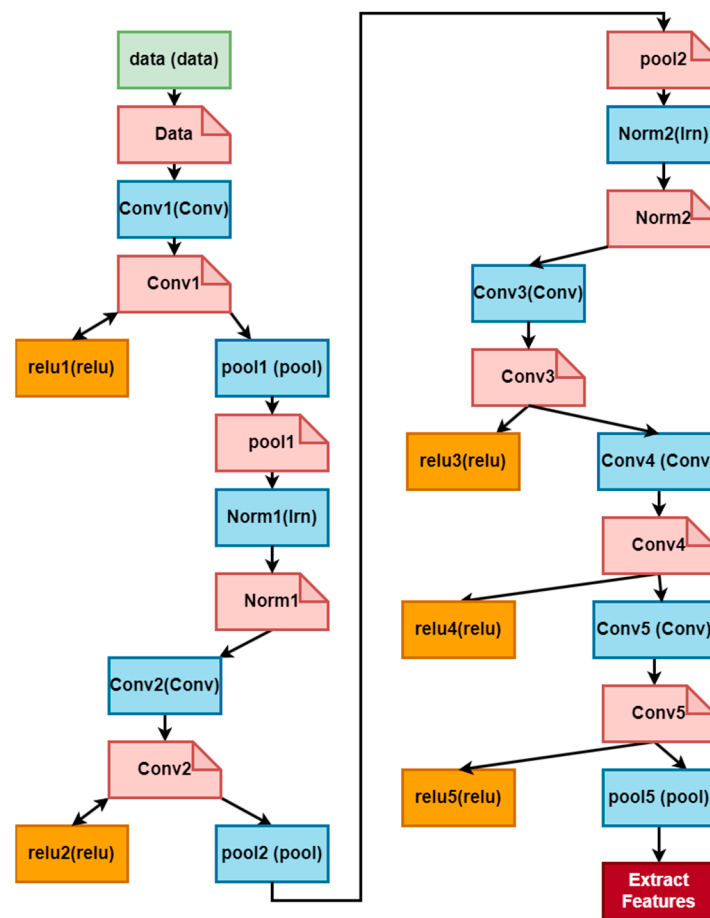


Figure 6. DCNN architecture to extract features from images.

3.5. Experimental Models

3.5.1. ML Models

After the extraction of features from preprocessed and augmenting training images, a diverse set of ML models is employed for binary and multiclass classification. The training parameters are configured for each model to ensure optimal learning and performance. The MLP utilizes a feedforward neural network architecture to map extracted image features to leukemia classes through the learning of weighted connections [42]. For MLP, we employed a feedforward neural network architecture with parameters set to a learning rate of 0.001, Adam optimizer, and 50 training epochs. RF, an ensemble of decision trees, leverages the collective decision-making of multiple trees to enhance accuracy and robustness [43,44]. RF was configured with 100 trees, no maximum depth constraint, and Gini as the criterion. SVM employs a non-linear classification approach, mapping the input features into a higher-dimensional space to find an optimal hyperplane for class separation [45]. We used a Radial Basis Function (RBF) kernel, with regularization parameter set to 1.0 and 'scale' as the gamma value. KNN adopts an instance-based learning strategy, classifying images based on the majority class of their k-nearest neighbors [46]. We employed the Euclidean distance metric with five neighbors. SGD, an optimization algorithm, iteratively adjusts model weights to minimize the loss function, aiding convergence towards optimal parameters [47,48]. We utilized a log loss function, a learning rate of 0.01, 100 training epochs, and L2 penalty.

When conducting our research, we carefully chose these models based on their unique strengths to ensure top performance in disease classification. MLP was selected for its ability to learn complex patterns and relationships within extracted image features. Then, RF was chosen for its strength and capacity to manage high-dimensional feature spaces, providing improved accuracy through ensemble decision-making. The non-linear classification method of SVM was favored since it was essential for translating input data into higher-dimensional spaces and determining the best class borders. Furthermore, KNN was chosen for its simple and intuitive classification approach, which is especially useful in situations with complex or ambiguous decision boundaries. Finally, SGD was implemented in order to optimize model training on large datasets while pre-serving reasonable processing times due to its scalability and computational efficiency.

3.5.2. TL Models

AlexNet

The AlexNet, a cutting-edge CNN framework, has a crucial impact on the progress of image classification. The model was first presented by Krizhevsky et al. [49] during the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) in 2012. It was selected for its proven track record in image classification tasks. Despite being one of the earlier CNN architectures, its deep layers are capable of learning rich hierarchical features from images, making it suitable for capturing intricate patterns in leukemia cell images [50]. The design consists of five convolutional layers, which are then followed by three fully connected layers. The convolutional layers employ convolution, ReLU activation, and max-pooling processes to extract hierarchical features from the input image. The fully linked layers further refine the retrieved features to perform the final classification. The architecture of the AlexNet model employed in our study is depicted in Figure 7.

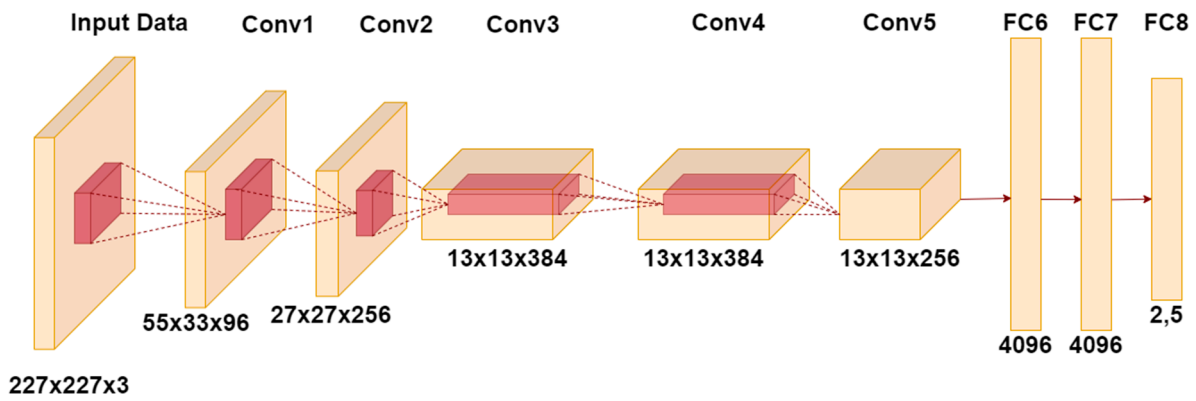


Figure 7. Employed architecture of AlexNet.

The mathematical representation of a convolutional layer in AlexNet is shown in Equation (5). Where, $h(x, y, i)$ is the output of the (i) -th feature map at position (x, y) , $w(a, b, i)$ represents the weights of the convolutional kernel, $x(x + a, y + b)$ denotes the input image, and $b(i)$ is the bias term. Models fully connected layers are represented in Equation (6). Where, (f_i) is the output of the (i) -th neuron, (w_{ij}) are the weights, and (h_j) represents the input from the previous layer. The softmax activation function is applied to the final fully connected layer for multiclass classification (Equation (7)). Where, $P(Y = i | x)$ represents the probability of the input (x) belonging to class (i) , and (j) is the number of classes.

$$h(x, y, i) = \sigma \left(\sum_{a=0}^{A-1} \sum_{b=0}^{B-1} w(a, b, i) \cdot x(x + a, y + b) + b(i) \right) \quad (5)$$

$$f_i = \sigma \left(\sum_{j=0}^{J-1} w_{ij} \cdot h_j + b_i \right) \quad (6)$$

$$P(Y = i | x) = \frac{e^{f_i}}{\sum_{j=0}^{J-1} e^{f_j}} \quad (7)$$

RetinaNet

RetinaNet, a highly regarded object detection model developed by Lin et al., is widely recognized for its exceptional performance in accurately detecting objects of different sizes [51]. The main innovation of RetinaNet is its unique focal loss function, which efficiently tackles the problem of class imbalance in object detection [52]. RetinaNet’s architecture utilizes a backbone network, like ResNet/ResNeXt, to extract feature maps [53]. As early detection of leukemia often involves identifying subtle abnormalities in cell morphology, RetinaNet’s object detection capabilities are valuable for locating and classifying leukemia cells within blood smear images [54]. The FPN utilizes the feature maps to create a multi-scale pyramid of feature maps, contributing to improved diagnostic accuracy enabling reliable object detection for objects of different sizes [55]. The model is designed with a focus on object detection, and its original formulation includes binary classification for anchor boxes, distinguishing between foreground and background. The architecture of the RetinaNet models in action is illustrated in Figure 8.

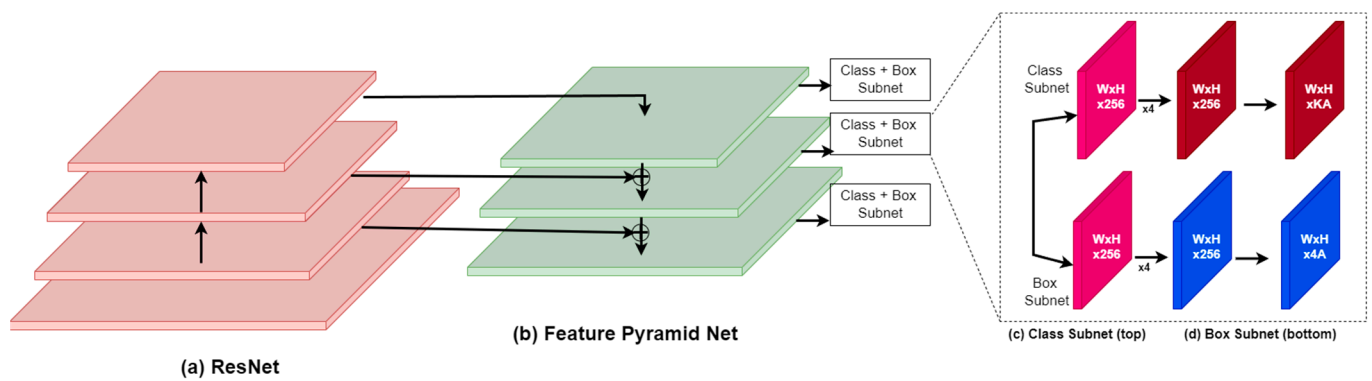


Figure 8. Architecture of RetinaNet model.

For multiclass classification, where multiple object classes are considered, the adaptation involves extending the binary focal loss to handle multiple classes. The multiclass focal loss is an extension of the binary focal loss and is shown in Equation (8). Where, (C) is the number of classes, $(p_{t,i})$ is the predicted probability of the correct class (i) , and (γ) is the focusing parameter. This formulation considers the sum of focal losses across all classes. The total multiclass loss for RetinaNet is then computed as a sum of focal losses for each anchor box and the smooth L1 loss for bounding box regression (Equation (9)). Where, $(p_{t,i})$, (p) , and (p^*) represent the predicted probability of the correct class (i) , predicted class probabilities, and ground truth class probabilities, respectively. Similarly, (b) , (b^*) , and (λ) denote predicted bounding box coordinates, ground truth bounding box coordinates, and the balancing parameter for smooth L1 loss. This adaptation allows RetinaNet to handle multiclass object detection tasks by extending its original binary classification framework to multiple classes.

$$FL(p_t) = -\sum_{i=1}^C (1 - p_{t,i})^\gamma \cdot \log(p_{t,i}) \tag{8}$$

$$Loss(p, p^*, b, b^*) = \sum_{i=1}^C FL(p_{t,i}) + \lambda \cdot SmoothL1(b, b^*) \tag{9}$$

XceptionNet

Depthwise separable convolutions are used in XceptionNet, an augmentation of the Inception architecture, to improve model performance and efficiency [56]. In order to classify leukemia into multiple classes, the network is modified to produce probabilities for every class. The model was selected for its innovative depth-wise separable convolutions, which enable efficient feature extraction while reducing computational complexity. Its pre-trained weights on ImageNet provide a strong foundation for transfer learning, enabling effective fine-tuning and adaptation to leukemia classification tasks [57]. The anticipated probability for every class is calculated by the last softmax layer. The general structure of XceptionNet incorporates numerous blocks of depthwise separable convolutions, followed by batch normalization and nonlinear activation functions. The XceptionNet architecture that we employed for our study is depicted in Figure 9.

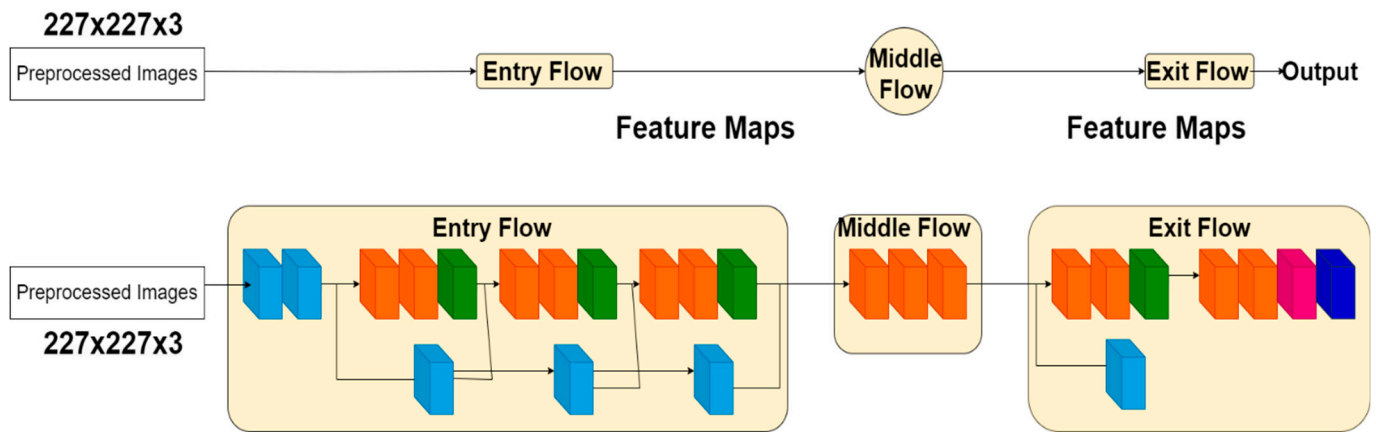


Figure 9. Architecture of XceptionNet model.

Let, (X) represent the input feature map, and (F_k) be the depthwise separable convolution operation in the k -th block. The output of the k -th block is computed using Equation (10). where (X_k) and (X_{k+1}) denote the input and output feature maps of the k -th block, respectively. The final classification layer is typically a fully connected layer with a softmax activation function for multiclass problems. For a given input feature vector (z) and the corresponding class probabilities (p_i) , the softmax function is defined in Equation (11). The training objective involves minimizing the categorical cross-entropy loss, which measures the dissimilarity between predicted and true class probabilities. For a given instance with ground truth label (y) and predicted probabilities (p_i) , the categorical cross-entropy loss is defined Equation (12). Training parameters, such as learning rate, batch size, and optimizer, are configured to optimize the model’s performance on the training dataset while preventing overfitting.

$$X_{k+1} = F_k(X_k) + X_k \tag{10}$$

$$p_i = \frac{e^{z_i}}{\sum_{j=1}^C e^{z_j}} \tag{11}$$

$$\text{Loss} = -\sum_{i=1}^C y_i \cdot \log(p_i) \tag{12}$$

InceptionResNet

InceptionResNet, an amalgamation of the Inception and ResNet architectures, combines the advantages of both networks for improved performance in multiclass leukemia classification. This architecture incorporates residual connections from the ResNet architecture, enhancing the network’s ability to capture intricate features [58,59]. Its multi-scale feature extraction capabilities are particularly beneficial for capturing nuanced details in cell images, enhancing classification accuracy. The structure involves multiple blocks, each comprising Inception modules and residual connections. Moreover, its residual connections facilitate efficient knowledge transfer and adaptation to leukemia diagnostic tasks [60,61]. Figure 10 illustrates the structure of the InceptionResNet employed in our study.

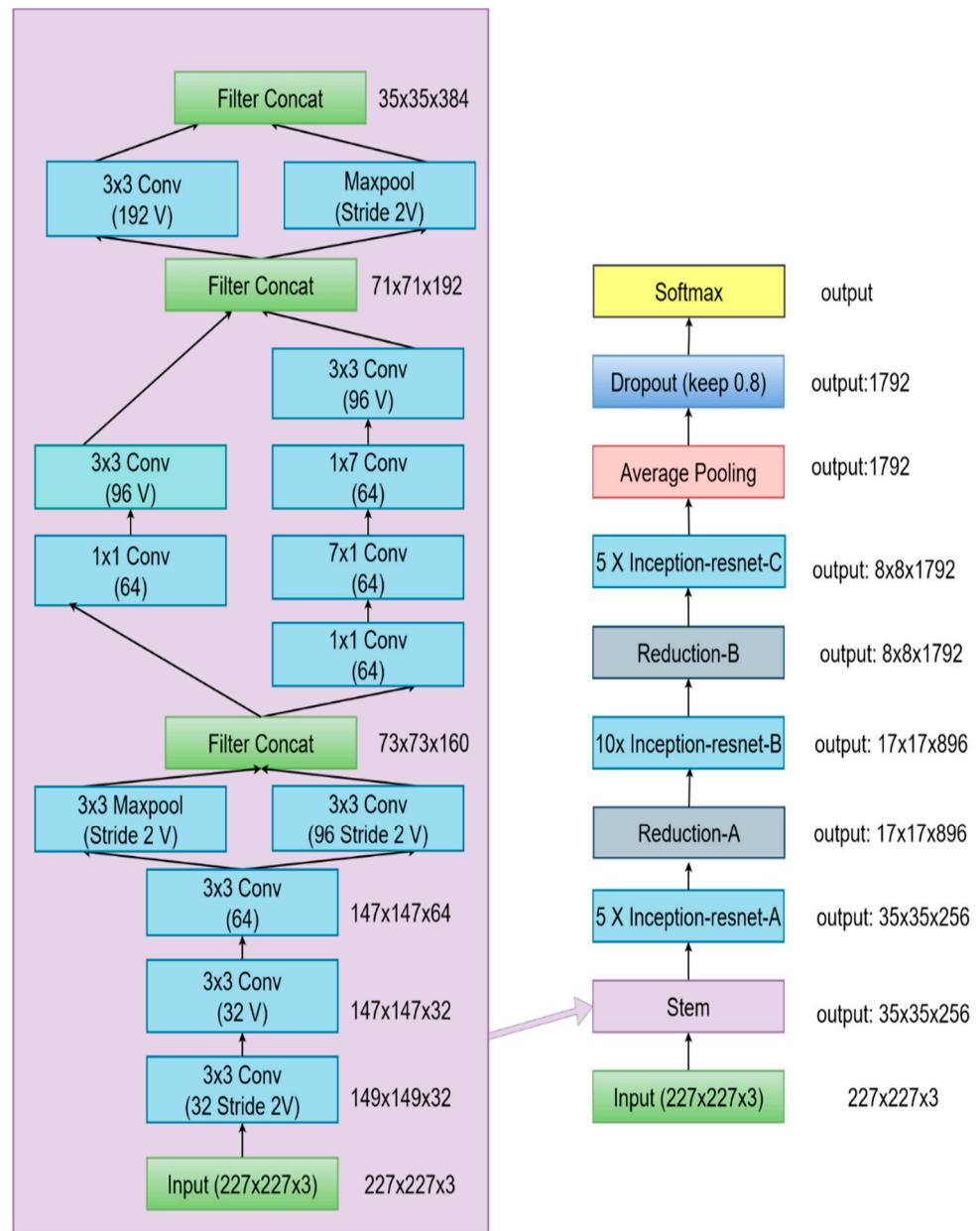


Figure 10. Inception-ResNet-v2 model architecture.

Let, (X) denote the input feature map, and (M_k) represent the (k) -th Inception module in the network. The output of the (k) -th block is computed using Equation (13). Where, (X_k) and (X_{k+1}) denote the input and output feature maps of the (k) -th, respectively. In the Inception module, multiple convolutional pathways with different filter sizes are employed to capture features at various scales. Let, (C_i) represent the output feature maps of the (i) -th convolutional pathway within the Inception module. The output $M_k(X_k)$ is obtained by concatenating the feature maps from all pathways and applying a bottleneck layer. Mathematically, this can be expressed as Equation (14). Where, (B) denotes the bottleneck layer. For multiclass classification, the final classification layer involves a fully connected layer with a softmax activation function (Equation (11)).

$$X_{k+1} = M_k(X_k) + X_k \tag{13}$$

$$M_k(X_k) = \text{Concat}(C_1, C_2, \dots, C_n) \cdot B \tag{14}$$

The training objective is to minimize the categorical cross-entropy loss, quantifying the dissimilarity between predicted and true class probabilities. For an instance with ground truth label (y) and predicted probabilities (p_i), the categorical cross-entropy loss is shown in Equation (12).

CenterNet

CenterNet, an advanced object detection framework, has been modified for leukemia classification. Figure 11 demonstrates the CenterNet operation. It focuses on predicting object centers and regressing bounding box coordinates to make detection easier [62,63]. Its novel architecture simplifies the object detection pipeline while achieving competitive accuracy. In the context of medical imaging, the ability to pinpoint and precisely locate relevant features within cell samples is critical for accurate classification [64]. The problem of multiclass leukemia classification entails predicting the presence of many leukemia subtypes. The detection head in CenterNet is in charge of predicting object centers and regressing bounding box coordinates, shown in Figure 11. By leveraging CenterNet, we aimed to enhance our model’s capacity to identify and distinguish specific regions of interest within leukemia cell images, contributing to more re-liable disease diagnosis.

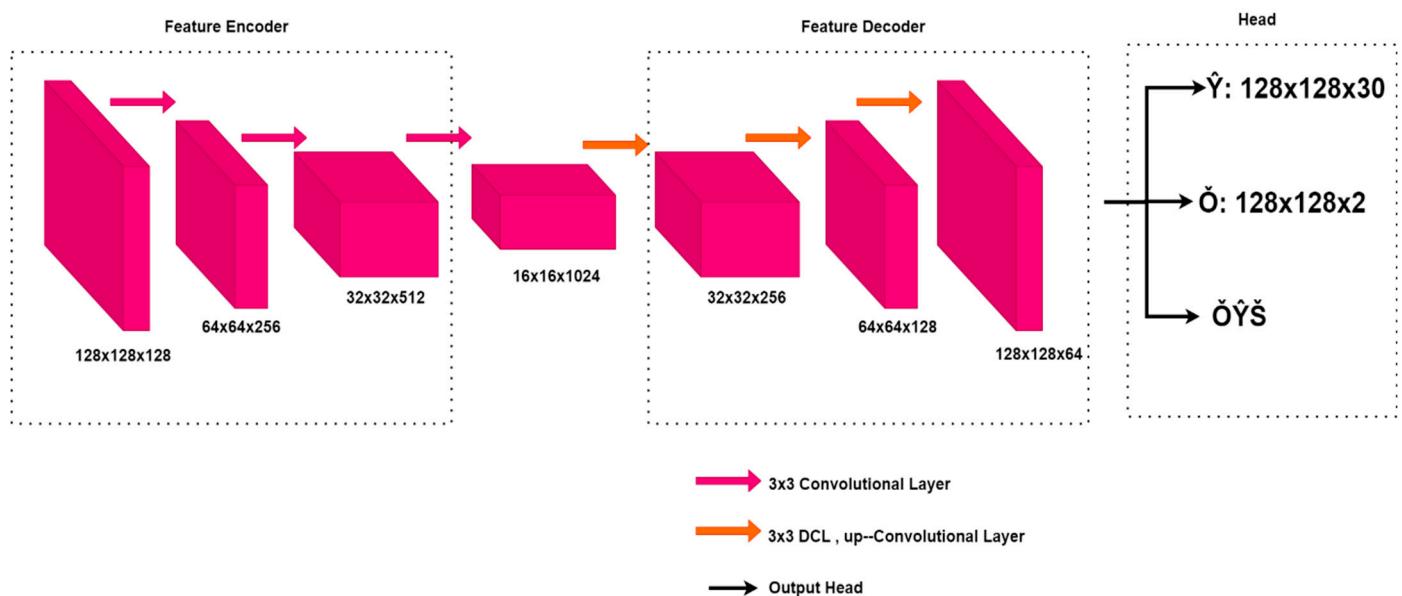


Figure 11. Architecture of CenterNet model.

The output feature map (O) is obtained by applying a 3×3 convolutional layer on (X_{k+1}). The bounding box width (W), and height (H) are then predicted using subsequent convolutional layers. Mathematically, the object center heatmap is generated using a sigmoid activation function, shown in Equation (15). The bounding box width (W) and height (H) are predicted through separate convolutional layers, and regression is performed using a linear activation function (Equations (16) and (17)). The final classification layer involves a fully connected layer with a softmax activation function (Equation (11)) for multiclass classification. The training objective is to minimize a combined loss function, which includes the focal loss (Equation (18)) for heatmap prediction and the smooth L1 loss for bounding box regression. The smooth L1 loss is shown in Equation (19). Where, (N) is the number of positive samples, (y_{ij}) is the binary indicator for the presence of class (j), (p_{ij}) is the predicted probability for class (j), (t_{ij}) and (\hat{t}_{ij}) are the ground truth and predicted bounding box coordinates, respectively.

$$H = \sigma(O) \tag{15}$$

$$W = \text{Linear}(O) \tag{16}$$

$$H = \text{Linear}(O) \quad (17)$$

$$L_{\text{heatmap}} = -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^C y_{ij} \cdot (1 - p_{ij})^\gamma \cdot \log(p_{ij}) \quad (18)$$

$$L_{\text{regression}} = \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^4 \text{SmoothL1}(t_{ij} - \hat{t}_{ij}), \quad (19)$$

3.6. Training Parameters

The training parameters for the TL models in our study are carefully chosen to ensure effective training and optimal model performance. We employed a 5-fold cross-validation strategy to robustly assess the models' generalization capabilities. The batch size is set to 256, determining the number of samples processed in each iteration during training. An epoch of 50 is defined, representing the number of times the entire dataset is passed forward and backward through the neural network. The optimizer chosen for the TL models is Adam, a popular optimization algorithm known for its efficiency and quick convergence. Adam adapts the learning rates for each parameter individually, combining the advantages of both the AdaGrad and RMSProp optimizers.

For multiclass classification, where we predict multiple leukemia subtypes, the categorical cross-entropy loss function is utilized. This loss function is suitable for scenarios with more than two classes and encourages the model to assign high probability to the correct class. In the case of binary classification, distinguishing between leukemia and normal samples, the Cross-Entropy loss function is employed. This loss function is well-suited for binary classification tasks, penalizing the model for predicting probabilities away from the true class. These hyperparameters are selected through a combination of empirical experimentation and considerations for the dataset's characteristics. The goal was to strike a balance between model complexity, training efficiency, and the ability to capture intricate patterns in the data.

3.7. Evaluation

The evaluation of our models is a meticulous process that encompasses a range of metrics to ensure a comprehensive assessment of their performance. These metrics provide a thorough understanding of our models' capabilities, allowing us to make well-informed decisions about their effectiveness in leukemia classification.

- Accuracy: This metric evaluates the overall correctness of the model predictions by calculating the ratio of correctly predicted instances to the total instances.
- Precision: It evaluates the correctness of positive predictions made by the model, by determining the proportion of true positive predictions to the overall predicted positives.
- Recall (Sensitivity): Evaluates the model's ability to accurately identify positive instances by comparing true positives to the total actual positives.
- The F1 score is a metric that combines precision and recall, providing a balanced evaluation. It is especially beneficial in situations where there is an uneven distribution of classes.
- The confusion matrix offers a comprehensive breakdown of the model's predictions, showcasing accurate positives, accurate negatives, incorrect positives, and incorrect negatives. It provides a comprehensive analysis of the model's advantages and disadvantages.
- Learning Curve: Learning curves illustrate the model's performance over epochs, indicating the rate at which the model is acquiring knowledge from the training data. They assist in identifying problems such as overfitting or underfitting.

We utilized Stratified K-fold cross-validation for each transfer learning model to ensure a thorough evaluation of the model's performance. This method ensures that the class distribution of the original dataset is preserved in each fold, which is essential for dealing

with the common issue of class imbalances. Dividing the dataset into $K = 5$ folds was done to ensure proportional representation of leukemia subtypes in each fold. For every iteration, $K-1$ folds were utilized for training, with the remaining fold used for validation. This process was repeated multiple times, enabling each fold to serve as both training and validation data. Through averaging performance metrics across folds, Stratified K -fold cross-validation minimized variance in performance estimates, offering more dependable insights into model effectiveness. Moreover, the method enabled hyperparameter optimization by assessing model performance with various parameter configurations.

4. Result Analysis

4.1. Binary Classification Results

The conducted experiment on binary classification yielded promising results across various ML and TL models, shown in Table 1. Notably, traditional ML models demonstrated strong performance with accuracies ranging from 85.36% to 92.26%. MLP achieves an accuracy of 92.26%, with high precision, recall, and F1 score values, indicating balanced performance in correctly classifying instances. RF follows closely with an accuracy of 91.71%, demonstrating robust performance across multiple metrics. SVM and KNN also exhibit strong performance, while SGD shows slightly lower accuracy but maintains reasonable precision and recall. On the other hand, the TL models, including Inception-ResNet, XceptionNet, AlexNet, RetinaNet, and CenterNet, consistently outperform the ML models. Inception-ResNet stands out as the top performer with an impressive accuracy of 96.89%, accompanied by high precision (96.43%), recall (96.61%), and F1 score (96.07%) values. The superiority of TL algorithms in achieving higher accuracy and overall performance underscores their efficacy in the intricate task of leukemia classification.

Table 1. Performance of binary classifiers.

Technique	Models	Accuracy	Precision	Recall	F1 Score
ML	MLP	92.26%	92.59%	92.38%	92.81%
	RF	91.71%	91.23%	91.66%	91.27%
	SVM	88.37%	88.11%	88.44%	88.79%
	KNN	87.51%	87.66%	87.51%	87.46%
	SGD	85.36%	85.59%	85.74%	85.65%
TL	Inception-ResNet	96.89%	96.43%	96.61%	96.07%
	XceptionNet	95.41%	95.24%	95.93%	95.21%
	AlexNet	94.01%	94.17%	94.21%	94.11%
	RetinaNet	94.55%	94.61%	94.31%	94.76%
	CenterNet	93.35%	93.68%	93.29%	93.45%

Figure 12a,b depicts the confusion matrix of highest performing ML and TL model for binary classification, respectively. The MLP model excels in precision for the "ALL" category but exhibits a lower precision for "Ham", indicating a higher rate of false positives. Additionally, the model demonstrates high recall for "Ham" but lower recall for "ALL", suggesting a higher rate of false negatives for "ALL". The resulting F1-scores indicate a good balance between precision and recall for "ALL" but reveal challenges in achieving a similar balance for the "Ham" class. On the other hand, highest performing TL model (Inception-ResNet) has demonstrated an impressive precision for both categories. It indicates a low rate of false positives, emphasizing the model's precision in identifying positive instances. Furthermore, the model exhibits high recall rates for both categories, with 96.9% for "ALL" and 96.3% for "Ham", indicating a low rate of false negatives. The F1-score, which signifies a balance between precision and recall, is also noteworthy, reaching 96.8% for "ALL" and 95.1% for "Ham". These findings underscore the robust performance of the TL in accurately classifying leukemia's existence while maintaining a balance between minimizing false positives and false negatives.

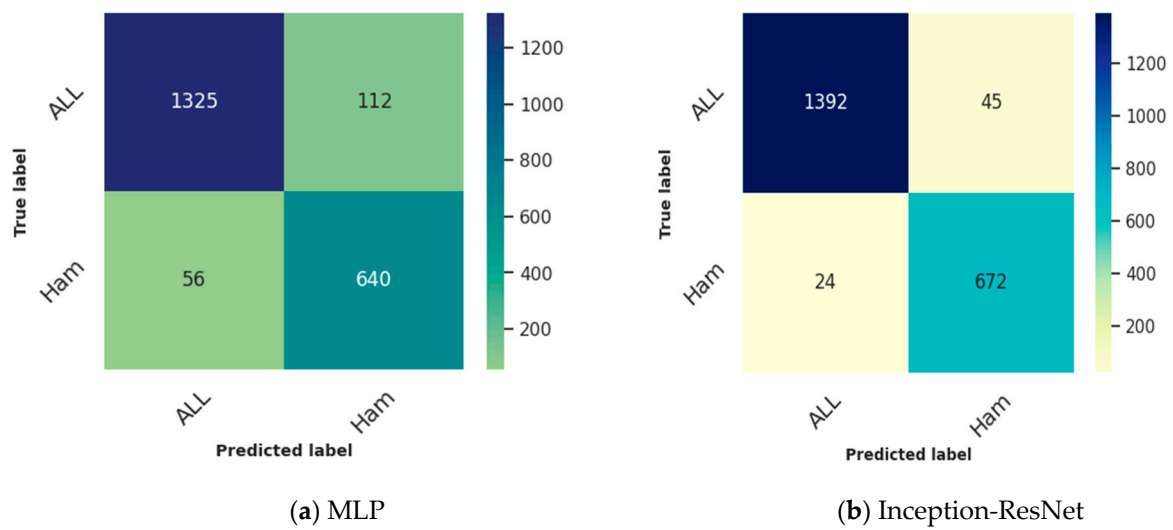


Figure 12. Confusion matrix of highest performing ML (a) and TL (b) model for binary classification.

In Figure 13, the learning curve of the top-performing TL model for binary classification is depicted. Across each fold, the model demonstrates positive learning trends, evident in the increasing accuracy and decreasing loss observed in both the training and validation sets. The high accuracy and low loss values underscore the model’s efficacy in the classification task. The initial fluctuations in both accuracy and loss curve observed in the first three folds of the learning curve may stem from factors such as random weight initialization, dataset complexity, and variations in data distribution. These factors can influence the model’s convergence rate and performance during early training epochs [65–67]. In contrast, the stability seen in the last two folds suggests that the model has converged to more optimal weight configurations and learning patterns over time. With continued training, the model becomes better equipped to generalize across diverse samples, resulting in smoother learning curves and consistent performance across folds. Furthermore, the minimal disparity between training and validation metrics for each fold suggests negligible generalization error, signifying the model’s ability to generalize well to unseen data. This alignment reinforces the model’s stability and implies balanced learning, mitigating concerns of overfitting or underfitting.

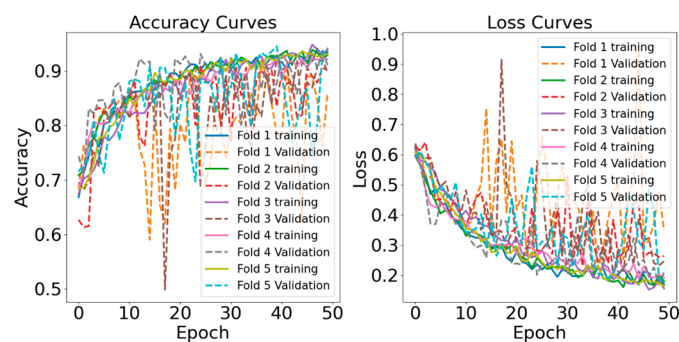


Figure 13. Learning curve of Inception-ResNet model for binary classification.

4.2. Multiclass Classification Results

Table 2 demonstrate notable performance variations among the models for multiclass classification. Among ML models, RF achieved an accuracy of 88.11%, showcasing a balanced precision, recall, and F1-score. Similarly, MLP, KNN, and SGD models exhibited competitive performance, with accuracies around 87–86%. SVM showed a slightly lower accuracy of 84.21% with balanced precision, recall, and F1-score. On the other hand, all the TL models outperformed traditional ML models. InceptionResNet demonstrated

the highest accuracy at 95.79%, emphasizing its efficiency in capturing intricate features. AlexNet and XceptionNet closely followed, achieving accuracy rates of 94.29% and 93.56%, respectively. CenterNet and RetinaNet also delivered respectable performance, though slightly lower than the top-performing TL models.

Table 2. Performance of multi-class classifiers.

Technique	Models	Accuracy	Precision	Recall	F1 Score
ML	RF	88.11%	88.34%	88.11%	88.23%
	MLP	87.78%	87.86%	87.36%	87.62%
	KNN	86.35%	86.35%	86.64%	86.91%
	SVM	84.21%	84.32%	84.59%	84.88%
	SGD	85.44%	85.87%	85.49%	85.94%
TL	InceptionResNet	95.79%	95.33%	95.41%	95.89%
	AlexNet	94.29%	94.49%	94.92%	94.29%
	XceptionNet	93.56%	93.32%	93.93%	93.66%
	CenterNet	93.98%	93.93%	93.91%	93.96%
	RetinaNet	91.91%	91.93%	91.76%	91.66%

Figure 14a,b depicts the confusion matrix of highest performing ML and TL model for multi-class classification, respectively. The error percentage of each class is the proportion of instances that the model misclassified for that class. It can be calculated by dividing the sum of the off-diagonal cells in the row or column of that class by the total number of instances for that class. In comparison to RF, which exhibits relatively higher error percentages ranging from 11.2% to 12% for various subtypes, Inception-ResNet demonstrates significantly improved accuracy. The misclassification percentages by RF on the 20% test data for ALL, AML, CML, CLL, and H are 12.03%, 11.92%, 11.98%, 11.95%, and 12.03%, respectively. In contrast, Inception-ResNet achieves significantly lower error percentages, such as 4.31% for ALL, 4.62% for AML, 4.58% for CML, 4.68% for CLL, and 4.20% for H. The use of transfer learning results in a further reduction in error percentages for each subtype to 7%, with reductions of 7.41% for ALL, 7.34% for AML, 7.3% for CML, 7.27% for CLL, and 7.83% for H. The significant reduction in error percentages achieved by the TL model across all leukemia subtypes indicates its robustness in differentiating between various classes. These affirm the effectiveness of the TL models, showcasing its potential as a reliable choice for medical imaging applications. Our findings contribute valuable insights for future research and applications in the field of early leukemia diagnostics, emphasizing the importance of advanced models for improved disease classification.

The learning curve depicted in Figure 15 demonstrates a graph of the training and validation accuracy and loss of an Inception-ResNet model. The learning curve in Figure 15 illustrates the training and validation accuracy and loss of an Inception-ResNet model. Across the five folds, the model consistently demonstrates high accuracy and low loss on both training and validation datasets, showcasing its effectiveness. The slight differences between training and validation curves suggest a small generalization error. Folds 2, 3, and 4 exhibit smooth and stable curves, indicating consistent learning with minimal fluctuations. However, in Folds 3 and 4, some variations appear in the validation curves, particularly in later epochs, indicating potential noise or randomness. Nevertheless, the overall trend shows an improvement in accuracy and a decrease in loss over time. Fold 5 illustrates convergence, as the model reaches a plateau in both accuracy and loss, with closely aligned curves implying low generalization error. The minimal fluctuation observed in the learning curve can be attributed to several factors contributing to stable learning behavior. Firstly, the multi-class dataset used for training have been well-preprocessed, reducing noise and variability that could lead to fluctuations in the learning process. Moreover, careful selection and tuning of optimization algorithms and hyperparameters contributed to smoother convergence by minimizing oscillations and erratic behavior during training [68].

Furthermore, the use of techniques like regularization and early stopping have helped prevent overfitting and stabilize the learning process.

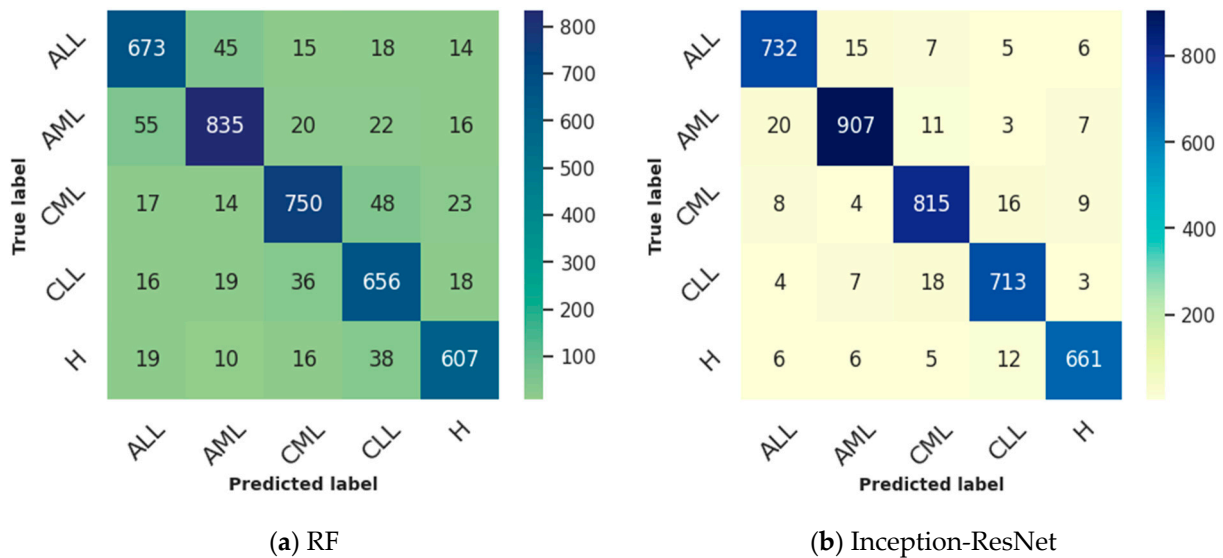


Figure 14. Confusion matrix of highest performing ML (a) and TL (b) model for multi-class classification.

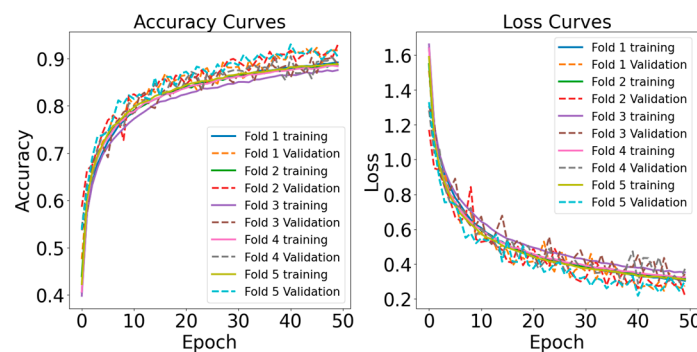


Figure 15. Learning curve of Inception-ResNet model for multi-class classification.

4.3. State-of-the-Art Comparison

Our research presents superior results compared to previous studies focusing on classifying leukemia with more than four categories, as shown in Table 3. Dealing with a greater number of classes naturally results in heightened complexity and difficulties. By utilizing the InceptionResNet model, we attained a remarkable accuracy of 95.79% across five classes. This exceeds the accuracies reported in previous studies, even when handling a larger number of leukemia categories. For example, in [29], a CNN model accurately classifies six leukemia categories with an 81.09% accuracy rate. On the other hand, our model achieves superior accuracy while managing a comparable number of classes. The VGG16 model, as mentioned in [34], achieved an accuracy of 94% across four classes, while another CNN model discussed in reference [31] reached an accuracy of 88.25% in distinguishing between five classes. Furthermore, the strong performance of our method highlights the effectiveness of the InceptionResNet architecture in capturing intricate patterns and features in leukemia cell images.

Table 3. Comparison of Leukemia Classification Models and Their Accuracy.

Reference	Model	Number of Classes	Accuracy
[25]	FSVM	3	96.92%
[30]	ResNet50	4	99.84%
[34]	VGG16	4	94%
[31]	CNN	5	88.25%
[29]	CNN	6	81.09%
Ours	InceptionResNet	5	95.79%

5. Discussion

Our study represents a major breakthrough in the field of medical imaging, particularly in the accurate diagnosis and classification of leukemia. The effective application of ML and TL models in accurately identifying and classifying leukemia subtypes shows great potential in improving pathological diagnosis. Our research highlights the capabilities of these models to optimize pathology workflows, offering fast and accurate leukemia classification that is essential for prompt and personalized treatment plans. By utilizing advanced image processing techniques and state-of-the-art models, the implementation of automated pathological diagnosis helps overcome the limitations of manual identification, such as subjectivity and time constraints. Our findings hold significant potential for improving the quality of care in clinical settings. The prospect of tailoring treatment strategies based on precise subtype identification could contribute to enhanced survival rates. Crucially, integrating these advanced models into pathology workflows has the potential to streamline the diagnostic process, enabling healthcare professionals to focus on intricate aspects of patient care. Our web application, offering rapid and accurate leukemia classification, serves as an invaluable tool for healthcare professionals, particularly oncologists and hematologists.

To seamlessly integrate our models for diagnosis and treatment planning into existing healthcare systems, several key steps must be taken. Firstly, compatibility with current standards should be ensured by adopting standardized data formats and communication protocols. Moreover, decision support systems leveraging the models' output can aid healthcare providers in making informed decisions about patient care. Training programs are also essential to equip healthcare professionals with the necessary skills to interpret their outputs accurately. Then, continuous quality assurance measures and monitoring are crucial to maintaining the reliability of the models in real-world clinical settings. Furthermore, adherence to regulatory standards and ethical guidelines is imperative to safeguard patient privacy. Through these measures, healthcare systems can effectively leverage our models to improve leukemia diagnosis and treatment planning, ultimately enhancing patient outcomes and optimizing clinical workflow efficiency.

The binary and multiclass classification results reveal compelling insights into model performance. In binary classification, the Inception-ResNet model stands out as the top performer with an accuracy of 96.89%, showcasing its superior ability to distinguish between two classes. Conversely, the lowest-performing model in binary classification is SGD, with an accuracy of 85.36%. In multiclass classification, Inception-ResNet also emerges as the highest performer with an accuracy of 95.79%, highlighting its efficacy in handling multiple leukemia classes. The lowest-performing model in this category is SVM, achieving an accuracy of 84.21%. Inception-ResNet's consistent superiority in both scenarios positions it as a promising candidate for impacting clinical decision support systems in various medical imaging applications.

The unique architecture of Inception-ResNet, which combines the best features of both ResNet and Inception, is responsible for its outstanding performance in our study. This combination resolves issues such as disappearing gradients and permits training very deep networks, enabling the model to capture complex features that are essential for precise categorization [69]. While residual connections reduce information loss during training,

the inception modules enable multi-level feature extraction. By using prior information, TL improves performance even further.

Although our study shows promising results, it is crucial to acknowledge a few limitations that need to be resolved. Initially, there may be difficulties in applying our results to various populations, as the effectiveness of the model can differ among different ethnicities and genetic backgrounds. It is crucial to conduct additional validation in various patient groups to confirm the relevance of our findings. Moreover, biases in the datasets used may affect the model's performance by over- or under-representing specific classes [70]. In addition, the applicability of our study findings may be restricted, particularly when extending results to various healthcare environments. In addition, although our models show high accuracy, their interpretability continues to be a challenge, which could impede their clinical adoption. Utilizing explainable AI methods like Grad-CAM can improve the interpretability of our models by visually displaying the key regions in input images that impact the predictions [71,72]. Finally, the implementation of AI models in clinical environments brings up ethical and regulatory concerns related to patient privacy, data security, and accountability. It is essential to follow strict ethical guidelines and regulatory standards to ensure the responsible deployment of our web application in healthcare. It is crucial to overcome these limitations to further the progress and ethical use of machine learning models in medical settings.

Future research directions and potential improvements to the proposed methodology are essential to guide subsequent studies and advance the field of cancer diagnosis and medical imaging. Firstly, exploring the integration of Vision Transformer (ViT) could enhance the performance and robustness of the proposed methodology. Investigating the applicability of ViT across different imaging modalities may also yield valuable insights into improving model generalization. Additionally, incorporating multimodal data fusion approaches, such as combining imaging data with clinical, genomic, or proteomic data, could provide a more comprehensive understanding of cancer pathophysiology and facilitate more accurate diagnosis and prognosis. Furthermore, developing interpretable and transparent AI models, such as incorporating explainable AI techniques like Grad-CAM, can enhance the trust and acceptance of AI-driven diagnostics in clinical practice. Addressing the computational efficiency of the proposed methodology is also crucial for its practical deployment in real-world healthcare settings. Finally, exploring the integration of emerging technologies such as federated learning and edge computing could enable decentralized model training and inference while ensuring data privacy and security. By focusing on these future research directions and potential improvements, subsequent studies can build upon the proposed methodology to advance the field of cancer diagnosis and medical imaging, ultimately improving patient outcomes and healthcare delivery.

6. Conclusions and Future Works

Our research showcases a notable advancement in tackling the difficulties related to the pathological diagnosis of leukemia by utilizing ML and TL models. The study offers a comprehensive analysis of different classifiers, including both traditional ML and cutting-edge TL models, for leukemia classification in both binary and multiclass scenarios. Notably, Inception-ResNet demonstrates outstanding performance in both scenarios, showcasing its accuracy and versatility. The significance of advanced models in revolutionizing leukemia classification cannot be overstated, as they offer the potential for more efficient and precise diagnostic processes. The findings highlight the significance of utilizing these technologies to address the limitations of manual identification, thereby improving the overall diagnostic precision in leukemia. While our methodology holds promise in cancer diagnosis and medical imaging, limitations exist. Generalization to diverse populations may be challenging due to ethnic and genetic variations, while dataset biases could affect model accuracy. Interpretability remains limited, hindering clinical adoption, and ethical considerations must be addressed. By incorporating additional clinical information, such as patient history, genetic data, and treatment responses, the

predictive power of our TL models can be significantly enhanced. Future endeavors should prioritize the seamless integration of diverse data sources to enhance diagnostic capabilities. Ensuring the validation of our developed models in various clinical settings, assessing their practical utility, and seamlessly integrating them into existing diagnostic workflows are crucial for their real-world applicability. Incorporating explainable AI techniques can improve the reliability of our models, allowing clinicians to understand and have confidence in the decision-making process. Our research opens up new possibilities for incorporating TL into medical diagnostics. In general, our work serves as a foundation for the advancement of medical imaging. We are committed to constantly improving our methods, gaining deeper understanding of clinical data, and striving for transparency in AI. We have made a notable breakthrough in precision diagnosis, potentially resulting in improved patient outcomes.

Author Contributions: R.H.—Conceptualization, Methodology, Writing—original draft, Writing—review & editing; A.A.S.—Writing—original draft, Formal analysis, Data Curation, Software, Validation; M.F.H.—Data Curation, Software, Visualization, Resources; F.I.—Data Curation, Software, Visualization, Resources; F.I.A.—Writing—review & editing, Resources, Validation; M.R.A.—Data Curation, Software, Visualization, Resources; S.K.—Investigation, Validation, Resources; A.R.—Investigation, Validation, Resources; M.J.H.—Writing—review & editing, Resources, Supervision, Funding acquisition, Validation. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study consist of the C-NMC dataset [36], obtainable from <https://www.kaggle.com/datasets/avk256/cnmc-leukemia>, accessed on 10 January 2024, and the Leukemia Dataset 0.2 [37], accessible via <https://www.kaggle.com/datasets/priyaadharshinivs062/leukemia-dataset-02>, accessed on 10 January 2024.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Wang, R.C.; Wang, Z. Precision Medicine: Disease Subtyping and Tailored Treatment. *Cancers* **2023**, *15*, 3837. [CrossRef]
2. Koliijn, P.M.; Langerak, A.W. Immune dysregulation as a leading principle for lymphoma development in diverse immunological backgrounds. *Immunol. Lett.* **2023**, *263*, 46–59. [CrossRef]
3. Lee, S.Y.; Miller, K.M.; Kim, J.J. Clinical and Mechanistic Implications of R-Loops in Human Leukemias. *Int. J. Mol. Sci.* **2023**, *24*, 5966. [CrossRef]
4. Klafke, N.; Bossert, J.; Kröger, B.; Neuberger, P.; Heyder, U.; Layer, M.; Winkler, M.; Idler, C.; Kaschdailewitsch, E.; Heine, R.; et al. Prevention and Treatment of Chemotherapy-Induced Peripheral Neuropathy (CIPN) with Non-Pharmacological Interventions: Clinical Recommendations from a Systematic Scoping Review and an Expert Consensus Process. *Med. Sci.* **2023**, *11*, 15. [CrossRef]
5. Al Khatib, A.O.; El-Tanani, M.; Al-Obaidi, H. Inhaled Medicines for Targeting Non-Small Cell Lung Cancer. *Pharmaceutics* **2023**, *15*, 2777. [CrossRef]
6. Yusufjanovich, E.U.; Irisbaevich, M.G.; Rafiqovich, Z.A.; Irsaliyevich, E.K. Evaluation of Effectiveness of Splenectomy in Chronic Leukemias. *World Bull. Public Health* **2023**, *19*, 79–83.
7. Senapati, J.; Jabbour, E.; Kantarjian, H.; Short, N.J. Pathogenesis and management of accelerated and blast phases of chronic myeloid leukemia. *Leukemia* **2022**, *37*, 5–17. [CrossRef]
8. Mustaqim, T.; Faticah, C.; Suciati, N. Deep Learning for the Detection of Acute Lymphoblastic Leukemia Subtypes on Microscopic Images: A Systematic Literature Review. *IEEE Access* **2023**, *11*, 16108–16127. [CrossRef]
9. Al-Bayati, A.M.S.; Al-Bayti, A.A.H.; Husain, V.I. A short review about chronic myeloid leukemia. *J. Life Bio Sci. Res.* **2023**, *4*, 15–19. [CrossRef]
10. Elhadary, M.; Elshoeibi, A.M.; Badr, A.; Elsayed, B.; Metwally, O.; Elshoeibi, A.M.; Mattar, M.; Alfarsi, K.; AlShammari, S.; Alshurafa, A.; et al. Revolutionizing chronic lymphocytic leukemia diagnosis: A deep dive into the diverse applications of machine learning. *Blood Rev.* **2023**, *62*, 101134. [CrossRef]
11. Rinaldi, I.; Winston, K. Chronic Myeloid Leukemia, from Pathophysiology to Treatment-Free Remission: A Narrative Literature Review. *J. Blood Med.* **2023**, *14*, 261–277. [CrossRef]

12. Wang, S.A.; Jorgensen, J.L.; Hu, S.; Jia, F.; Li, S.; Loghavi, S.; Ok, C.Y.; Thakral, B.; Xu, J.; Medeiros, L.J.; et al. Validation of a 12-color flow cytometry assay for acute myeloid leukemia minimal/measurable residual disease detection. *Cytom. B Clin. Cytom.* **2023**, *104*, 356–366. [CrossRef]
13. Timakova, A.; Ananov, V.; Fayzullin, A.; Makarov, V.; Ivanova, E.; Shekhter, A.; Timashev, P. Artificial Intelligence Assists in the Detection of Blood Vessels in Whole Slide Images: Practical Benefits for Oncological Pathology. *Biomolecules* **2023**, *13*, 1327. [CrossRef]
14. Tan, P.; Chen, X.; Zhang, H.; Wei, Q.; Luo, K. Artificial intelligence aids in development of nanomedicines for cancer management. *Semin. Cancer Biol.* **2023**, *89*, 61–75. [CrossRef]
15. Kassani, S.H.; Kassani, P.H.; Wesolowski, M.J.; Schneider, K.A.; Deters, R. A Hybrid Deep Learning Architecture for Leukemic B-lymphoblast Classification. In Proceedings of the ICTC 2019 10th International Conference on ICT Convergence: ICT Convergence Leading the Autonomous Future, Jeju, Republic of Korea, 16–18 October 2019; pp. 271–276. [CrossRef]
16. Jiwani, N.; Gupta, K.; Pau, G.; Alibakhshikenari, M. Pattern Recognition of Acute Lymphoblastic Leukemia (ALL) Using Computational Deep Learning. *IEEE Access* **2023**, *11*, 29541–29553. [CrossRef]
17. Simsek, E.; Badem, H.; Okumus, I.T. Leukemia Sub-Type Classification by Using Machine Learning Techniques on Gene Expression. In *Proceedings of Sixth International Congress on Information and Communication Technology; Lecture Notes in Networks and Systems*; Springer: Singapore, 2022; Volume 217, pp. 629–637. [CrossRef]
18. Jagadev, P.; Virani, H.G. Detection of leukemia and its types using image processing and machine learning. In Proceedings of the International Conference on Trends in Electronics and Informatics, ICEI 2017, Tirunelveli, India, 11–12 May 2017; pp. 522–526. [CrossRef]
19. Babaso, S.P.; Mishra, S.K.; Junnarkar, A. Leukemia Diagnosis Based on Machine Learning Algorithms. In Proceedings of the 2020 IEEE International Conference for Innovation in Technology, INOCON 2020, Bangluru, India, 6–8 November 2020. [CrossRef]
20. Rahmani, A.M.; Yousefpoor, E.; Yousefpoor, M.S.; Mehmood, Z.; Haider, A.; Hosseinzadeh, M.; Ali Naqvi, R. Machine Learning (ML) in Medicine: Review, Applications, and Challenges. *Mathematics* **2021**, *9*, 2970. [CrossRef]
21. Linardatos, P.; Papastefanopoulos, V.; Kotsiantis, S. Explainable AI: A Review of Machine Learning Interpretability Methods. *Entropy* **2021**, *23*, 18. [CrossRef]
22. Zhuang, F.; Qi, Z.; Duan, K.; Xi, D.; Zhu, Y.; Zhu, H.; Xiong, H.; He, Q. A Comprehensive Survey on Transfer Learning. *Proc. IEEE* **2021**, *109*, 43–76. [CrossRef]
23. Peirelinck, T.; Kazmi, H.; Mbuwir, B.V.; Hermans, C.; Spiessens, F.; Suykens, J.; Deconinck, G. Transfer learning in demand response: A review of algorithms for data-efficient modelling and control. *Energy AI* **2022**, *7*, 100126. [CrossRef]
24. Hamidah; Rustam, Z.; Utama, S.; Siswantining, T. Multiclass classification of acute lymphoblastic leukemia microarrays data using support vector machine algorithms. *J. Phys. Conf. Ser.* **2020**, *1490*, 012027. [CrossRef]
25. Fauzi, I.R.; Rustam, Z.; Wibowo, A. Multiclass classification of leukemia cancer data using Fuzzy Support Vector Machine (FSVM) with feature selection using Principal Component Analysis (PCA). *J. Phys. Conf. Ser.* **2021**, *1725*, 012012. [CrossRef]
26. Dasariraju, S.; Huo, M.; McCalla, S. Detection and Classification of Immature Leukocytes for Diagnosis of Acute Myeloid Leukemia Using Random Forest Algorithm. *Bioengineering* **2020**, *7*, 120. [CrossRef]
27. More, P.; Sugandhi, R. Automated and Enhanced Leucocyte Detection and Classification for Leukemia Detection Using Multi-Class SVM Classifier. *Eng. Proc.* **2023**, *37*, 36. [CrossRef]
28. Kashaf, A.; Khatibi, T.; Mehrvar, A. Treatment outcome classification of pediatric Acute Lymphoblastic Leukemia patients with clinical and medical data using machine learning: A case study at MAHAK hospital. *Inf. Med. Unlocked* **2020**, *20*, 100399. [CrossRef]
29. Schouten, J.P.E.; Matek, C.; Jacobs, L.F.P.; Buck, M.C.; Bošnački, D.; Marr, C. Tens of images can suffice to train neural networks for malignant leukocyte detection. *Sci. Rep.* **2021**, *11*, 7995. [CrossRef]
30. Rahman, W.; Faruque, M.G.G.; Roksana, K.; Sadi, A.H.M.S.; Rahman, M.M.; Azad, M.M. Multiclass blood cancer classification using deep CNN with optimized features. *Array* **2023**, *18*, 100292. [CrossRef]
31. Ahmed, N.; Yigit, A.; Isik, Z.; Alpkocak, A. Identification of Leukemia Subtypes from Microscopic Images Using Convolutional Neural Network. *Diagnostics* **2019**, *9*, 104. [CrossRef]
32. Saeed, U.; Kumar, K.; Khuhro, M.A.; Laghari, A.A.; Shaikh, A.A.; Rai, A. DeepLeukNet—A CNN based microscopy adaptation model for acute lymphoblastic leukemia classification. *Multimed. Tools Appl.* **2024**, *83*, 21019–21043. [CrossRef]
33. Loey, M.; Naman, M.; Zayed, H. Deep Transfer Learning in Diagnosing Leukemia in Blood Cells. *Computers* **2020**, *9*, 29. [CrossRef]
34. Ghongade, O.S.; Reddy, S.K.S.; Gavini, Y.C.; Tokala, S.; Enduri, M.K. Acute Lymphoblastic Leukemia Blood Cells Prediction Using Deep Learning & Transfer Learning Technique. *Indones. J. Electr. Eng. Inform.* **2023**, *11*, 778–790. [CrossRef]
35. Abir, W.H.; Uddin, M.F.; Khanam, F.R.; Tazin, T.; Khan, M.M.; Masud, M.; Aljahdali, S. Explainable AI in Diagnosing and Anticipating Leukemia Using Transfer Learning Method. *Comput. Intell. Neurosci.* **2022**, *2022*, 5140148. [CrossRef]
36. Gupta, R.; Gehlot, S.; Gupta, A. C-NMC: B-lineage acute lymphoblastic leukaemia: A blood cancer dataset. *Med. Eng. Phys.* **2022**, *103*, 103793. [CrossRef]
37. Dharshini, P.; Guru, S. LEUKEMIA DATASET 0.2. Available online: <https://www.kaggle.com/datasets/priyaadharshiniv062/leukemia-dataset-02> (accessed on 14 January 2024).

38. Perveen, S.; Alourani, A.; Shahbaz, M.; Ashraf, U.; Hamid, I. A framework for Early Detection of Acute Lymphoblastic Leukemia and its Subtypes from Peripheral Blood Smear Images Using Deep Ensemble Learning Technique. *IEEE Access* **2024**, *12*, 29252–29268. [[CrossRef](#)]
39. Sahu, A.; Das, P.K.; Meher, S.; Panda, R.; Abraham, A. An Efficient Deep Learning-Based Breast Cancer Detection Scheme with Small Datasets. In *Intelligent Systems Design and Applications. ISDA 2022; Lecture Notes in Networks and Systems*; Springer: Cham, Switzerland, 2023; Volume 717, pp. 39–48. [[CrossRef](#)]
40. Sahu, A.; Das, P.K.; Meher, S. An automatic sparse-based deep cascade framework with multilayer representation for detecting breast cancer. *Measurement* **2024**, *228*, 114375. [[CrossRef](#)]
41. Sahu, A.; Das, P.K.; Meher, S. An efficient deep learning scheme to detect breast cancer using mammogram and ultrasound breast images. *Biomed. Signal Process. Control* **2024**, *87*, 105377. [[CrossRef](#)]
42. Das, P.K.; Diya, V.A.; Meher, S.; Panda, R.; Abraham, A. A Systematic Review on Recent Advancements in Deep and Machine Learning Based Detection and Classification of Acute Lymphoblastic Leukemia. *IEEE Access* **2022**, *10*, 81741–81763. [[CrossRef](#)]
43. Haque, R.; Islam, N.; Tasneem, M.; Das, A.K. Multi-class sentiment classification on Bengali social media comments using machine learning. *Int. J. Cogn. Comput. Eng.* **2023**, *4*, 21–35. [[CrossRef](#)]
44. Haque, R.; Parameshachari, B.D.; Hasan, M.K.; Sakib, A.H.; Rahman, A.U.; Islam, M.B. Scientific Article Classification: Harnessing Hybrid Deep Learning Models for Knowledge Discovery. In Proceedings of the 2023 International Conference on Ambient Intelligence, Knowledge Informatics and Industrial Electronics (AIKIIIE), Ballari, India, 2–3 November 2023; pp. 1–7. [[CrossRef](#)]
45. Haque, R.; Laskar, S.H.; Khushbu, K.G.; Hasan, M.J.; Uddin, J. Data-Driven Solution to Identify Sentiments from Online Drug Reviews. *Computers* **2023**, *12*, 87. [[CrossRef](#)]
46. Haque, R.; Al Mamun, M.A.; Ratul, M.H.; Aziz, A.; Mitra, T. A Machine Learning Based Approach to Analyze Food Reviews from Bengali Text. In Proceedings of the 12th International Conference on Electrical and Computer Engineering, ICECE 2022, Dhaka, Bangladesh, 21–23 December 2022; pp. 80–83. [[CrossRef](#)]
47. Haque, R.; Islam, N.; Islam, M.; Ahsan, M.M. A Comparative Analysis on Suicidal Ideation Detection Using NLP, Machine, and Deep Learning. *Technologies* **2022**, *10*, 57. [[CrossRef](#)]
48. Islam, N.; Haque, R.; Pareek, P.K.; Islam, M.B.; Sajeeb, I.H.; Ratul, M.H. Deep Learning for Multi-Labeled Cyberbully Detection: Enhancing Online Safety. In Proceedings of the 2023 International Conference on Data Science and Network Security, ICDSNS 2023, Tiptur, India, 28–29 July 2023. [[CrossRef](#)]
49. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. *Commun. ACM* **2017**, *60*, 84–90. [[CrossRef](#)]
50. Shaheen, M.; Khan, R.; Biswal, R.R.; Ullah, M.; Khan, A.; Uddin, M.I.; Zareei, M.; Waheed, A. Acute Myeloid Leukemia (AML) Detection Using AlexNet Model. *Complexity* **2021**, *2021*, 6658192. [[CrossRef](#)]
51. Lin, T.Y.; Goyal, P.; Girshick, R.; He, K.; Dollar, P. Focal Loss for Dense Object Detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *42*, 318–327. [[CrossRef](#)]
52. Tan, G.; Guo, Z.; Xiao, Y. PA-RetinaNet: Path Augmented RetinaNet for Dense Object Detection. In *Artificial Intelligence and Lecture Notes in Bioinformatics; Lecture Notes in Computer Science*; Springer: Berlin/Heidelberg, Germany, 2019; Volume 11728, pp. 138–149. [[CrossRef](#)]
53. Haryono, A.; Jati, G.; Jatmiko, W. Oriented object detection in satellite images using convolutional neural network based on ResNeXt. *ETRI J.* **2023**. [[CrossRef](#)]
54. Zhou, M.; Wu, K.; Yu, L.; Xu, M.; Yang, J.; Shen, Q.; Liu, B.; Shi, L.; Wu, S.; Dong, B. Development and Evaluation of a Leukemia Diagnosis System Using Deep Learning in Real Clinical Scenarios. *Front. Pediatr.* **2021**, *9*, 693676. [[CrossRef](#)]
55. Wang, Y.; Wang, C.; Zhang, H.; Dong, Y.; Wei, S. Automatic Ship Detection Based on RetinaNet Using Multi-Resolution Gaofen-3 Imagery. *Remote Sens.* **2019**, *11*, 531. [[CrossRef](#)]
56. Shaheed, K.; Abbas, Q.; Hussain, A.; Qureshi, I. Optimized Xception Learning Model and XgBoost Classifier for Detection of Multiclass Chest Disease from X-ray Images. *Diagnostics* **2023**, *13*, 2583. [[CrossRef](#)]
57. Chand, S.; Vishwakarma, V.P. A novel Deep Learning Framework (DLF) for classification of Acute Lymphoblastic Leukemia. *Multimed Tools Appl.* **2022**, *81*, 37243–37262. [[CrossRef](#)]
58. Kodipalli, A.; Guha, S.; Dasar, S.; Ismail, T. An inception-ResNet deep learning approach to classify tumours in the ovary as benign and malignant. *Expert Syst.* **2022**, e13215. [[CrossRef](#)]
59. Wang, H.; Xu, S.; Fang, K.B.; Dai, Z.S.; Wei, G.Z.; Chen, L.F. Contrast-enhanced magnetic resonance image segmentation based on improved U-Net and Inception-ResNet in the diagnosis of spinal metastases. *J. Bone Oncol.* **2023**, *42*, 100498. [[CrossRef](#)]
60. Anilkumar, K.K.; Manoj, V.J.; Sagi, T.M. Automated detection of leukemia by pretrained deep neural networks and transfer learning: A comparison. *Med. Eng. Phys.* **2021**, *98*, 8–19. [[CrossRef](#)]
61. Szegedy, C.; Ioffe, S.; Vanhoucke, V.; Alemi, A.A. Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. *Proc. AAAI Conf. Artif. Intell.* **2017**, *31*, 4278–4284. [[CrossRef](#)]
62. Guo, Y.; Lu, X. ST-CenterNet: Small Target Detection Algorithm with Adaptive Data Enhancement. *Entropy* **2023**, *25*, 509. [[CrossRef](#)]
63. Xia, G.-S.; Bao, C.; Cao, J.; Hao, Q. AOGC: Anchor-Free Oriented Object Detection Based on Gaussian Centerness. *Remote Sens.* **2023**, *15*, 4690. [[CrossRef](#)]

64. Nazir, T.; Nawaz, M.; Rashid, J.; Mahum, R.; Masood, M.; Mehmood, A.; Ali, F.; Kim, J.; Kwon, H.Y.; Hussain, A.; et al. Detection of Diabetic Eye Disease from Retinal Images Using a Deep Learning Based CenterNet Model. *Sensors* **2021**, *21*, 5283. [[CrossRef](#)]
65. Karimi, D.; Warfield, S.K.; Gholipour, A. Transfer learning in medical image segmentation: New insights from analysis of the dynamics of model parameters and learned representations. *Artif. Intell. Med.* **2021**, *116*, 102078. [[CrossRef](#)]
66. Xiao, T.; Singh, M.; Mintun, E.; Darrell, T.; Dollár, P.; Girshick, R. Early Convolutions Help Transformers See Better. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 30392–30400.
67. Narkhede, M.V.; Bartakke, P.P.; Sutaone, M.S. A review on weight initialization strategies for neural networks. *Artif. Intell. Rev.* **2022**, *55*, 291–322. [[CrossRef](#)]
68. Cohen, O.; Malka, O.; Ringel, Z. Learning curves for overparametrized deep neural networks: A field theory perspective. *Phys. Rev. Res.* **2021**, *3*, 023034. [[CrossRef](#)]
69. Ferreira, C.A.; Melo, T.; Sousa, P.; Meyer, M.I.; Shakibapour, E.; Costa, P.; Campilho, A. Classification of Breast Cancer Histology Images Through Transfer Learning Using a Pre-trained Inception Resnet V2. In *Artificial Intelligence and Lecture Notes in Bioinformatics*; Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2018; Volume 10882, pp. 763–770. [[CrossRef](#)]
70. Claro, M.L.; de MSVeras, R.; Santana, A.M.; Vogado, L.H.S.; Junior, G.B.; de Medeiros, F.N.; Tavares, J.M.R. Assessing the impact of data augmentation and a combination of CNNs on leukemia classification. *Inf. Sci.* **2022**, *609*, 1010–1029. [[CrossRef](#)]
71. Xiao, M.; Zhang, L.; Shi, W.; Liu, J.; He, W.; Jiang, Z. A visualization method based on the Grad-CAM for medical image segmentation model. In Proceedings of the 2021 International Conference on Electronic Information Engineering and Computer Science, EIECS 2021, Changchun, China, 23–26 September 2021; pp. 242–247. [[CrossRef](#)]
72. Zhang, Y.; Hong, D.; McClement, D.; Oladosu, O.; Pridham, G.; Slaney, G. Grad-CAM helps interpret the deep learning models trained to classify multiple sclerosis types using clinical brain magnetic resonance imaging. *J. Neurosci. Methods* **2021**, *353*, 109098. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.