



(RESEARCH ARTICLE)



Privacy-preserving multi-class classification of acute lymphoblastic leukemia subtypes using federated learning

Md Ashraful Alam ¹, Amir Sohel ², Anik Biswas ³, Shaid Bin Md Sifat ⁴, Ashrafur Rahman Nabil ², Niladry Chowdhury ⁵, Asma Hafiz ⁶ and Md Aslamuzzaman Bappy ^{4,*}

¹ Department of Computer Science, Colorado State University, Fort Collins, USA.

² Information Technology Project Management, St. Francis College, New York, USA.

³ Information Studies, Trine University, Indiana, USA.

⁴ Department of Computer Science and Engineering, Daffodil International University, Dhaka, Bangladesh.

⁵ Business Analytics, Trine University, Indiana, USA.

⁶ Department of Biology, National University - Patiya Government College, Chittagong, Bangladesh.

World Journal of Advanced Research and Reviews, 2024, 24(03), 2795-2806

Publication history: Received on 16 November 2024; revised on 28 December 2024; accepted on 30 December 2024

Article DOI: <https://doi.org/10.30574/wjarr.2024.24.3.3971>

Abstract

Early and accurate diagnosis for a highly aggressive hematological malignancy: Acute Lymphoblastic Leukemia. This is where automated, privacy-preserving diagnostic solutions can not only ease the burden of current diagnostic approaches but also avoid invasive, time-intensive, and prone to error. In this study, we present a Federated Learning framework for the Multi-Class classification of Acute Lymphoblastic Leukemia subtypes based on Peripheral Blood Smear images. To deal with class imbalance, data augmentation techniques were applied, and then pre-trained convolution neural networks such as InceptionV3, DenseNet121, and Xception were fine-tuned to extract features. Of these, InceptionV3 performed the best with an accuracy of 95.49% in the Federated Learning framework guaranteeing the privacy of patient data through differential privacy mechanisms. Through comparative analysis, it was confirmed that in using the Federated Learning approach, the high diagnostic accuracy and robust generalization against different datasets were preserved, while outperforming centralized learning. By proposing a scalable, privacy-compliant solution for all diagnoses, Acute Lymphoblastic Leukemia diagnoses may be transformed into the new practice of hematological oncology.

Keywords: Acute Lymphoblastic Leukemia; Peripheral Blood Smear; Federated Learning; Transfer Learning; Data Privacy; Medical Image Classification

1. Introduction

Acute lymphoblastic leukemia (ALL) is a malignant abnormality in which small lymphoblasts are produced in excess quantities in the bone marrow and blood. Between 1990 and 2017, globally, the incidence of ALL rose from about 49,070 cases to 64,190 cases, with males being more often affected than females [1]. Also, ALL is well below 4% of all blood cancers in the United States, estimating 6,660 new cases in 2022 in the US [2]. While treatment has improved, the five-year relative survival for ALL is still about 71%, indicating the imperative need to find early and precise methods of diagnosis [3].

Flow cytometry and bone marrow biopsy are traditional diagnostic procedures for ALL, but they are invasive, costly and time-consuming [4]. As an alternative to PBS analysis, initial screening can be performed by Peripheral Blood Smear (PBS) [5]. However, clinical presentation is non-specific for ALL, and laboratory expert's manual inspection of PBS

* Corresponding author: Md Aslamuzzaman Bappy

images is also prone to diagnostic error. Thus, there is a great need for automated, reliable, and efficient diagnostic tools that can assist clinicians in confirming ALL subtypes.

Over the last few decades, Machine Learning (ML) and Deep Learning (DL) techniques have become increasingly popular tools for medical image analysis and help detect and classify multiple diseases in an automated manner [6]. In particular, CNN-based DL models have been widely used for ALL's image classification into benign and malignant categories and features to distinguish among different ALL types. However, these models perform very well but rely heavily on large, varied datasets typically located at multiple medical institutions. However, strict privacy regulations and ethical concerns prevent such sharing, limiting centralized model training with such data.

To address this challenge, Federated Learning (FL) is an innovative solution that allows collaborative model training across multiple institutions without sharing sensitive patient data [7]. In an FL framework, local models are trained on-site using institutional data, and only the model parameters are shared with a central server to create a global model. This decentralized approach ensures data privacy and security while leveraging the collective knowledge from distributed datasets. FL has already demonstrated its effectiveness in various medical imaging tasks, providing a privacy-preserving framework for collaborative model development.

In this study, we propose a privacy-preserving, federated deep learning framework for the multi-class classification of ALL subtypes using PBS images. To enhance the performance and generalization of our models, we first apply data augmentation techniques, which help to address the problem of data imbalance and improve feature extraction. Additionally, we evaluate the performance of multiple transfer learning models, including pre-trained architectures such as ResNet, EfficientNet, and DenseNet. After identifying the best-performing transfer learning model, we implement it within the federated learning framework to enable decentralized, privacy-preserving training across multiple institutions. Moreover, our methodology integrates advanced privacy-preserving mechanisms, such as differential privacy, to ensure compliance with data protection regulations during the model training process. The proposed framework facilitates accurately classifying benign and malignant ALL subtypes, including Early Pre-B, Pre-B, and Pro-B ALL. It ensures patient data remains secure throughout the process. Finally, we evaluate the federated model's performance against traditional centralized learning approaches to demonstrate its superiority in preserving data privacy while achieving high diagnostic accuracy. By combining federated learning with state-of-the-art transfer learning models and data augmentation techniques, this research aims to develop a robust and scalable diagnostic tool for ALL. This tool can potentially enhance the early detection and classification of ALL subtypes significantly, ultimately improving patient outcomes and advancing the field of hematologic oncology.

2. Related Works

ALL classification and diagnosis have attempted to gain the attention of machine learning and privacy-preserving methods. This section looks at key contributions to this domain, with a focus on specific methodologies and links to privacy-preserving frameworks, including FL.

In the FL framework, Das and Kuppan [8] suggested a fractional mayfly optimization (FMO)-DenseNet approach for leukemia classification. While their model yielded admirable accuracy (94.3%) yet again, it was limited to binary classification, and practically depended on FMO to train. In contrast to our study, the scope was expanded to consider the multi-class classification of ALL subtypes, which is a more challenging diagnostic problem. In addition, our framework employs differential privacy mechanisms to also provide robust data security which their research overlooks. Similarly, Ghongade et al. [9] used deep learning and transfer learning approaches with pre-trained CNNs for binary classification all achieving high accuracy. Yet they did not apply their study to important problems like class imbalance or data privacy. Instead, our method can address class imbalance effectively via data augmentation techniques that are sophisticated enough to guarantee patient data privacy through differential privacy mechanisms and make it more suitable for real-world clinical settings.

In their study, Ullah, et al [10] also introduced an Efficient Channel Attention (ECA) module within a VGG16-based model for ALL classification with 91.1% accuracy. To overcome the scalability and privacy constraints in distributed environments, their attention-based strategy served for feature extraction but was left unresolved. However, any privacy deserved by this federated learning framework is kept, as it proves to obtain higher accuracy (95.49% for the example setting with InceptionV3) and robust generalization across diverse datasets thus rendering it more versatile and secure. On the other hand, Haque et al. [11] employed transfer learning models (like Inception-ResNet) with image processing algorithms and reached F1 scores of 96.07% for binary classification and 95.89% for multiclass classification. Their centralized approach worked wonderfully until they encountered the privacy risk of handling

sensitive medical data. On the other hand, our federated learning model provides a decentralized, privacy-preserving solution, with no loss of diagnostic accuracy, which is better for clinical applications.

Additionally, Saeed et al. [12] used Multi-Attention EfficientNet architectures with 99.73% accuracy rates (EfficientNetV2S). Yet, high reliance on centralized learning exposes user data to privacy vulnerabilities and is not scalable. Additionally, they limit applicability in real-world multi-class diagnostic settings with binary classification. On the other hand, our federated learning framework can work for multi-class classification and protect the data's security, which curbs these important limitations. Moreover, Ansari et al. [13] built a customized CNN for ALL and AML classification with high (99%) accuracy out of a unique dataset and Tversky loss. However, their work focused on the expansion of the dataset and accuracy but was limited to centralized learning. In contrast, our federated learning framework exploits decentralized training by maintaining privacy compliance and robust generalization across heterogeneous datasets.

Pałczyński et al. [14] presented a hybrid AI system for leukemia classification based on MobileNetV2, machine learning, reaching 97.4% accuracy. While their lightweight models are suitable for IoT settings, their system offers limited privacy guarantees for the sensitive clinical applications we study. Compared to our federated learning approach, our approach not only maintains competitive diagnostic accuracy but also provides differential privacy mechanisms to create a secure and scalable solution. We conclude that while current studies in leukemia classification have improved remarkably, they often miss important problem areas such as privacy-preserving, multi-class scalability, and robustness to intra-class variance. To close these gaps, we propose a framework that integrates pre-trained CNNs like InceptionV3, DenseNet121, and Xception with differential privacy mechanisms and provides a superior privacy-compliant and scalable solution for ALL subtype diagnoses.

3. Methodology

This research employs a comprehensive and structured approach to classify ALL subtypes using a privacy-preserving FL framework. Figure 1 illustrates the overall methodology taking us step by step from dataset preparation to Federated Learning implementation. The methodology is divided into five major stages: Dataset preparation, Data augmentation, Transfer learning with multiple pre-trained models, Selection of the best-performing transfer learning model, and Integration of the selected model into the FL framework. Preparation of the PBS image dataset annotated and divided into benign and malignant classes and subtypes for the malignant classes is the first step of the process. To improve model robustness, we augment the input and label data, such as rotation, flipping, and scaling to create a richer training set. We then apply multiple pre-trained transfer learning models to find the architecture that provides the best test accuracy to classify ALL subtypes. After selecting the best model, the FL framework is implemented, such that distributed datasets collaboratively train the model, whilst maintaining data privacy. To further safeguard sensitive information during the model training process, our work also integrates differential privacy mechanisms.

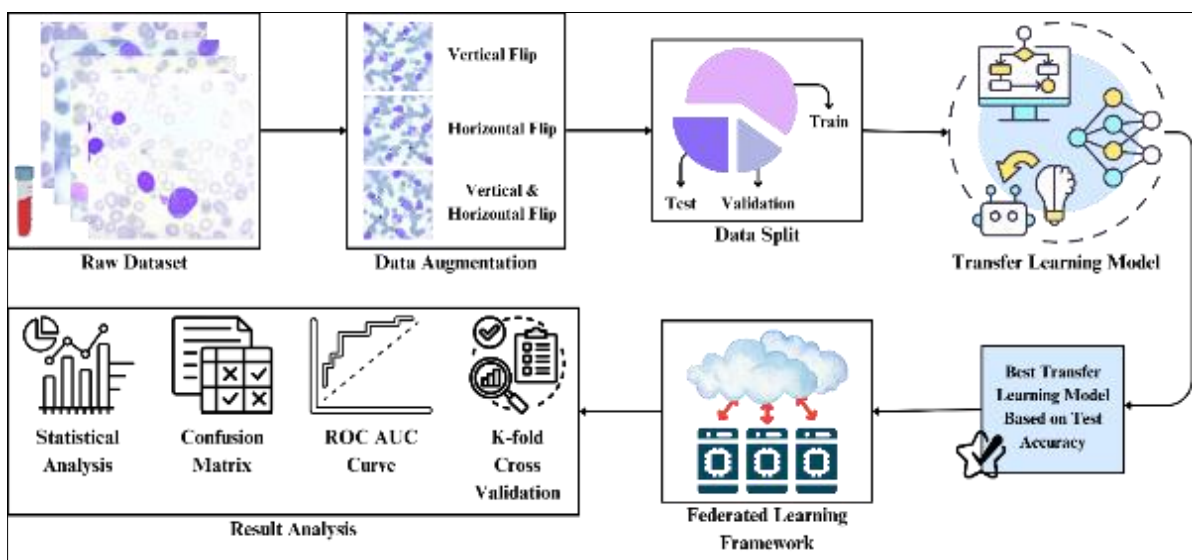


Figure 1 Methodology flowchart of this work.

3.1. Dataset

The dataset used in this study is a comprehensive collection of PBS images, meticulously prepared for ALL detection. This dataset [15,16] was curated at the bone marrow laboratory of Taleqani Hospital, Tehran, Iran, and comprises 3,242 PBS images collected from 89 patients suspected of ALL. The images were captured using a Zeiss camera mounted on a microscope with 100x magnification and saved in JPG format. Each sample was carefully stained and prepared by experienced laboratory staff to ensure high-quality imaging. The dataset is divided into two primary classes: benign and malignant. The benign class consists of normal hematogenous cells, while the malignant class represents ALL-positive cases further subdivided into three subtypes: Early Pre-B, Pre-B, and Pro-B ALL. The definitive classification of these PBS images was performed by a hematologist using a flow cytometry tool, ensuring precise categorization of the classes and subtypes. While the dataset offers a diverse representation of PBS images, capturing variations in morphology and staining patterns, it suffers from an inherent class imbalance, particularly in the distribution of certain ALL subtypes. Figure 2 showcases sample PBS images from this dataset, highlighting the variations in morphology and staining quality across the samples.

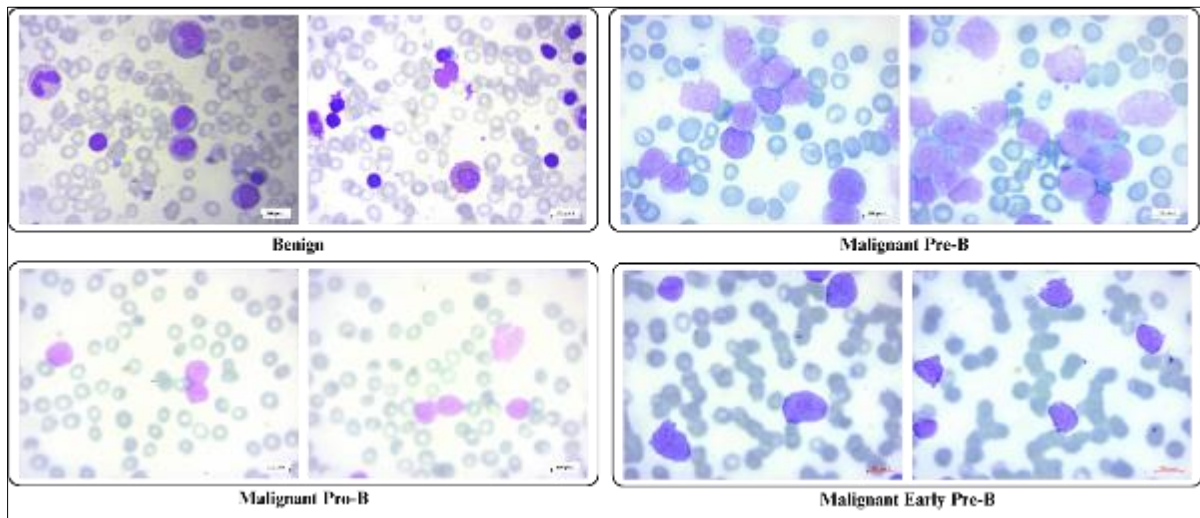


Figure 2 Sample Peripheral Blood Smear images from the dataset.

3.2. Data Augmentation

Some data augmentation techniques were applied to address the class imbalance in the dataset to have a better generalization capability of the model. In particular, geometric augmentation methods [17] were used to artificially expand the dataset, creating additional variations from existing images. The techniques used in this problem were: vertical flip, horizontal flip, and both vertical and horizontal flip together. All images, including the original and augmented, were resized to a consistent resolution of 224x224 pixels. Change in the spatial orientation for images effectively augmented the diversity of the training dataset but mimicked real-world variation in the image acquisition conditions. The number of images in the dataset was greatly increased through these geometric augmentations. Thus, for every original image, three additional augmented versions were created equaling 12,968 images (3,242 original images \times 4 variations with original). The class imbalance has been alleviated by this expanded dataset and the model is further robust because new additional samples were also added for the training. Figure 3 illustrates examples of these geometric augmentations, showcasing the variations generated by the applied transformations.

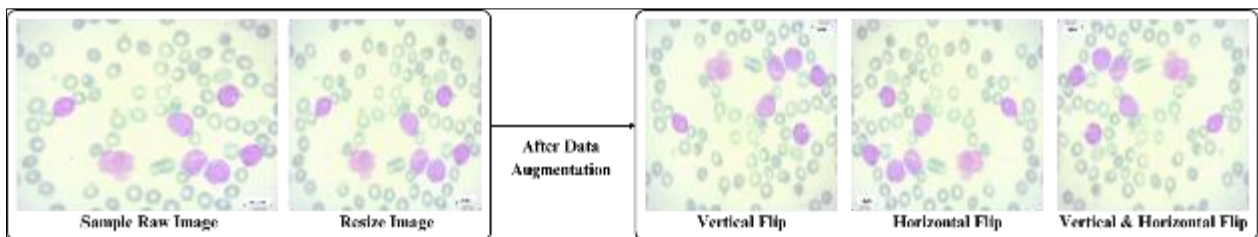


Figure 3 Example of geometric data augmentation techniques applied to PBS images.

3.3. Data Split

To facilitate an effective training and evaluation process, the augmented dataset, consisting of a total of 12,968 images, was systematically divided into three subsets: Test, validation and training set sizes of 70%, 20%, and 10%, respectively. This split will guarantee a robust framework of model training, yet retain enough samples for testing model performance on unseen data as well as tuning in validation. In particular, the training set contains 9,078 images, allowing the model to be trained and generalized sufficiently well. The 2,594 image images in the testing set are used to allow reliable evaluation of the model's performance. Third, a 1,296 image validation set is finally supplied to enable model tuning and prevent overfitting during training. It provides a balanced split that helps optimize the learning process and provides a comprehensive scale for measuring the perfection and generalization of the model.

3.4. Transfer Learning Model

Utilizing transfer learning, a well-known deep learning technique, we train to obtain an accurate classification of ALL subtypes. The transfer learning is used and taken advantage of to employ pre-trained convoluted neural network (CNN) models that have been trained on such large-scale datasets and then train them for domain-specific tasks such as the classification of PBS images. By exploiting features learned by general image datasets, this approach achieves performance improvements while substantially reducing the training time. Moreover, the use of transfer learning is particularly advantageous when there is insufficient data, which is ideal given the limited data for this study. We explored the performance of 10 different state-of-the-art pre-trained CNN models widely popular in image recognition and classification tasks. The first model is VGG16 [18], a deep CNN with 16 layers which is a simple model with excellent feature extraction power. After that, we used VGG19 [19], a deeper (19 layers) version, which improved performance with its deeper architecture. These two models have continued to be proven accurate in visual recognition tasks. We further used ResNet50 [20], a 50-layer residual network that introduces skip connections to mitigate vanishing gradient problems, to train deeper architectures more efficiently. This is complemented by looking at more complex features with a deeper variant of ResNet101 [21] with 101 layers. The performance of these ResNet models is also robust across multiple applications.

Additionally, we added InceptionV3 [22], which supports factorized convolutions, into our models for effective accuracy and reduced computation. Secondly, we explore this by also attempting Xception [23] (knowledgeably an adaptation of the Inception, but employing depth-wise separable convolutions to learn both performance and parameter improvements. To further improve the diversity in our approach we used DenseNet121 [24], a densely connected network, which encourages feature reuse by connecting each layer to all other layers. The architecture here is to reduce the number of parameters but still achieve high accuracy. It was also explored that EfficientNetB0 [25] due to its unusual scaling approach, that is, it balances width, depth and resolution, will optimize both accuracy and computational efficiency. We also included MobileNetV2 and NASNetMobile [26], two models intended for mobile and constrained resources. Depth-wise separable convolutions on MobileNetV2 have produced lightweight yet effective feature extraction, and NASNetMobile is a neural architecture search model with competitive accuracy at reduced computational costs.

3.5. Federated Learning Framework

Federated Learning (FL) is an innovative paradigm that enables decentralized model training while ensuring data privacy [27]. Unlike existing cloud data pipelines, FL doesn't move sensitive data to a central server; instead, local datasets stay on their respective client devices, ensuring privacy and compliance with data regulations [28]. Model updates are enabled by the FL framework, which aggregates the model parameters trained locally from several clients into a strong global model, making use of diverse data sources without performing computations on the raw data itself. The federated framework is iterative. Finally, the global model is initialized and distributed by the central server to the local clients. Each client then trains the model on its local data using some proprietary data analysis. The trained model parameters are subsequently sent back to the central server, which aggregates them to update the global model. This process continues until the global model converges.

For our implementation, we used this best transfer learning model from the centralized evaluation phase. This model was deployed to multiple clients (institutions or hospitals) where the model was trained locally using their datasets from Peripheral Blood Smear (PBS) images. The locally trained models were sent back to a central server and their weights were aggregated with the weights of the current global model. Through this framework, we make sure the model learns collaboratively across institutions while keeping data private. Figure 4 provides the image for the FL framework showing how the clients communicate with the central server. The distributed global model is denoted as downward arrows whereas training parameters are moved from the clients to the server per upward arrows. This is a

complete analogue of the process of cooperating training in a decentralized environment with a privacy preservation background.

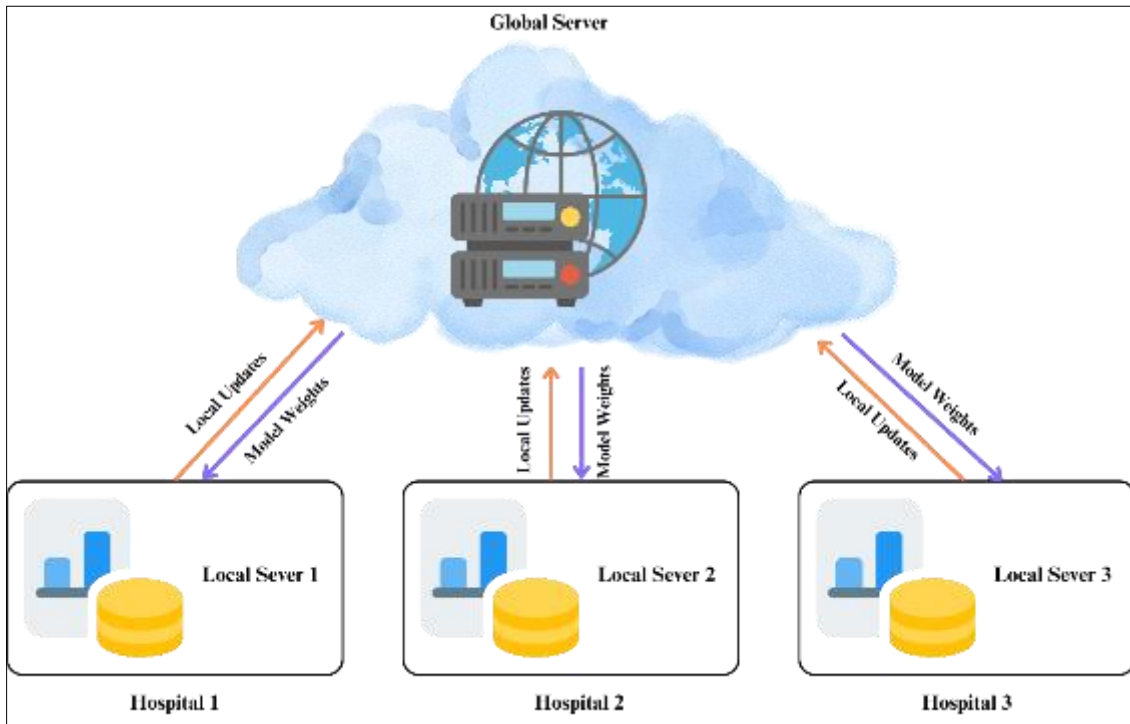


Figure 4 Federated Learning framework for ALL subtype classification

4. Experimental Analysis and Discussion

Computational resources including a T4 GPU were provided by Google Colab to conduct the experiments. With 12.7GB of RAM and 15GB of VRAM on hand, we had plenty of capacity for model processing and training within this environment. We implement and evaluate all models using Python-based libraries, such as TensorFlow and PyTorch. To evaluate the performance of each model, we employed four standard metrics: They are Accuracy [29], Precision [30], Recall [31] and F1 score [32]. Finally, measures of the models' ability to categorize benign and malignant PBS images under class imbalance were fully assessed by these metrics.

4.1. Result Analysis

Table 1 Performance of Transfer Learning Models

Models	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
VGG16	85.42	86.1	84.8	85.45
VGG19	86.31	87	85.9	86.45
ResNet50	88.23	88.8	87.5	88.15
ResNet101	89.4	90.1	88.9	89.5
DenseNet121	91.23	91.8	90.9	91.35
NASNetMobile	89.85	90.4	89.1	89.75
InceptionV3	92.42	92.9	91.8	92.35
Xception	91.8	92.4	91.1	91.75
EfficientNetB0	87.52	88.1	86.5	87.3
MobileNetV2	85.93	86.5	85.1	85.8

Table 1 presents the performance results of ten transfer learning models applied to the ALL dataset. Each model was fine-tuned on the training set, validated, and tested to determine its effectiveness in classifying ALL subtypes.

The results in Table 1 highlight the performance of various transfer learning models on the test dataset. Among these, InceptionV3 emerged as the top-performing model, achieving an accuracy of 92.42%, precision of 92.90%, recall of 91.80%, and an F1-Score of 92.35%. This model demonstrated superior classification capability compared to others, likely due to its advanced architecture, which effectively captures both spatial and channel information. On the other hand, models like VGG16 and MobileNetV2 showed relatively lower performance, with accuracies of 85.42% and 85.93%, respectively. While these models are known for their lightweight architectures, they may lack the depth required to process complex medical images effectively. Models such as ResNet50, DenseNet121, and Xception also delivered a strong performance, with accuracies exceeding 88%, indicating their robustness in handling challenging classification tasks.

The Federated Learning framework was implemented using the top-performing models identified in the experimental analysis, specifically those achieving over 90% accuracy in standalone experiments. The framework consisted of 5 local training epochs on client devices and 10 communication rounds between the clients and the central server, resulting in a total of 50 optimization epochs for each model. Table 2 highlights the results of the Federated Learning experiments using the three top-performing transfer learning models.

Table 2 Federated Learning Performance of Top Transfer Learning Models

Models	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
DenseNet121	93.21	93.8	92.9	93.35
InceptionV3	95.49	96.1	94.8	95.45
Xception	94.15	94.7	93.8	94.25

Among these models, InceptionV3 demonstrated the highest accuracy of 95.49%, along with a precision of 96.10%, recall of 94.80%, and an F1-Score of 95.45%. This result underscores the robustness of InceptionV3's architecture in adapting to the decentralized Federated Learning environment, maintaining its superior classification ability even in a distributed setting. The Xception model achieved commendable results with an accuracy of 94.15%, closely followed by DenseNet121 at 93.21%. Both models showcased their potential to generalize well across diverse data distributions while effectively leveraging Federated Learning principles. These findings further validate the efficacy of Federated Learning in utilizing powerful transfer learning models, particularly InceptionV3, to achieve highly accurate and reliable medical image classification in a decentralized setup.

4.2. Performance Analysis

The model was thoroughly evaluated using the confusion matrix, ROC-AUC curve, and k-fold cross-validation to ensure that classification capabilities and robustness are achieved. The confusion matrix illustrated in Figure 5 provides an overview of the model's classification performance across four categories: The model correctly classifies these categories (Benign, Malignant Early Pre-B, Malignant Pre-B, and Malignant Pro-B.) with high values on the diagonal, such as 389 for "Benign" and 758 for "Malignant Early Pre-B." Areas of error such as the 19 cases when "Malignant Pre-B" was mistakenly classed as "Malignant Pro-B" are highlighted, providing information on the model's weaknesses. Finally, the confusion matrix demonstrates the model's strong performance confirming the majority of predictions were made based on the true labels.

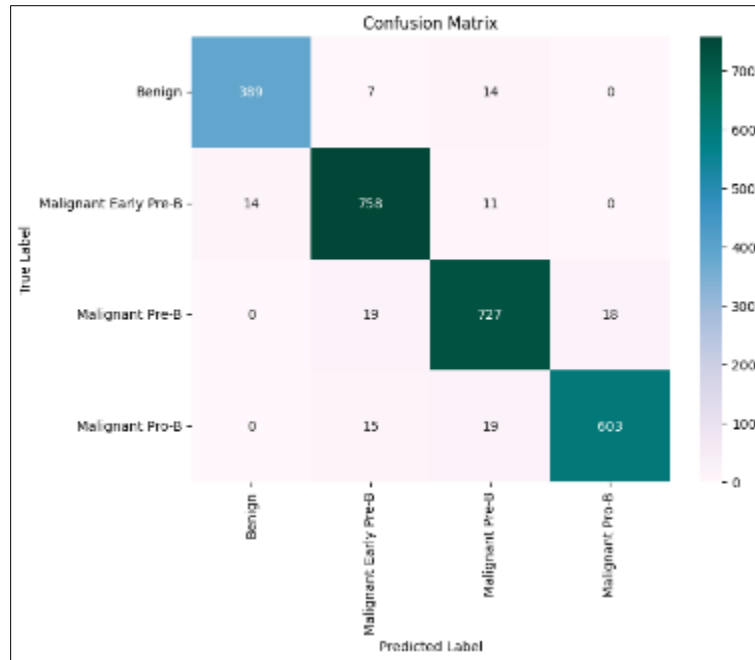


Figure 5 Federated Learning framework for ALL subtype classification.

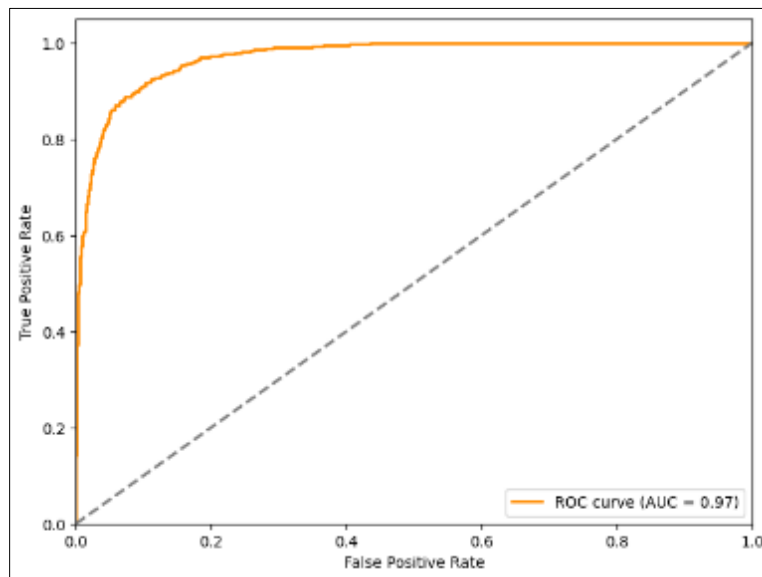


Figure 6 Federated Learning framework for ALL subtype classification.

Figure 6 further shows the ROC-AUC curve which indicates how well the model can differentiate two classes. The curve has an AUC of 0.97 and is an excellent tradeoff between true positive and false positive rates. We demonstrate that the curve is placed close to the top-left corner, which further confirms the high discriminative capability of the model in its ability to distinguish different categories.

For the robustness of the model, 5-fold cross-validation was performed and the result is presented in Figure 7. The test accuracy for each fold was as follows: Fold 1 got 95.13, Fold 2 got 98.15, Fold 3 got 95.88, Fold 4 got 96.54, and Fold 5 got 96.41. The mean accuracy across all folds shows the reliability of the trained model (the model has minimal performance variation), indicating the model’s stability and consistency.

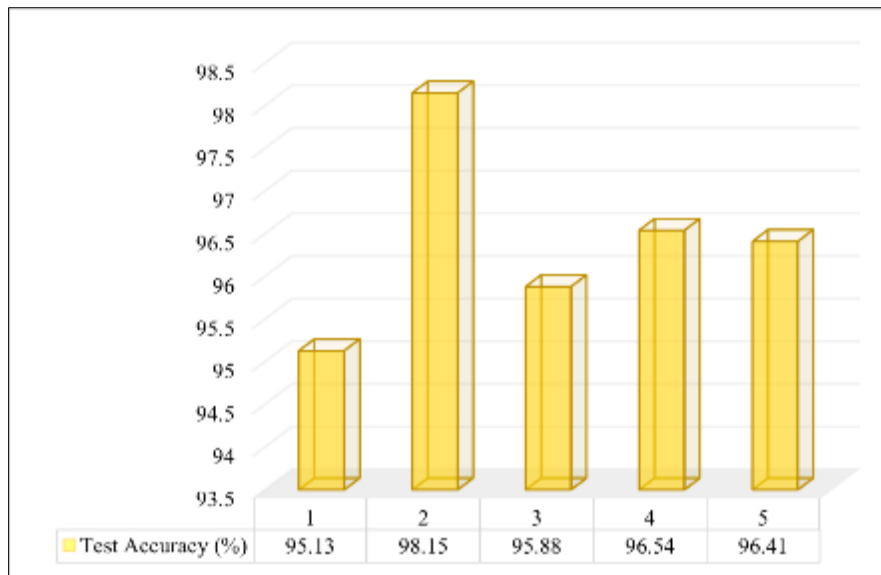


Figure 7 Federated Learning framework for ALL subtype classification.

4.3. Discussion

The results from the experiments and performance evaluations confirm the effectiveness and robustness of the InceptionV3 model in classifying ALL subtypes, especially in the Federated Learning setting. In standalone transfer learning experiments, InceptionV3 performed much better with 92.42% accuracy, 92.90% precision, 91.80% recall and 92.35% F1 score. The fact that it is likely due to a superior spatial and channel acquisition (and therefore spatial and channel-dependent information) that the advanced architecture provides is likely the cause of this increase. In a Federated Learning setting InceptionV3 maintained high performance resulting in 95.49% accuracy, 96.10% precision, 94.80% recall, and 95.45% F1 score. In these results, we show that our model not only holds up well in centralized training but also succeeds in Federated Learning, where training is executed over multiple client devices and the parameters are aggregated on a central server. For medical applications, these requirements mean that this is significant.

While both the Xception and DenseNet121 also performed well in the Federated Learning framework their results were slightly poorer than InceptionV3. Where their accuracies of 94.15% and 93.21, respectively, show that although they are strong models, they do not reach the overall performance of Inception V3 on this specific data. The ROC-AUC curve shows the good discriminatory performance of the InceptionV3 model ($AUC=0.97$) indicating that there is excellent separation between the distinct classes. In combination with the analysis of the confusion matrix, this metric reveals two things: first, that InceptionV3 is making a majority of correct predictions; second, that these mistakes are not being made with large errors (by having a low number of misclassified cases). The model is stable and consistent on other subsets of data confirmed by 5-fold Cross-validation. Test accuracy varied a minimal amount (from 95.13 to 98.15%) across the folds, making the model very reliable and perceivable on unseen data. Finally, for ALL subtype classification, InceptionV3 finds to be the most effective model, both in a centralized, as well as, a Federated Learning environment. The robust performance, high accuracy, and stability in different evaluation metrics confirm its suitability for real-world medical image classification tasks.

5. Conclusion

We present a privacy-preserving and robust framework for machine learning-based classification of ALL subtypes with FL and state-of-the-art deep learning models in this study. Using a large dataset of PBS images and using advanced transfer learning models such as InceptionV3 the proposed framework achieves high diagnostic accuracy while providing security and privacy for the sensitive patient data. This research demonstrates how decentralized learning can contribute to medical image analysis by providing solutions to certain challenges like class imbalance with data augmentation and enforcing data protection regulations by using differential private mechanisms. Experimental results of the FL architecture are significantly better than centralized learning and are shown to be scalable and robust to diverse distributed datasets. The FL paradigm integrates cutting-edge modelling to ensure that learning happens efficiently amongst institutions without compromising the privacy of the data, and this is a tremendous step forward in hematological oncology. The findings suggest the need for adopting privacy-preserving AI technologies in healthcare to

improve early detection, accuracy of diagnosis, and where possible better health outcomes. This work can still be extended into future research by adding more data sources, using ensemble methods, and using this framework on other medical imaging tasks.

Compliance with ethical standards

Disclosure of conflict of interest

The authors declare that they have no competing financial interests or personal relationships that may have influenced the work reported in this study.

References

- [1] Dong Y, Shi O, Zeng Q, Lu X, Wang W, Li Y, Wang Q, Wang Q, Wang Q. Leukemia incidence trends at the global, regional, and national level between 1990 and 2017. *Exp Hematol Oncol* 2020; 9:1–11. <https://doi.org/10.1186/S40164-020-00170-6/FIGURES/5>.
- [2] Saeed A, Shoukat S, Shehzad K, Ahmad I, Eshmawi AA, Amin AH, Tag-Eldin E. A Deep Learning-Based Approach for the Diagnosis of Acute Lymphoblastic Leukemia. *Electronics* 2022, Vol 11, Page 3168 2022;11:3168. <https://doi.org/10.3390/ELECTRONICS11193168>.
- [3] Sánchez Suárez MDM, Martín Roldán A, Alarcón-Payer C, Rodríguez-Gil MÁ, Poquet-Jornet JE, Puerta Puerta JM, Jiménez Morales A. Treatment of Chronic Lymphocytic Leukemia in the Personalized Medicine Era. *Pharmaceutics* 2024, Vol 16, Page 55 2023;16:55. <https://doi.org/10.3390/PHARMACEUTICS16010055>.
- [4] Hosseini A, Eshraghi MA, Taami T, Sadeghsalehi H, Hoseinzadeh Z, Ghaderzadeh M, Rafiee M. A mobile application based on efficient lightweight CNN model for classification of B-ALL cancer from non-cancerous cells: A design and implementation study. *Inform Med Unlocked* 2023; 39:101244. <https://doi.org/10.1016/J.IMU.2023.101244>.
- [5] K.T N, Prasad K, Singh BMK. Analysis of red blood cells from peripheral blood smear images for anemia detection: a methodological review. *Med Biol Eng Comput* 2022; 60:2445–62. <https://doi.org/10.1007/S11517-022-02614-Z/FIGURES/8>.
- [6] Lasker A, Obaidullah SM, Chakraborty C, Roy K. Application of Machine Learning and Deep Learning Techniques for COVID-19 Screening Using Radiological Imaging: A Comprehensive Review. *SN Computer Science* 2022 4:1 2022;4:1–34. <https://doi.org/10.1007/S42979-022-01464-8>.
- [7] Qayyum A, Ahmad K, Ahsan MA, Al-Fuqaha A, Qadir J. Collaborative Federated Learning for Healthcare: Multi-Modal COVID-19 Diagnosis at the Edge. *IEEE Open Journal of the Computer Society* 2022; 3:172–84. <https://doi.org/10.1109/OJCS.2022.3206407>.
- [8] Das S, Kuppan P. Optimisation with deep learning for leukaemia classification in federated learning. *Int J Bioinform Res Appl* 2024; 20:556–83. <https://doi.org/10.1504/IJBRA.2024.142549>.
- [9] Ghongade OS, Reddy SKS, Gavini YC, Tokala S, Enduri MK. Acute Lymphoblastic Leukemia Blood Cells Prediction Using Deep Learning & Transfer Learning Technique. *Indonesian Journal of Electrical Engineering and Informatics (IJEI)* 2023; 11:778–90. <https://doi.org/10.52549/IJEI.V11I3.4855>.
- [10] Ullah MZ, Zheng Y, Song J, Aslam S, Xu C, Kiazolu GD, Wang L. An Attention-Based Convolutional Neural Network for Acute Lymphoblastic Leukemia Classification. *Applied Sciences* 2021, Vol 11, Page 10662 2021;11:10662. <https://doi.org/10.3390/APP112210662>.
- [11] Haque R, Al Sakib A, Hossain MF, Islam F, Ibne Aziz F, Ahmed MR, Kannan S, Rohan A, Hasan MJ. Advancing Early Leukemia Diagnostics: A Comprehensive Study Incorporating Image Processing and Transfer Learning. *BioMedInformatics* 2024, Vol 4, Pages 966-991 2024;4:966–91. <https://doi.org/10.3390/BIOMEDINFORMATICS4020054>.
- [12] Saeed A, Shoukat S, Shehzad K, Ahmad I, Eshmawi AA, Amin AH, Tag-Eldin E. A Deep Learning-Based Approach for the Diagnosis of Acute Lymphoblastic Leukemia. *Electronics* 2022, Vol 11, Page 3168 2022;11:3168. <https://doi.org/10.3390/ELECTRONICS11193168>.

- [13] Ansari S, Navin AH, Sangar AB, Gharamaleki JV, Danishvar S. A Customized Efficient Deep Learning Model for the Diagnosis of Acute Leukemia Cells Based on Lymphocyte and Monocyte Images. *Electronics* 2023, Vol 12, Page 322 2023;12:322. <https://doi.org/10.3390/ELECTRONICS12020322>.
- [14] Pałczyński K, Śmigiel S, Gackowska M, Ledziński D, Bujnowski S, Lutowski Z. IoT Application of Transfer Learning in Hybrid Artificial Intelligence Systems for Acute Lymphoblastic Leukemia Classification. *Sensors* 2021, Vol 21, Page 8025 2021;21:8025. <https://doi.org/10.3390/S21238025>.
- [15] Hosseini A, Eshraghi MA, Taami T, Sadeghsalehi H, Hoseinzadeh Z, Ghaderzadeh M, Rafiee M. A mobile application based on efficient lightweight CNN model for classification of B-ALL cancer from non-cancerous cells: A design and implementation study. *Inform Med Unlocked* 2023; 39:101244. <https://doi.org/10.1016/J.IMU.2023.101244>.
- [16] Blood Cells Cancer (ALL) dataset n.d. <https://www.kaggle.com/datasets/mohammadamireshraghi/blood-cell-cancer-all-4class/data> (accessed December 23, 2024).
- [17] Khalifa NE, Loey M, Mirjalili S. A comprehensive survey of recent trends in deep learning for digital images augmentation. *Artif Intell Rev* 2022; 55:2351–77. <https://doi.org/10.1007/S10462-021-10066-4/TABLES/5>.
- [18] Chand S, Vishwakarma VP. A novel Deep Learning Framework (DLF) for classification of Acute Lymphoblastic Leukemia. *Multimed Tools Appl* 2022; 81:37243–62. <https://doi.org/10.1007/S11042-022-13543-2/METRICS>.
- [19] Mustaqim T, Fatichah C, Suciati N. Deep Learning for the Detection of Acute Lymphoblastic Leukemia Subtypes on Microscopic Images: A Systematic Literature Review. *IEEE Access* 2023; 11:16108–27. <https://doi.org/10.1109/ACCESS.2023.3245128>.
- [20] Chand S, Vishwakarma VP. Acute Leukaemia Diagnosis Using Transfer Learning on Resnet-50. *ACM International Conference Proceeding Series* 2022. <https://doi.org/10.1145/3590837.3590839>.
- [21] Jawahar M, Anbarasi LJ, Narayanan S, Gandomi AH. An attention-based deep learning for acute lymphoblastic leukemia classification. *Scientific Reports* 2024 14:1 2024;14:1–20. <https://doi.org/10.1038/s41598-024-67826-9>.
- [22] Mulya RF, Utami E, Ariatmanto D. Classification of Acute Lymphoblastic Leukemia based on White Blood Cell Images using InceptionV3 Model. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)* 2023; 7:947–52. <https://doi.org/10.29207/RESTI.V7I4.5182>.
- [23] Ananthu KS, Krishna Prasad P, Nagarajan S, Vimina ER. Acute Lymphoblastic Leukemia Detection Using Transfer Learning Techniques. *Lecture Notes in Networks and Systems* 2022; 213:679–92. https://doi.org/10.1007/978-981-16-2422-3_53.
- [24] Ahmed IA, Senan EM, Shatnawi HSA, Alkhraisha ZM, Al-Azzam MMA. Hybrid Techniques for the Diagnosis of Acute Lymphoblastic Leukemia Based on Fusion of CNN Features. *Diagnostics* 2023, Vol 13, Page 1026 2023;13:1026. <https://doi.org/10.3390/DIAGNOSTICS13061026>.
- [25] Alim MS, Bappon SD, Sabuj SM, Islam MJ, Tarek MM, Azam MS, Islam MM. Integrating convolutional neural networks for microscopic image analysis in acute lymphoblastic leukemia classification: A deep learning approach for enhanced diagnostic precision. *Systems and Soft Computing* 2024; 6:200121. <https://doi.org/10.1016/J.SASC.2024.200121>.
- [26] Kumar Y, Shrivastav S, Garg K, Modi N, Wiltos K, Woźniak M, Ijaz MF. Automating cancer diagnosis using advanced deep learning techniques for multi-cancer image classification. *Scientific Reports* 2024 14:1 2024; 14:1–25. <https://doi.org/10.1038/s41598-024-75876-2>.
- [27] Putzu L, Loddo A, Porcu S. White Blood Cell Classification Via Distributed Collaborative Machine Learning 2024. <https://doi.org/10.2139/SSRN.4732432>.
- [28] Kapustina O, Burmakina P, Gubina N, Serov N, Vinogradov V. User-friendly and industry-integrated AI for medicinal chemists and pharmaceuticals. *Artificial Intelligence Chemistry* 2024; 2:100072. <https://doi.org/10.1016/J.AICHEM.2024.100072>.
- [29] Sheakh MA, Tahosin MstS, Akter L, Jahan I, Islam MN, Siddiky MR, Hasan MM, Hasan S. Comparative analysis of machine learning algorithms for ECG-based heart attack prediction: A study using Bangladeshi patient data. <https://WjarrCom/Sites/Default/Files/WJARR-2024-2928Pdf> 2024; 23:2572–84. <https://doi.org/10.30574/WJARR.2024.23.3.2928>.

- [30] Islam T, Sheakh MA, Tahosin MS, Hena MH, Akash S, Bin Jordan YA, FentahunWondmie G, Nafidi HA, Bourhia M. Predictive modeling for breast cancer classification in the context of Bangladeshi patients by use of machine learning approach with explainable AI. *Scientific Reports* 2024 14:1 2024; 14:1-17. <https://doi.org/10.1038/s41598-024-57740-5>.
- [31] Tahosin MS, Sheakh MA, Islam T, Lima RJ, Begum M. Optimizing brain tumor classification through feature selection and hyperparameter tuning in machine learning models. *Inform Med Unlocked* 2023; 43:101414. <https://doi.org/10.1016/J.IMU.2023.101414>.
- [32] Sheakh MA, Sazia Tahosin M, Hasan MM, Islam T, Islam O, Rana MM. Child and Maternal Mortality Risk Factor Analysis Using Machine Learning Approaches. *ISDFS 2023 - 11th International Symposium on Digital Forensics and Security 2023*. <https://doi.org/10.1109/ISDFS58141.2023.10131826>.