Iraqi Journal of Science, 2025, Vol. 66, No. 11, pp: 5159- 5167
DOI: 10.24996/ijs.2025.66.11.37

ISSN: 0067-2904

Acute Lymphoblastic Leukemia Classification Using Modified VGG16 Architecture

Fallah H. Najjar[1, 2]*, Salman Abd Kadum[1], Ola N. Kadhim[3], Ali J. Ramadhan[4]
[1]*Department of Computer System Techniques, Technical Institute of Najaf, Al-Furat Al-Awsat Technical University, 54001 Najaf, Iraq*
[2]*Department of Emerging Computing, Faculty of Computing, Universiti Teknologi Malaysia, Johor Bahru, Malaysia*
[3]*Technical Institute of Al-Mussaib, Al-Furat Al-Awsat Technical University, 54001 Najaf, Iraq*
[4]*Department of Computer Techniques Engineering, College of Technical Engineering, University of Alkafeel, 54001 Najaf, Iraq*

Received: 27/4/2024 Accepted: 12/10/2024 Published: 30/11/2025

Abstract

Acute lymphoblastic leukemia (ALL) diagnosis is a challenge, including invasive classical methods, which are time-consuming, inaccurate, and error-prone. In this paper, we propose modifying the VGG16 architecture to improve its performance in the classification task. We utilize the Acute Lymphoblastic Leukemia (ALL) image dataset to train the proposed modified VGG16 model. The dataset was split into training and testing sets at a ratio of 80% for training data, 10% for validation data, and 10% for testing data. The ALL dataset consists of four classes: Benign, Early, Pre, and Pro. The results of the proposed modified VGG16 model were very satisfactory: accuracy, 96.59%; precision, 96.61%; sensitivity, 96.59%; F1-score, 96.58%; and Matthew's correlation coefficient of 95.35%. It was demonstrated that the image size also influences the model, indicating a trade-off based on how efficient a computational one can be concerning classification accuracy. These findings highlight the promise of deep learning algorithms to revolutionize all characterizations and offer potential utility for future applications in medical imaging.

Keywords: Acute Lymphoblastic Leukemia, ALL, VGG16, medical imaging, deep learning.

تصنيف سرطان الدم الليمفاوي الحاد باستعمال بنية VGG16 المعدلة

فلاح حسن نعمة[1، 2]*, سلمان عبد كاظم[1], عُلا نجاح كاظم[3] , علي جاسم رمضان[4]
[1]قسم تقنيات أنظمة الحاسوب، المعهد التقني النجف، جامعة الفرات الأوسط التقنية، النجف الاشرف، العراق
[2]قسم الحوسبة الناشئة، كلية الحوسبة ، الجامعة التقنية الماليزية، جوهور باهرو، ماليزيا
[3]المعهد التقني المسيب، جامعة الفرات الأوسط التقنية، النجف الاشرف، العراق
[4]قسم تقنيات هندسة الحاسوب, كلية التقنيات الهندسية, جامعة الكفيل, النجف الاشرف, العراق

*Email: fallahnajjar@atu.edu.iq

الخلاصة

يعد تشخيص سرطان الدم الليمفاوي الحاد (ALL) تحديًا، بما في ذلك للطرق الكلاسيكية الغازية، والتي تستغرق وقتًا طويلاً وغير دقيقة وعرضة للخطأ. في هذه الورقة، نقترح تعديل بنية VGG16 لتحسين أدائها في مهمة التصنيف. نستعمل مجموعة بيانات صور سرطان الدم الليمفاوي الحاد (ALL) لتدريب نموذج VGG16 المعدل المقترح. تم تقسيم مجموعة البيانات إلى مجموعات تدريب واختبار بنسبة 80٪ لبيانات التدريب، و10٪ لبيانات التحقق و10٪ لبيانات الاختبار. تتكون مجموعة بيانات ALL من أربع فئات: حميدة ومبكرة ومسبقة واحترافية. كانت نتائج نموذج VGG16 المعدل المقترح مرضية للغاية: الدقة 96.59٪؛ والدقة 96.61٪؛ والحساسية 96.59٪؛ ودرجة F1 96.58٪؛ ومعامل ارتباط ماثيو 95.35٪. وقد تبين أن حجم الصورة يؤثر أيضًا على النموذج، مما يشير إلى وجود تنازلات بناءً على مدى كفاءة الحوسبة فيما يتعلق بدقة التصنيف. وتسلط هذه النتائج الضوء على وعد خوارزميات التعلم العميق بإحداث ثورة في توصيف ALL وتقديم فائدة محتملة للتطبيقات المستقبلية في التصوير الطبي.

1. Introduction

The classification of ALL diseases would improve diagnostic accuracy using supervised machine learning techniques, particularly in identifying patterns and anomaly features that might have escaped the notice of the human eye during manual assessment [1, 2]. Given cutting-edge algorithms and state-of-the-art machine learning, machine and deep learning techniques can scrutinize blood smear images in a way faster than human experts can [3, 4]. A drastic reduction in diagnosis time is likely to be the result [5, 6]. Moreover, the AI system is trainable to pick out even the most minor differences in the features of the cells that may indicate the conditions of ALL, hence being very important in early detection and intervention. AI has, therefore, been helpful in the process of diagnosing ALL by making it more efficient, exact, repeatable, and reliable [7]. AI systems, as a result, can maintain the same level of analysis, while human experts, out of fatigue and subjective judgments, may interpret differently [8, 9]. Such systems are further observed to be continuously updated with new data, hence developing and adapting with time in consideration of the accumulated diagnostic information.

Furthermore, the application of AI in medical diagnostics pushes interdisciplinary research [10, 11]. For instance, a medical practitioner is likely to work closely with a data scientist or engineer to improve the AI model to perfection [12, 13]. This promotes interdisciplinary research and fosters innovation, as it allows for the advancement of a single field of study. In the end, using machine learning and/or deep learning models to diagnose diseases like ALL will not only make medical diagnosis faster and more accurate, but it will also give patients more hope for better outcomes because they can get help earlier and more accurately [14, 15]. In line with AI development, its role in medicine will be one of rapid increase and ever more critical, if not vital, to modern and future medical practice [16, 17].

However, our main contributions to our work are: 1) We adapt the pre-trained VGG16 model for ALL detection by persistently fine-tuning necessary convolution layers from block1_conv1 to block4_conv3. These networks together incorporate the feature extraction capabilities of VGG16 while enabling the model to learn leukemia-specific features. 2) We modified the VGG16 architecture by appending a new fully connected layer with 1024 units, followed by a dropout layer (0.25 dropout rate). Using a dropout layer can prevent overfitting, and incorporating a convolution layer with common physiological customizations results in a model capable of fitting complex morphological patterns across ALL.

2. Related Works

These days, medical science is witnessing an extraordinary growth in the automatic diagnosis and treatment of leukemia disease through the integration of artificial intelligence (AI) and deep learning. There have been several studies in this direction, and they can be thematically categorized according to the techniques and models used. One common thread across multiple studies is using dimension-reduction techniques to address high-dimensional medical data. Mahdi et al. [18] proposed SGD-SPCA, combining supervised principal component analysis (SPCA) with the classic stochastic gradient descent (SGD) algorithm. They implemented a scalable application intended for a low number of samples and high-dimensional data, suitable for cancer diagnosis and changing parameters without delay. It is most beneficial in a big data environment when updates need to be done all the time. However, Tusar and Anik [10] successfully investigated deep learning to detect leukemia using images. They used CNNs as feature extractors of leukemia images, allowing them to find benign and malignant cases when tried against the data. Similarly, Rahman et al. [19] used the DarkNet-19 model and the Mrmr algorithm. With 1000 features, their method achieves an accuracy rate of 99.94% for the ALL diagnosis, which is among the highest reported in the literature.

DOĞAN et al. [20] applied a deep neural network (DNN) to automate the detection of all subtypes of ALL from blood smear images, which achieved 98% accuracy. This study also presents an avenue for remote diagnosis that could improve access to healthcare. The other approach utilized pre-trained CNNs for classification tasks. Gokulkrishnan et al. [21] utilized the ResNet-50 and ResNet-101 models for training on a public ALL dataset, yielding results exceeding 98% accuracy. Hagar et al. [22] evaluated two models, VGG16 and DenseNet-121, to distinguish up to eight leukemia classes, achieving a reported classification accuracy of 98.2%. The integration of hybrid AI techniques has also been highly successful. Atteia et al. [23] suggested a way to use a hybrid approach that combined PCA-PSO and neural models that had already been trained to pull out features. Their technique was able to achieve 97.4%, demonstrating the good performance of bioinspired methods in medical image processing applications.

We verify the operation of our technique by exploiting the recent advancements in data augmentation and deep learning-based leukemia classification. Most models today only use pre-trained architectures, but we use a VGG16 model that has already been trained and updates certain convolutional layers to find differences between leukemia subtypes. This fine-tuning, combined with customized fully connected layers and a strong data augmentation strategy that suits our dataset, is enough for the model to be more accurate than many existing approaches or equal its weight without creating an overfitting scenario. Unlike techniques like SGD-SPCA [18] and bio-inspired algorithms [23], our technique directly classifies images using CNNs rather than reducing data dimensionality or optimizing feature selection. Also, our work is different from that of Rahman et al. [19] and Gokulkrishnan et al. [21], who used pre-trained networks and created more data by fine-tuning at the layer level to stop overfitting and train a more general model. While previous studies have achieved high performance with pre-trained models or hybrid techniques, our study focuses on teaching the model to be flexible and adaptable specifically for the early detection and classification of ALL, thereby enhancing diagnostic precision and improving generalization.

3. Methodology

Our proposed methodology, shown in Figure 1, consists of four main stages: pre-processing, data augmentation, VGG16 architecture design, and evaluation. In the pre-processing stage, the dataset images were resized with dimensions of 300 × 300 pixels and were used as the primary input for the deep learning model. The second stage is data

augmentation, which involves applying random transformations to the training images, incorporating adjustments in brightness within the specified range between 0.8 and 1.2, rotation, and reflection. The study utilized a dataset consisting of 3,256 peripheral blood smear (PBS) images [24, 25]. The investigators evaluated images from 89 patients, all of whom were referred for detection and treatment of ALL. The main classes of the dataset are two: benign, 504 images, and malignant, 2752 lymphoblast images. The malignant is divided into three subtypes: early (985 images), pre (963 images), and pro (804 images). However, the code was executed in a Jupyter Notebook environment on a Windows 11 MSI laptop with the following hardware specifications: Intel Core i7, 11th Generation processor, 32 GB RAM, 1 TB SSD storage, and an NVIDIA GPU with 8 GB VRAM. Furthermore, the TensorFlow, NumPy, OpenCV, and Matplotlib libraries were utilized to build and train the deep learning model.

Further, in order to enhance the flexibility of the model, the dataset was split into three sets: 80% is allocated to training, 10% is allocated to validation, and an additional 10% is for testing. In contrast, to increase the generalization ability of the model, a data augmentation strategy was implemented. In the third stage, VGG16 architecture design, the pre-trained VGG16 model is imported, and the top layer is executed.

Subsequently, specific convolutional layers were identified as trainable, enabling the model to adapt to the nuances of leukemia subtypes. To fit the model to the complexities of ALL classification, new fully connected layers were appended to the VGG16 structure. The flattening layer was successful in creating a dense layer of 1024 units, utilizing a Rectified Linear Unit (ReLU) activation function. The sealant layer, which features a leakage rate of 0.25, addressed concerns about overfitting. Finally, the SoftMax activation function produced a classification output to distinguish between leukemia subtypes. Furthermore, the model was trained with a learning rate of 1e-05, a batch size of 32, and an epoch number of 25. The goal of training the model from the beginning, instead of using pre-existing weights, was to infuse it with the ability to discriminate the complex morphological patterns inherent in cancer images.

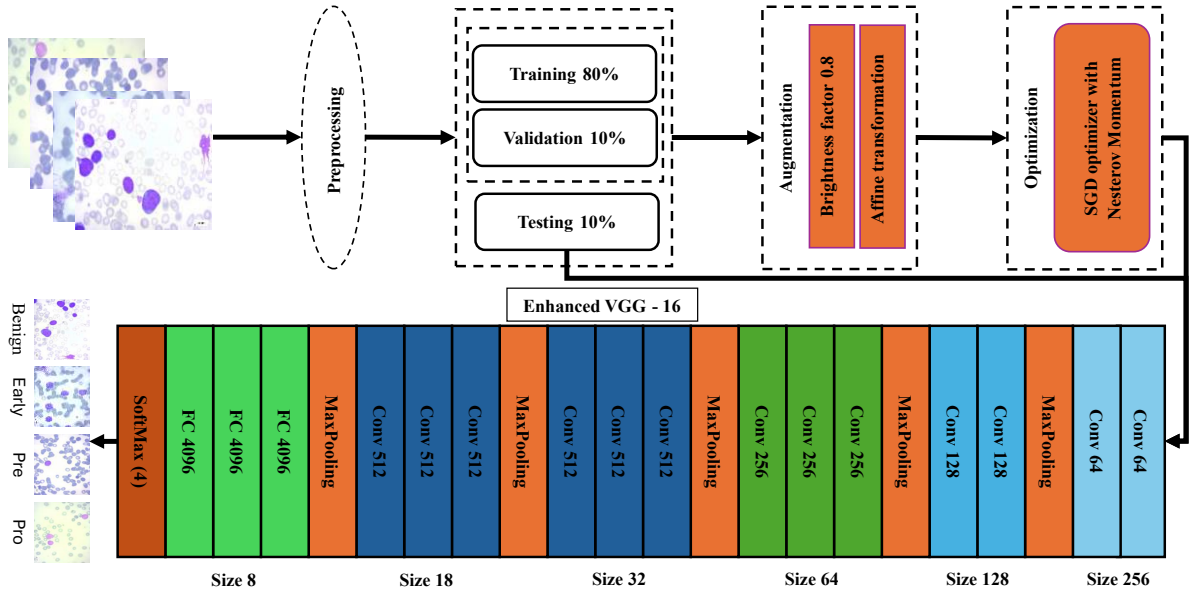

Figure 1: Our Proposed Method

After training, the proposed modified VGG16 undergoes evaluation using a set of commonly used metrics. The metrics being evaluated are accuracy [24], precision [24], recall [24], F1 score [24], and Matthews Correlation Coefficient (MCC) [24]. Each metric is calculated as the following:

$$accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{1}$$

$$recall = \frac{TP}{TP + FN} \tag{2}$$

$$precision = \frac{TP}{TP + FP} \tag{3}$$

$$F1 = \frac{2 \times precision + recall}{precision + recall} \tag{4}$$

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{5}$$

Where TP and FP represent true and false positives, TN and FN represent true and false negatives.

4. Experimental Results

As previously mentioned, the model demonstrated outstanding performance across various evaluation metrics after training with resized images. It is worth noting that its accuracy is 96.59%, which highlights its skill in accurately classifying leukemia classes. The model effectively reduces both false positives and false negatives, as evidenced by its precision score of 96.61 and recall score of 96.59%. The F1 score, which is a combination of precision and recall, reaches 96.58%, indicating a well-balanced ability to capture both positive and negative cases.

The MCC scores 95.35%, confirming the effectiveness of the model in overcoming the inherent complexities of classifying hematological malignancies. Furthermore, the training and validation accuracies and losses are displayed in Figure 2.

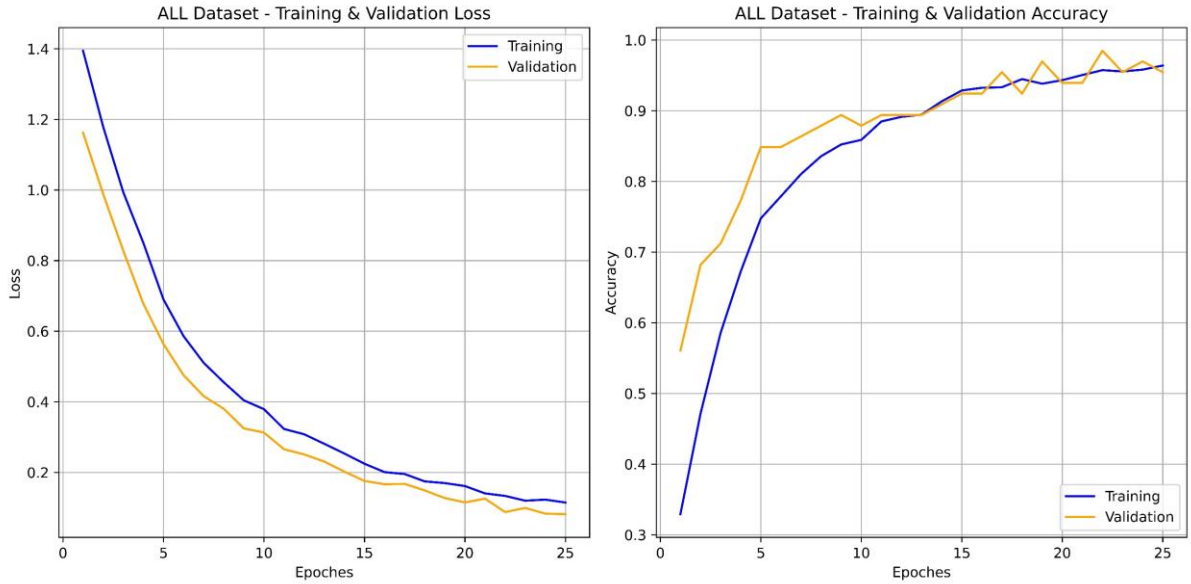

Figure 2: Training and validation results (loss and accuracy).

The entire experimentation process was executed on computational resources, and the total time, in seconds (s), elapsed during training versus accuracy results are displayed in Table 1.

Table 1: Training time and accuracy comparison across different image sizes.

Experimental	Image size	Training time (s)	Accuracy (%)
1	300 × 300	12,024	96.59
2	250 × 250	10,928	96.31
3	200 × 200	9,237	95.88
4	150 × 150	8,024	95.11

From Table 1, a clear pattern emerges with the model's accuracy gradually decreasing as the image dimensions decrease; as the image size decreases, so does the training time. However, further exploration of improvements in the balance between classification accuracy and computational efficiency is required. Figure 3 also displays the confusion matrix of the test set.

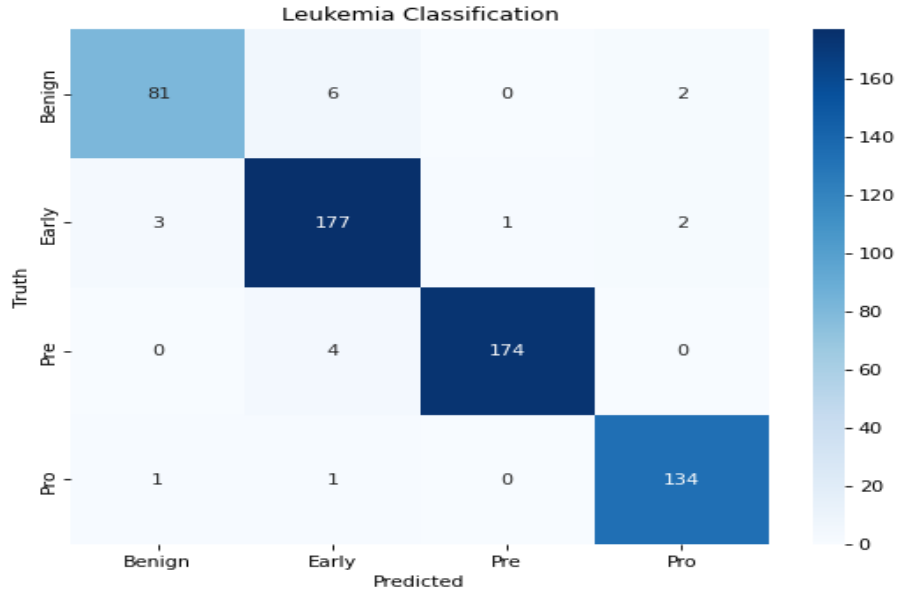

Figure 3: ALL Classification - Confusion matrix.

The confusion matrix in Figure 3 shows the exceptional quality of our model to detect leukemia cases: high accuracy for all four classes, Benign, Early, Pre, and Pro. The diagonal's high values indicate the correct classification of many samples. The model was particularly effective at false-positive cases, where it correctly identified 81 out of 89 benign cases. However, there were instances where the model misclassified early-stage malignant subtypes like Early and Pro. This was particularly visible in the misclassifications of Early and Pre subtypes, which showed very similar morphology and some overlap in predictions. The Pro subtype has the least overlap and the highest accuracy in its predicted class.

These results demonstrate the model's ability to accurately classify even the most challenging cases as either benign or malignant, with only a few exceptions where confusion arises, a common occurrence among closely related leukemic subtypes. Collectively, these results indicate that our model could serve as a valuable tool for isolating leukemia from PBS images, albeit requiring some adjustments to enhance its ability to distinguish between various malignant subtypes.

The proposed method not only performs well but also positions as a promising and computational method compared to the state-of-the-art (SOTA) in the field of ALL classification. Furthermore, Table 2 displays the performance metrics of the proposed method compared to the SOTA.

Table 2: Comparative analysis of the performance metrics on the testing set of the ALL dataset [25, 26].

Method (Year)	Accuracy	Precision	Sensitivity	F1	MCC
Tusar and Anik (2022) [10]	96.13	-	-	-	-
Mahdi et al. (2021) [18]	94.40	-	-	-	-
Kadhim et al. (2023) [14]	98.15	96.00	94.68	95.24	-
Our proposed method	96.59	96.61	96.59	96.58	95.35

The obtained results for classifying the ALL dataset show impressive performance when compared to the SOTA. Exceeding the 96.13% for Tosar and Anik [10] and Mahdi et al. [18] in terms of accuracy and 98.15% of Kadhim et al. [14] in terms of precision, recall, and F1, our model achieves an accuracy of 96.59%. The accuracy of 96.61% confirms a high degree of accuracy in mitigating false positives, while the recall of 96.59% demonstrates a strong ability to capture true positive cases. The F1 score, which combines precision and recall, reaches 96.58%, highlighting the balanced performance of our model. The MCC is 95.35%, indicating a solid ability to address stratification imbalances in the classification of hematological malignancies.

Comparing the results achieved with this model to those of SOTA techniques, performance is equal for most indicators. For example, Kadhim et al. [14] had a slightly higher precision of 98.15%. Our proposed model gave good grades in contiguous tests: its accuracy at 96.59%, or even its precision and sensitivity were lower than but close to 96.61% for both measures. In other words, this is true of the probability score for F1 of 96.58%. The score of F1 represents that the model is good at picking up positives, as indicated by its precision, and ensures most true positives are caught, as shown by its sensitivity. Compared with Tusar and Anik [10] or Mahdi et al. [18], who achieved 96.13% and 94.40%, respectively, this method outperforms them, especially in keeping high accuracy and a balanced perspective across different ALL subtypes.

In contrast to other SOTA research, the proposed method's performance is attributable to several key factors. First, the modified VGG16 architecture was used to selectively fine-tune specific convolutional layers, which allowed the model to capture detailed and relevant features of leukemia subtypes. Second, various data augmentation techniques, random brightness adjustments, rotations, and reflections, enhanced the generalization capacity of models by expanding training data diversity. Finally, combining a dropped layer for preventing overfitting and a well-structured, fully connected layer was beneficial in ensuring that the model learned strong patterns. Altogether, these factors contributed to the model's 96.59% accuracy, outweighing other SOTA methods in terms of both accuracy and balance performance among evaluation metrics.

5. Conclusions

Our study suggested a modified VGG16 deep learning model to classify ALL in PBS images. The model captured the complex morphological features of different leukemia subtypes utilizing the modifications employed, such as fine-tuning (convolutional layers), fully connected layers, and dropout layers. Moreover, after performing data augmentation

like random brightness changes, rotations, and flips, the model showed better generalization, avoiding overfitting.

The proposed model tested on the experimental results got 96.59% accuracy, whereas precision is 96.61%, recall is 96.59%, F1-score is about 96.58%, and MCC is approximately 95.35%. These findings demonstrate the proposed model's well-controlled specificity and sensitivity in identifying Benign, Early, Pre, or Pro classes and distinguishing between their subgroups. The model performance was competitive with/improved upon several SOTA methods for accuracy and other evaluation metrics.

The study also investigated how image size can affect computational cost and classification performance, finding a trade-off between these. It remains an open problem to strike a better trade-off between effectiveness and efficiency, which will be explored in our future work. In general, the proposed method provides a reliable and efficient large-scale solution for ALL classification that could potentially assist in diagnosing medical imaging more accurately and earlier, ultimately improving patient outcomes.

References

[1] R. Najjar, "Redefining radiology: a review of artificial intelligence integration in medical imaging," *Diagnostics,* vol. 13, no. 17, pp. 1-25, 2023, doi: https://doi.org/10.3390/diagnostics13172760.

[2] F. H. Najjar, H. M. Al-Jawahry, M. S. Al-Khaffaf, and A. T. Al-Hasani, "A novel hybrid feature extraction method using LTP, TFCM, and GLCM," presented at the journal of physics: conference series, 2021.

[3] B. Abhisheka, S. K. Biswas, B. Purkayastha, D. Das, and A. Escargueil, "Recent trend in medical imaging modalities and their applications in disease diagnosis: a review," *Multimedia Tools and Applications,* vol. 83, no. 2, pp. 43035–43070, 2023, doi: https://doi.org/10.1007/s11042-023-17326-1.

[4] M. Hadjouni, H. Elmannai, A. Saad, A. Altahe, and A. Elaraby, "A Novel Deep Learning Approach for Brain Tumors Classification Using MRI Images," *Traitement du Signal,* vol. 40, no. 3, pp. 1137-1144, 2023, doi: https://doi.org/10.18280/ts.400329.

[5] P. K. Talupuri, B. Naseeba, N. P. Challa, and A. S. Sathwik, "Analysis of Acute Lymphoblastic Leukemia Detection Methods Using Deep Learning," presented at the International Conference on Self Sustainable Artificial Intelligence Systems (ICSSAS), Erode, India, 2023.

[6] S. Franco, X. Geng, V. Korostyshevskiy, J. E. Karp, and C. Lai, "Systematic review and meta-analysis: Prognostic impact of time from diagnosis to treatment in patients with acute myeloid leukemia," *Cancer,* vol. 129, no. 19, pp. 2975-2985, 2023.

[7] N. Das, S. Gajendra, and R. Gupta, "Analytical Appraisal of Hematogones in B-ALL MRD Assessment Using Multidimensional Dot-Plots by Multiparametric Flow Cytometry: A Critical Review and Update," *Indian Journal of Hematology and Blood Transfusion,* vol. 40, pp. 12-24, 2023, doi: https://doi.org/10.1007/s12288-023-01696-5.

[8] G. Madhu, A. W. Mohamed, S. Kautish, M. A. Shah, and I. Ali, "Intelligent diagnostic model for malaria parasite detection and classification using imperative inception-based capsule neural networks," *Scientific Reports,* vol. 13, no. 1, pp. 1-12, 2023, doi: https://doi.org/10.1038/s41598-023-40317-z.

[9] B. H. Kann, A. Hosny, and H. J. Aerts, "Artificial intelligence for clinical oncology," *Cancer Cell,* vol. 39, no. 7, pp. 916-927, 2021.

[10] M. T. H. K. Tusar and R. K. Anik, "Automated Detection of Acute Lymphoblastic Leukemia Subtypes from Microscopic Blood Smear Images using Deep Neural Networks," Bachelor, Computer Science and Engineering, City university, 2022.

[11] F. H. Najjar, K. T. Khudhair, Z. N. Khudhair, H. H. Alwan, and A. Al-khaykan, "Acute lymphoblastic leukemia image segmentation based on modified HSV model," presented at the Journal of Physics: Conference Series, 2023.

[12] S. Ganesh, T. Chithambaram, N. R. Krishnan, D. R. Vincent, J. Kaliappan, and K. Srinivasan, "Exploring Huntington's Disease Diagnosis via Artificial Intelligence Models: A Comprehensive

Review," *Diagnostics,* vol. 13, no. 23, pp. 1-39, 2023, doi: https://doi.org/10.3390/diagnostics13233592.
[13] Y. Wang, C. Wang, and L. Liu, "Trends in using deep learning algorithms in biomedical prediction systems," *Frontiers in Neuroscience,* vol. 17, pp. 1-32, 2023, doi: https://doi.org/10.3389/fnins.2023.1256351.
[14] K. A. Kadhim, F. H. Najjar, A. A. Waad, I. H. Al-Kharsan, Z. N. Khudhair, and A. A. Salim, "Leukemia Classification using a Convolutional Neural Network of AML Images," *Malaysian Journal of Fundamental and Applied Sciences,* vol. 19, no. 3, pp. 306-312, 2023, doi: https://doi.org/10.11113/mjfas.v19n3.2901.
[15] W. Walter *et al.*, "How artificial intelligence might disrupt diagnostics in hematology in the near future," *Oncogene,* vol. 40, no. 25, pp. 4271-4280, 2021.
[16] A. Mittal, S. Dhalla, S. Gupta, and A. Gupta, "Automated analysis of blood smear images for leukemia detection: a comprehensive review," *ACM Computing Surveys (CSUR),* vol. 54, no. 11s, pp. 1-37, 2022, doi: https://doi.org/10.1145/3514495.
[17] W. Walter, N. Pfarr, M. Meggendorfer, P. Jost, T. Haferlach, and W. Weichert, "Next-generation diagnostics for precision oncology: Preanalytical considerations, technical challenges, and available technologies," *Seminars in Cancer Biology,* vol. 84, no. 9, pp. 3-15, 2022, doi: https://doi.org/10.1016/j.semcancer.2020.10.015.
[18] G. J. Mahdi, B. A. Kalaf, and M. A. Khaleel, "Enhanced supervised principal component analysis for cancer classification," *Iraqi Journal of Science,* vol. 62, no. 40, pp. 1321-1333, 2021, doi: https://doi.org/10.24996/ijs.2021.62.4.28.
[19] W. Rahman, M. G. G. Faruque, K. Roksana, A. S. Sadi, M. M. Rahman, and M. M. Azad, "Multiclass blood cancer classification using deep CNN with optimized features," *Array,* vol. 18, pp. 1-23, 2023, doi: https://doi.org/10.1016/j.array.2023.100292.
[20] Ş. DOĞAN, B. TAŞCI, and T. TUNCER, "Detection of acute lymphocytic leukemia (ALL) with a pre-trained deep learning model," presented at the International Conference on Recent Academic Studies, 2023.
[21] N. Gokulkrishnan, T. Nayak, and N. Sampathila, "Deep Learning-Based Analysis of Blood Smear Images for Detection of Acute Lymphoblastic Leukemia," presented at the International Conference on Electronics, Computing and Communication Technologies (CONECCT), Bangalore, India, 2023.
[22] M. Hagar, F. K. Elsheref, and S. R. Kamal, "A New Model for Blood Cancer Classification Based on Deep Learning Techniques," *International Journal of Advanced Computer Science and Applications,* vol. 14, no. 6, pp. 422-429, 2023, doi: https://dx.doi.org/10.14569/IJACSA.2023.0140645.
[23] G. Atteia, R. Alnashwan, and M. Hassan, "Hybrid Feature-Learning-Based PSO-PCA Feature Engineering Approach for Blood Cancer Classification," *Diagnostics,* vol. 13, no. 16, pp. 1-13, 2023, doi: https://doi.org/10.3390/diagnostics13162672.
[24] G. M. Foody, "Challenges in the real world use of classification accuracy metrics: From recall and precision to the Matthews correlation coefficient," *Plos one,* vol. 18, no. 10, p. e0291908, 2023, doi: https://doi.org/10.1371/journal.pone.0291908.
[25] M. Ghaderzadeh, M. Aria, A. Hosseini, F. Asadi, D. Bashash, and H. Abolghasemi, "A fast and efficient CNN model for B-ALL diagnosis and its subtypes classification using peripheral blood smear images," *International Journal of Intelligent Systems,* vol. 37, no. 8, pp. 5113-5133, 2022, doi: https://doi.org/10.1002/int.22753.
[26] M. G. Mehrad Aria, Davood Bashash, Hassan Abolghasemi, Farkhondeh Asadi, and Azamossadat Hosseini. *Acute Lymphoblastic Leukemia (ALL) image dataset.*, doi: https://doi.org/10.34740/KAGGLE/DSV/2175623.