Suchithra R, et. al. International Journal of Engineering Research and Applications
www.ijera.com
ISSN: 2248-9622, Vol. 13, Issue 12, December 2023, pp 94-102

RESEARCH ARTICLE OPEN ACCESS

Advancements in Leukemia Diagnosis: A Comprehensive Exploration of Machine Learning Approaches

Suchithra R*, Sri Sowmya Vishnubhotla* , Syeda Hafsa Naaz * , Samiksha Kumari *
*Students, Department of Information Science and Engineering, BMS Institute of Technology and Management, Bengaluru-64 **
*Mrs. Bhavya G ***
*Professor, Department of Information Science and Engineering, BMS Institute of Technology and Management, Bengaluru-64***

ABSTRACT:
This research paper provides a comprehensive review of the application of machine learning (ML) in leukemia diagnosis, focusing on acute lymphoblastic leukemia (ALL), chronic lymphocytic leukemia (CLL), acute myeloid leukemia (AML), and chronic myeloid leukemia (CML). Emphasizing the urgency to bridge the gap between algorithm development and clinical implementation, the paper evaluates existing literature, showcasing a diverse landscape of ML methodologies, including CNN, SVM, k-NN, Neural Networks, and Deep Networks. While acknowledging high reported accuracies in previous studies, the paper strategically highlights the real-world applicability of EfficientNet B3 and VGG16 models, addressing accessibility challenges in healthcare settings. The tailored nature of these models, coupled with considerations of efficiency, flexibility, and practical implementation, positions the research as a significant stride towards effective and accessible leukemia detection solutions. The paper concludes by advocating for ongoing refinement, exploration of alternative architectures, and addressing emerging challenges to propel ML applications in leukemia diagnosis into a new era of efficacy and widespread utility.

Index terms: Leukemia, Types of Leukemia-ALL, CLL, CML, AML, CNN, VGG16, EfficientNetB3

Date of Submission: 08-12-2023 Date of acceptance: 23-12-2023

I. INTRODUCTION

In the intricate landscape of hematopoiesis, the bone marrow takes center stage within the central cavity of bones, giving rise to all blood cells through a process known as hematopoiesis. Stem cells, fundamental to this process, undergo differentiation to yield various blood cell types, among which are white blood cells encompassing neutrophils, monocytes, eosinophils, basophils, and lymphocytes. The orchestration of these cellular transformations plays a critical role in maintaining homeostasis within the circulatory system.

Leukemia, a formidable adversary, manifests as cancer in blood cells, instigating the generation of abnormal white blood cells within the bone marrow. The urgency and severity of leukemia's impact vary, with acute leukemia precipitating rapid patient deterioration, while chronic leukemia is characterized by a more gradual progression, potentially falling within lymphocytic or myelogenous categories. Further delineating the landscape, leukemia can be classified into distinct subtypes, including Acute Lymphoblastic Leukemia (ALL), Acute Myelogenous Leukemia (AML), Chronic Lymphocytic Leukemia (CLL), and Chronic Myelogenous Leukemia (CML).

To navigate the complexity of leukemia classification, two systems have been pivotal: the French-American-British (FAB) classification and the World Health Organization (WHO) proposal. Notably, the identification of blast cells in peripheral blood smears serves as a hallmark for diagnosing Acute Myeloid Leukemia (AML), comprising seven subtypes (M-1 to M-7). The conventional methodology employed by hematologists involves meticulous microscopic examination of blood under a light microscope. However, this process is inherently tedious, time-

consuming, and ill-suited for analyzing a vast number of cells.

Acknowledging the limitations of traditional methods, a repertoire of mathematical approaches and technologies has emerged to discern blood cell differentials. This includes innovative techniques crucial for artifact extraction and the identification of leukemia, marking a transformative step towards more efficient and precise diagnostic methodologies. This research endeavors to explore and contribute to this evolving landscape, seeking to enhance our understanding and application of advanced technologies in leukemia diagnosis.

This review delves into the manifold applications of machine learning in both acute and chronic leukemias, encompassing lymphoid and myeloid diseases, and utilizing microscopy and flow cytometry for diagnosis. The overarching objective is to deepen comprehension of current models, identify limitations, and emphasize the importance of future research in optimizing machine learning utilization in this domain. In particular, the review meticulously examines working models, methodologies, and other aspects within the machine learning literature related to the image-based diagnosis of the four types of leukemia . The intent is to contribute to the ongoing exploration of the future of clinical care, with a focus on highlighting the latest technological advancements.

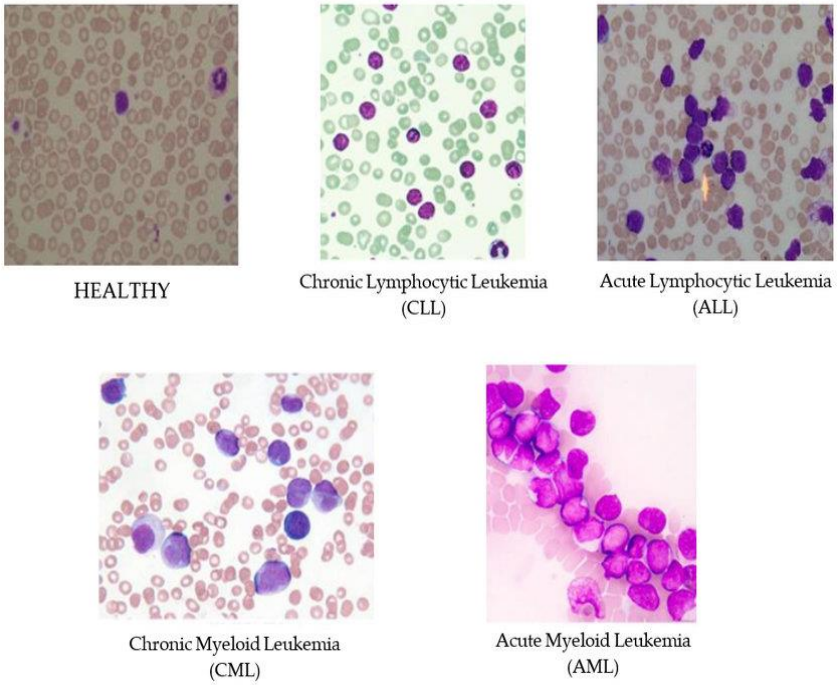


Fig. 1: Classification of Leukemic Cells

II. COLLECTING METHODOLOGY

Accurate and complete diagnosis of leukemia is important for the development of effective treatment strategies and follow-up of the disease. Among the many available technologies, the use of image-based data entry plays an important role in improving the accuracy and efficiency of leukemia diagnosis. This is especially true when compared to other methods such as blood smear microscopy, flow cytometry, and bone marrow aspiration.

Images encompass multidimensional information by combining quantitative and qualitative features and can detect dynamic changes that may be missed with traditional methods. Images provide a content summary that aids in the interpretation of digital data and supports a full understanding of leukemia pathology.

1. **Micro Imaging: Blood Smear Analysis**

- **Description**: Microscopic imaging by examining blood smears. A thin layer of blood is spread on a slide, stained, and monitored for abnormalities in the blood cells.
- **Application in Leukemia**: Microscopic examination is important for morphological analysis of blood. It provides important information

for the diagnosis of leukemia by helping to determine cell abnormalities, size and structure.

2. **Flow Cytometry: Cell Sorting and Counting**
- **Description:** Flow Cytometry is a method that uses lasers to measure the size, shape, and surface markings of a cell to identify and identify each other. It provides a lot of information about many cells in a single model.
- **Use in Leukemia:** Flow cytometry is especially useful in the analysis of blood and bone marrow in leukemia. It helps identify and identify different blood cells that indicate the abnormal cell disease specific to leukemia.

3. **Bone Marrow Aspiration: Sample Collection**
- **Description:** Bone marrow aspiration is the use of a needle to remove a small amount of bone marrow fluid. The samples collected contained a mixture of blood and bone marrow (BMC).
- **Application in Leukemia:** Bone marrow aspiration is a direct method to measure the cellular composition of the bone marrow. It can identify abnormal cell populations, such as leukemia blasts, and provide important information for leukemia diagnosis and classification.
Understanding the advantages and limitations of each technology is important for accurate and successful diagnosis of leukemia. Combining information from multiple imaging modalities can provide a complete picture of the disease and aid in treatment planning and monitoring. This multidisciplinary approach increases the sensitivity and specificity of the test, ultimately leading to accurate diagnosis and better patient outcomes.

III. LITERARY FINDINGS

1.Machine learning applications in the diagnosis of leukemia: Current trends and future directions, 2019

The study, authored by Haneen T. Salah, Ibrahim N. Muhsen, Mohamed E. Salama, Tarek Owaidah, and Shahrukh K. Hashmi, and first published on September 9, 2019, explores the current landscape and future prospects of machine learning applications in the diagnosis of leukemia.

By assessing the current advancements and potential future directions, the study offers a comprehensive overview of the role of machine learning in improving the accuracy and efficiency of leukemia diagnosis, presenting implications for the ongoing development and integration of these technologies in clinical practice.The review provides a comprehensive overview of the increasing integration of machine learning (ML) techniques in the diagnosis of leukemia, focusing on the four common types: acute lymphocytic leukemia (ALL), chronic lymphocytic leukemia (CLL), acute myeloid leukemia (AML), and chronic myelogenous leukemia (CML).

The authors conducted a thorough literature search, utilizing strict selection criteria and electronic databases, supplemented by manual searching and Google Scholar exploration. Out of the 58 articles reviewed, 22 studies met the inclusion criteria, with the majority of them employing supervised learning for classification tasks.Despite challenges such as digitalization, labeling, and costs,the review underscores the potential opportunities offered by augmented human intelligence (AHI) tools, including ML, in shaping the future of medical practice.

Table 1: Summarized view of Paper 1

Methodology	Algorithm	Supervised/ Unsupervised	Accuracy
I. Search Strategies and Data Sources	DCNN[9]	Both	>94%
II. Inclusion Criteria and Study Characteristics	LDA[10]	Supervised	80%
III. Data Collection and Extraction	HCA[6] and SVM[5]	HCA-unsupervised SVM-supervised	NR(not reported)
	GMLVQ[11]	Both	100%

There are however certain drawbacks identified in the paper.

- **Limited Use of Unsupervised Learning and Deep Neural Networks:**

The predominant use of supervised learning in pathology studies is noted as a drawback due to the time-consuming nature of sample labeling. The literature suggests a gap in the exploration of unsupervised learning methodologies, which could mitigate the need for labeled samples and enhance the utilization of larger datasets for more accurate model development.

- **Binary Approach and Complexity of Pathological Diagnosis:**

Many studies employ a "binary" approach, distinguishing between diseased and normal cases. This oversimplified method does not reflect the complex reality of pathological diagnosis. There is a gap in understanding how ML can handle the nuanced and diverse nature of pathological conditions, moving beyond a binary classification.

2.Machine Learning Detection and Classification of Leukemia using C_NMC Leukemia, 2023

The open-access paper, authored by Fatma M. Talaat and Samah A. Gamel, published on June 13, 2023, focuses on the application of machine learning in the detection and classification of leukemia, utilizing the C-NMC_Leukemiadataset.The study likely explores the use of machine learning algorithms for improved accuracy in diagnosing leukemia, utilizing advanced computational techniques to analyze and classify medical data. By leveraging the C-NMC_Leukemia dataset, the authors may investigate the effectiveness of machine learning models in differentiating between various types of leukemia, contributing valuable insights to the ongoing efforts in enhancing diagnostic capabilities for this critical medical condition.The presented systematic review underscores the critical challenge in leukemia diagnostics, emphasizing the need for early detection and precise differentiation of malignant leukocytes at minimal costs.

With limited availability of flow cytometer equipment and laborious methods in diagnostic centers, the review motivates the exploration of machine learning (ML) applications for more efficient and accurate leukemia diagnosis. This ML-based approach involves three main steps: Image Preprocessing, Feature Extraction, and Classification, utilizing an optimized Convolutional Neural Network (OCNN) with fuzzy optimization. The results indicate a remarkable 99.99% accuracy achieved with the C-NMC_Leukemia dataset, highlighting the potential of ML in revolutionizing leukemia diagnostics, particularly in the early stages. The review emphasizes the global impact of leukemia and the challenges in early diagnosis due to the limitations of current diagnostic methods. It discusses the potential of ML in overcoming these challenges, showcasing the success of an OCNN classifier in achieving high accuracy.

Table 2: Summarized view of Paper 2

Methodology	Algorithm	Supervised/ Unsupervised	Accuracy
I.Image Preprocessing Phase (IPP) II.Feature Extraction Phase (FEP) III.Classification Phase (CP).	CNN[8] SVM[5]	Both Supervised	94% 99%

There are however certain drawbacks identified in the paper.

- **Inadequate discussion of computational performance:**

While the article reports an accuracy of 99.99% for the optimized convolutional neural network (OCNN), it does not provide an insight into the performance of the designs. When considering the potential deployment of machine learning models in real clinical settings, it is important to discuss the computational requirements, processing time, and performance requirements of OCNNs, especially given the limitations in many clinical settings.

- **Image preprocessing and subtraction may be good at features:**

This article briefly says that image preprocessing and subtraction are still alarming steps in learning to play music, but do not delve into the injustices shown in these stages. . Biases in prioritizing or ignoring features can affect the model's ability to generalize across different patient populations, and

it is important to understand and reduce these biases, which is a concern for ethical and unethical diagnoses.

3.Machine Learning algorithms for the diagnosis of Leukemia

The authors, Italia Joseph Maria, T. Devi, and D. Ravi, likely discuss the application of various machine learning techniques in the context of leukemia diagnosis. Machine learning algorithms may be explored for their effectiveness in analyzing medical data, possibly including features extracted from blood samples or other diagnostic tests, to aid in the accurate and efficient identification of leukemia. The paper delves into the specifics of the algorithms used, their performance metrics, and their potential implications for improving leukemia diagnosis.

Leukemia, a form of cancer affecting the blood, bone marrow, and lymphatic tissues, is characterized by the abnormal proliferation of white blood cells without the formation of solid tumors. Machine learning algorithms have become integral in the realm of leukemia treatment, aiding in the classification of different leukemia types and the detection of leukemia presence in patients.

This paper focuses on Support Vector Machines, k-Nearest Neighbour, Neural Networks, Naïve Bayes, and Deep Learning algorithms, providing a comparative study of their effectiveness in classifying leukemia subtypes. The algorithms analyze color images of stained blood smears, employing preprocessing steps such as image segmentation and feature extraction to determine the presence of leukemia and classify the specific subtype, considering factors like cell morphology and cytochemical staining.

The study explores various criteria for algorithm comparison, including supervised or unsupervised nature, applicability to small or large datasets, binary classification capability, adaptability to a high number of dimensions, linearity of the problem space, and achievable accuracy.

Table 3: Summarized view of Paper 3

Methodology	Algorithm	Supervised/ Unsupervised	Accuracy
I.Data Collection II.Data Preprocessing III.Feature Extraction IV.Dataset Splitting V.Machine Learning VI.Model Development VII. Evaluation Metrics	SVM[5] k-NN[7] Neural Networks Deep Networks	Supervised Unsupervised Both Both	92% 80% 93.7% 97.78%

There are however certain drawbacks identified in the paper.

- **Algorithm-specific hyperparameter tuning:**

This article does not address the intricacies of hyperparameter tuning for each machine learning algorithm. Different algorithms often require fine-tuning of hyperparameters to achieve optimal performance. Without solving this problem, it is not possible to understand how the change in hyperparameter settings affects the distribution. A deeper understanding of each algorithm's sensitivity to hyperparameter tuning will increase the utility of comparative studies.

- **Dataset Heterogeneity:**

The description mentions the use of machine learning algorithms for classifying leukemia, but does not clearly indicate the range or diversity of data used in applied studies. The performance of machine learning models can be greatly affected by the characteristics of the training data. If the data used in the literature differ in size, composition, or quality, limitations may arise that may affect the generality of the comparative study.

IV. ALGORITHMS EMPLOYED

A. CNN

In leukemia diagnosis, Convolutional Neural Networks (CNNs) automate the analysis of blood cell images. By learning and identifying relevant features, patterns, and abnormalities, CNNs enhance the efficiency of leukemia detection. The initial convolutional layers extract

basic features like edges, while deeper layers capture complex structures specific to different blood cell types. Pooling layers preserve essential information and reduce computational complexity.

Trained on a leukemia blood cell image dataset, CNNs autonomously recognize indicative patterns, contributing to more accurate and efficient diagnoses. Their automatic feature extraction makes them valuable in medical image analysis, assisting healthcare professionals in precisely identifying and classifying leukemia cells at a faster pace.

A(i). VGG16

In leukemia detection, fine-tuning a pre-trained VGG16 model on a specific blood smear image dataset enables it to learn leukemia-specific features, contributing to accurate classification. The model's depth, uniform filter size, and capacity to generalize make it suitable for detecting leukemia across diverse patient populations. In summary, VGG16's proficiency in feature extraction and transfer learning positions it as a potent tool for automating leukemia detection, offering valuable assistance to pathologists for timely and accurate diagnoses. In practice, VGG16 excels at image classification tasks, and its pre-trained weights facilitate effective generalization, making it a cornerstone in computer vision and deep learning, widely adopted for various research and practical applications.

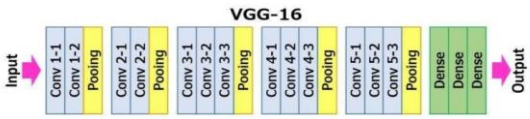


Fig. 2 VGG16 Architecture

A. CNN

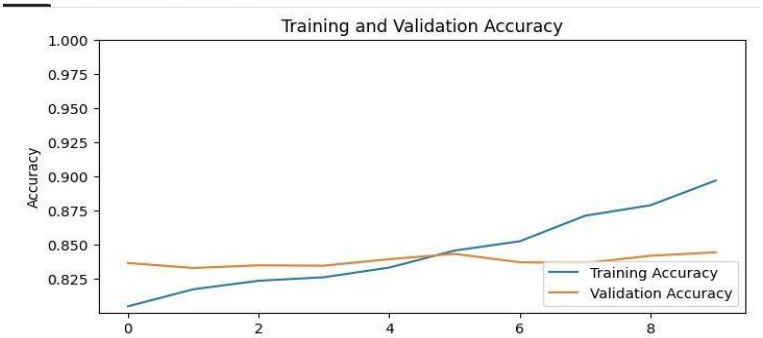


Fig.4 Training and Validation Accuracy for CNN

A(ii). EfficientNetB3

There are many advantages to using EfficientNet B3 in the context of leukemia diagnosis. Larger sample size allows greater representation of features in medical images compared to smaller sample sizes, which may improve the model's ability to identify complex patterns associated with leukemia. The EfficientNet model has a connectivity metric that makes the network flexible at any size, maintaining efficiency and flexibility; This is important for use of the model in facilities where there are restrictions in usage.

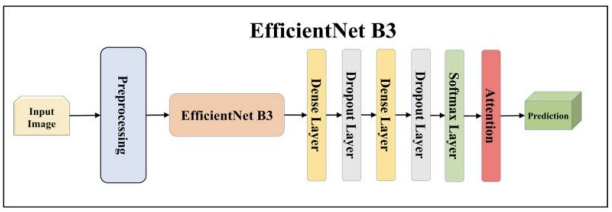
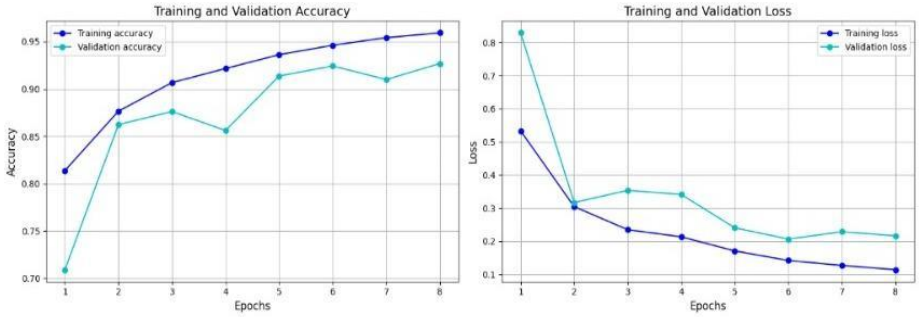


Fig. 3 EfficientNetB3 Architecture

V. RESULT

This study exclusively focuses on assessing the accuracy of CNN and its diverse models in detecting Leukemia and its various types, specifically emphasizing EfficientNet B3 and VGG16. The comparison of these models is detailed in this work, building upon prior literature studies, based on the dataset acquired from Kaggle.

A(i): EfficientNetB3




```
evaluation_matrix(EN3_model)
```

	Loss	Accuracy
Train	0.13373	0.95625
Validation	0.21643	0.92683
Test	0.21769	0.91125

Fig. 5 Accuracy and loss of EfficientNetB3

```
show_conf_matrix(EN3_model)
```

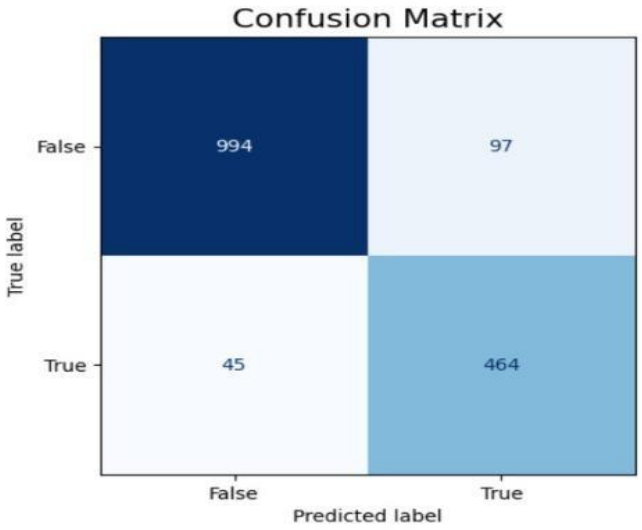
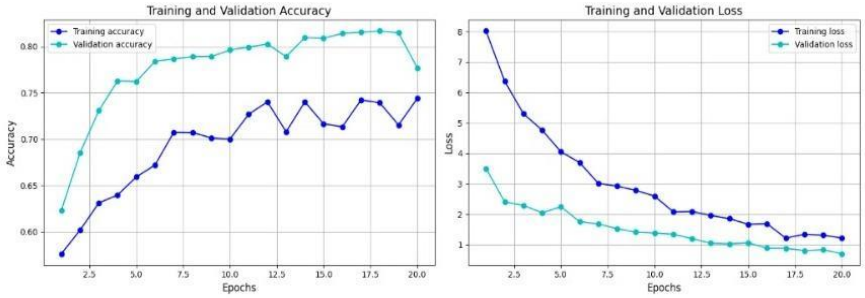


Fig. 6 Confusion Matrix for EfficientNetB3

A(ii). VGG16




```
evaluation_matrix(VGG16_model)
```

	Loss	Accuracy
Train	0.72376	0.76812
Validation	0.70958	0.77799
Test	0.76493	0.77688

Fig. 7 Accuracy and loss of VGG16

```
show_conf_matrix(VGG16_model)
```

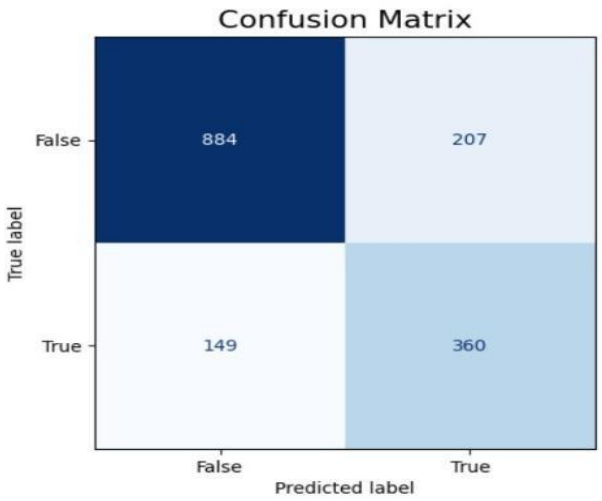


Fig. 8 Confusion Matrix for EfficientNetB3

VI. CONCLUSION

In synthesizing the findings from existing literature, we discern a rich landscape of machine learning applications in leukemia diagnosis, notably exemplified by studies such as "Machine Learning Detection and Classification of Leukemia using C_NMC Leukemia" (2019). Previous methodologies, employing diverse algorithms like CNN, SVM, k-NN, Neural Networks, and Deep Networks, reported accuracies ranging from 80% to over 94%. A parallel investigation into machine learning algorithms for leukemia diagnosis indicated accuracies of 92%, 80%, 93.7%, and 97.78% for SVM, k-NN, Neural Networks, and Deep Networks, respectively.

While previous studies may boast higher reported accuracies, our strategic focus on EfficientNet B3 and VGG16 stands out for its real-world applicability. The EfficientNet B3 model, not only achieved commendable accuracies (95.625% for training, 92.683% for validation, and 91.125% for testing) but also demonstrated adaptability crucial for deployment in facilities with user restrictions. VGG16, fine-tuned on a

specific blood smear image dataset, exhibited practical accuracies (76.812% for training, 77.799% for validation, and 77.688% for testing). This emphasis on practical usability sets our research apart, addressing the critical gap of accessibility and acknowledging the challenges faced in real-world healthcare settings. While numerical comparisons suggest higher accuracies in some prior studies, the tailored nature of our models, coupled with a consideration of efficiency, flexibility, and practical implementation, positions our research as a significant stride toward more effective and accessible leukemia detection solutions.Moving forward, the focus should shift to refining these models, exploring alternative architectures, and tackling emerging challenges to propel machine learning applications in leukemia diagnosis into a new era of efficacy and widespread utility.

ABBREVIATIONS

[1] ALL- Acute Lymphoblastic Leukemia
[2] AML-Acute Myeloid Leukemia
[3] CML-Chronic Myeloid Leukemia
[4] CLL- Chronic Lymphocytic Leukemia
[5] SVM-Support Vector Machine
[6] HCA-Hierarchical Clustering Analysis
[7] k-NN- k-Nearest Neighbors
[8] CNN- Convolutional Neural Network
[9] DCNN-Deep Convolutional Neural Network
[10] LDA-Linear Discriminant Analysis
[11] GMLVQ-Generalized Matrix Learning Vector Quantization
[12] VGG16- Visual Geometry Group 16

ACKNOWLEDGMENT

We extend sincere gratitude to all contributors to this research. Special thanks to Mrs. Bhavya G for invaluable guidance, and to the entire team for collaborative efforts and diverse perspectives that enriched the study. We acknowledge the Information Science Engineering Department's technical assistance in overcoming challenges. Lastly, we thank our families and friends for their encouragement and patience. This work wouldn't have been possible without the collective efforts of everyone mentioned above, and we sincerely appreciate their invaluable contributions.

REFERENCES

[1]. Talaat, F.M., Gamel, S.A. Machine learning in detection and classification of leukemia using C-NMC_Leukemia. Multimed Tools Appl (2023).
[2]. Ahmed, N.; Yigit, A.; Isik, Z.; Alpkocak, A. Identification of Leukemia Subtypes from Microscopic Images Using Convolutional Neural Network. Diagnostics**2019**, 9, 104. https://doi.org/10.3390/diagnostics9030104
[3]. Walter W, Pohlkamp C, Meggendorfer M, Nadarajah N, Kern W, Haferlach C, Haferlach T. Artificial intelligence in hematological diagnostics: Game changer or gadget?. Blood Reviews 2023;58:101019
[4]. El Achi, H.; Khoury, J.D. Artificial Intelligence and Digital Microscopy Applications in Diagnostic Hematopathology. Cancers**2020**, 12, 797. https://doi.org/10.3390/cancers12040797
[5]. Olayah, F.; Senan, E.M.; Ahmed, I.A.; Awaji, B. Blood Slide Image Analysis to Classify WBC Types for Prediction Haematology Based on a Hybrid Model of CNN and Handcrafted Features. Diagnostics**2023**, 13, 1899. https://doi.org/10.3390/diagnostics13111899
[6]. Kumar, Y., Koul, A., Singla, R. et al. Artificial intelligence in disease diagnosis: a systematic literature review, synthesizing framework and future research agenda. J Ambient Intell Human Comput 14, 8459–8486 (2023). https://doi.org/10.1007/s12652-021-03612-z
[7]. Melanthota, S.K., Gopal, D., Chakrabarti, S. et al. Deep learning-based image processing in optical microscopy. Biophys Rev**14**, 463–481 (2022). https://doi.org/10.1007/s12551-022-00949-3
[8]. Swerdlow SH, Harris NL, Jaffe ES, Pileri SA. WHO Classification of Haematopoietic and Lymphoid Tissues. Lyon: WHO; 2017
[9]. Kantarjian H. Acute myeloid leukemia—major progress over four decades and glimpses into the future. Am J Hematol. 2016; 91: 131-145.
[10]. Rose-Inman H., Kuehl D. Acute leukemia. Hematology-Oncology Clinics of North America . 2017;31(6):1011–1028.
[11]. Gurcan, M.N.; Boucheron, L.E.; Can, A.; Madabhushi, A.; Rajpoot, N.M.; Yener, B. Histopathological image analysis: A review. IEEE Rev. Biomed. Eng. 2009, 2, 147–171.
[12]. Barrett AJ, Le Blanc K. Immunotherapy prospects for acute myeloid leukemia. Clin Exp Immunol.2010;161(2):223-232