Indonesian Journal of Computer Science
ISSN 2549-7286 (*online*)
Jln. Khatib Sulaiman Dalam No. 1, Padang, Indonesia
Website: ijcs.stmikindonesia.ac.id | E-mail: ijcs@stmikindonesia.ac.id

Leukemia Detection and Classification Based on Machine Learning and CNN: A Review

Hakar Hasan Rasheed[1,2], Adnan Mohsin Abdulazeez[3]
hakar.hassan@gmail.com[1], hakar.hasan@auas.edu.krd[2], adnan.mohsin@dpu.edu.krd[3]
[1] Ararat Private Technical Institute, Duhok, Kurdistan Region, Iraq.
[2] Akre University for Applied Sciences/ Technical College of Informatics -Akre/ Department of Information Technology.
[3] Duhok Polytechnic University, Duhok, Kurdistan Region, Iraq.

Article Information

Submitted : 19 May 2024
Reviewed : 28 May 2024
Accepted : 15 Jun 2024

Keywords

DataMining, Classification Algorithms, Leukemia.

Abstract

Advancements in data mining methods have significantly improved disease diagnosis, particularly in the realm of leukemia detection. Leukemia, a complex cancer affecting white blood cells, poses significant challenges in diagnosis and management due to its diverse manifestations. Various machine learning algorithms, including Convolutional Neural Networks (CNNs), Support Vector Machines (SVMs), Random Forests (RF), Decision Trees (DTs), K-Nearest Neighbors (K-NN), Logistic regression (LR) and Naïve Bayes (NB) classifiers, have been employed to accurately classify leukemia cases based on diverse datasets and image analyses. This paper provides a comprehensive overview and comparison of these classification techniques, highlighting their effectiveness in diagnosing different leukemia subtypes. Additionally, the paper discusses the methodology and findings of several studies focusing on leukemia detection, emphasizing the significance of machine learning in enhancing diagnostic accuracy and treatment planning. Furthermore, it explores the challenges and future directions in leveraging machine learning for leukemia diagnosis, including the need for standardized datasets, algorithm refinement, and integration with clinical data for personalized treatment strategies.

A. Introduction

Advancements in data mining methods for diagnosis have shown significant improvement, particularly in disease diagnosis, employing a range of classification techniques like decision trees and rule classifiers. These techniques not only yield accurate conclusions but also aid in visualizing patterns within datasets. However, there isn't a singular classifier that outperforms others universally; factors such as the classification method, gene selection method, and datasets greatly influence classification accuracy, according to the World Health Organization (WHO), cancer was responsible for 8.2 million deaths in 2012, constituting 13% of total global mortality, with a projected 70% surge in new cancer cases over the next two decades [1], [2].

This emphasizes the critical need for distinct diagnostic and treatment strategies across over 100 types of cancer. The analysis of blood components—red blood cells (RBC), white blood cells (WBC) [3], and platelets—through microscopic blood smear tests is pivotal in diagnosing various blood diseases, given the integral role of WBCs in the immune system [32]. WBCs, categorized into lymphocytes, monocytes, neutrophils, and eosinophil as shown in fig 1. provide essential insights into an individual's health status, aiding in the diagnosis of conditions like leukemia [4]–[6]. Leukemia itself is classified into four primary types based on the infected cells and the disease's severity: Acute Lymphoblastic Leukemia (ALL), Acute Myeloid Leukemia (AML) [7], [8]. Chronic Myeloid Leukemia (CML), and Chronic Lymphocytic Leukemia (CLL), highlighting the diversity in blood cancer manifestations and the complexity of their management [34], [35].

a) Acute Lymphoblastic Leukemia (ALL), a type of white blood cell cancer characterized by the rapid growth of immature WBCs in the bone marrow, predominantly affects children. Diagnosing ALL is challenging due to its symptoms (fatigue, weakness, joint, and bone pain) closely mimicking viral fevers. ALL is classified into three subtypes: L1, L2, and L3. Diagnosis involves a combination of biopsies [9], imaging (X-ray, CT scan, MRI, Ultrasound), and blood chemistry analysis to assess the leukemia type, disease spread, and tumor growth rate [37], [38].

b) Acute Myeloid Leukemia (AML) represents the most frequent type of acute leukemia, characterized by the production of not only abnormal WBCs but, in some cases, also red blood cells (RBCs) and platelets. Symptoms of AML can vary widely, from fever and bone pain to unusual fatigue and spontaneous bleeding, thereby necessitating a differential diagnosis based on its eight recognized subtypes [10], [39]. On the other hand.

c) Chronic Myeloid Leukemia (CML) is a blood and bone marrow disease predominantly diagnosed in adults, with its onset rare in children. Initial diagnosis involves microscopic bone marrow analysis, while confirmation of CML relies on examining a peripheral blood smear microscopically [41]. d) Chronic Lymphocytic Leukemia (CLL) the most common form of leukemia in adults in the Western world, has witnessed a significant evolution in treatment protocols due to the introduction of novel biological agents and the discovery of molecular predictive markers. Notably [11], certain lifestyle and environmental factors, such as agricultural work, exposure to Agent Orange, and possibly ionizing radiation, have been associated with an increased risk of developing CLL, though the direct link to

radiation remains unconfirmed [12], [13]. The ongoing research and development in leukemia treatment highlight the complexity of this disease and underscore the importance of advancing our understanding and therapeutic approaches to manage its various forms more effectively. this abnormal blood cell proliferation disrupts normal bodily functions [14], leading to severe health complications, including anemia, excessive bleeding, infection, and the potential for pain or swelling to spread to lymph nodes and other organs. Leukemia, a cancer that predominantly targets white blood cells (WBCs) [15], plays a critical role in the human body's defense against infections and diseases. This malignancy interferes with the bone marrow's ability to produce healthy blood cells, instead generating large numbers of immature WBCs. Such disruption not only weakens the immune system but also has broader implications for a person's overall health. Leukemia manifests in two primary forms: acute and chronic. Acute leukemia is marked by a swift proliferation of abnormal WBCs, complicating their differentiation from healthy cells.

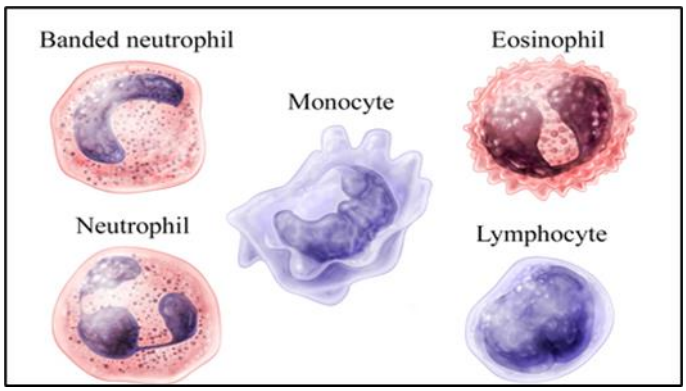

Figure 1. Types of White Blood Cells (WBCs)

The rest of this paper is organized to describe the background theory of deep learning and the algorithms to detect Leukemia in **section A**. After that, in **section B**, different research Classification algorithm on using CNN and K-NN, SVM, RF, DT and Naive Bayes to detect leukemia are described and compared in Table 1, Table 2, Table 3 and Table 4. Finally, **section C** describes a Result And Discussion, a conclusion and limitations, and a list of references are shown.

B. Research Methodology

Classification techniques represent a cornerstone in the accurate allocation of extracted data across distinct categories [16]. This process hinges on the adept training of classification algorithms using designated training sets, enabling them to discern intricate patterns within the data distribution. Once trained, these classifier models leverage the discerned characteristics of incoming data to assign them to appropriate classes. Within the realm of digital pathology, this classification endeavor undertakes the critical task of distinguishing between cancerous and healthy tissue samples [17]. To accomplish this, a comprehensive study has employed an array of seven distinct machine learning algorithms. Firstly, the Naïve Bayes algorithm [18], grounded in probabilistic reasoning, constructs a probability set by scrutinizing the frequency and combinations of values within datasets. Despite assuming independence among variables, Naïve Bayes swiftly

learns within controlled classification contexts. Additionally, the K-Nearest Neighbors (K-NN) method, a supervised learning model, predicts dataset classes based on the proximity of neighboring data points. By assessing similarities through diverse distance metrics like Euclidean distance, K-NN efficiently classifies data into appropriate categories. Furthermore [19], [20], Decision Tree (DT) methodology, characterized by its intuitive flowchart-like structure, proves invaluable for pattern classification. With internal nodes representing attribute tests and leaf nodes carrying class labels, decision trees employ various algorithms such as ID3 for feature threshold pair optimization, enhancing classification accuracy. Random Forest (RF) [21], [22], an ensemble learning technique, aggregates multiple classification trees to yield remarkably accurate results, particularly beneficial for large datasets with substantial input variables and missing data. Moreover, Support Vector Machine (SVM) models excel in creating hyperplanes within feature spaces to discern class boundaries, minimizing generalization errors. By strategically selecting hyperplanes with maximal margins, SVM optimizes classification accuracy while identifying crucial support vectors along class borders.

Finally, the advent of deep learning has ushered in revolutionary capabilities, particularly in image analysis and disease diagnosis. Convolutional Neural Networks (CNNs) [23], renowned for their prowess in computer vision tasks, including image recognition and segmentation, have been instrumental in leukemia detection. Leveraging transfer learning and pretrained models like AlexNet and ResNet [24], CNNs exhibit exceptional performance in automating feature extraction and achieving unparalleled classification accuracy. In essence, the convergence of diverse classification techniques, ranging from traditional algorithms like Naïve Bayes and Decision Trees to cutting-edge approaches like CNNs [25], [26], empowers researchers and healthcare professionals with robust tools for accurate leukemia diagnosis and treatment planning [27]. These methodologies not only facilitate early disease detection but also pave the way for personalized therapeutic interventions, thereby enhancing patient outcomes and advancing our understanding of hematologic malignancies. Data mining algorithms and classification algorithm as shown in figure 2.

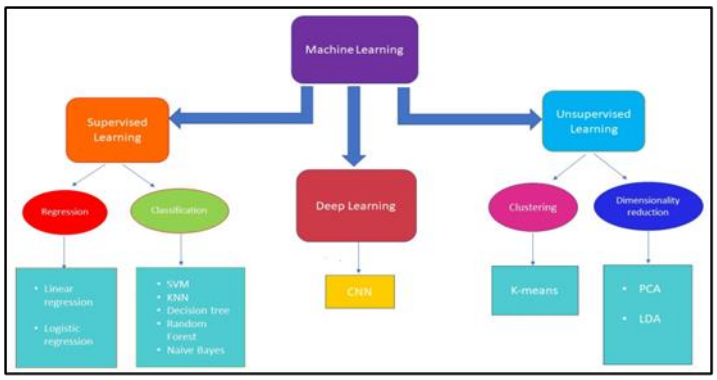

Figure 2. Classification techniques

A. Convolutional Neural Networks (CNN) and SVM, K-NN, RF, DT, Naïve Bayes for Detecting Leukemia:

In leukemia detection, various machine learning algorithms play pivotal roles in accurately classifying patient data. Convolutional Neural Networks (CNNs) excel in analyzing blood smear images, identifying abnormal cell morphology [28]. Support Vector Machines (SVMs) efficiently separate leukemia cases by finding optimal hyperplanes based on distinct features. Random Forests (RF) utilize decision tree ensembles to robustly classify leukemia instances, effectively managing high-dimensional data. Decision Trees (DTs) offer interpretable rules for leukemia diagnosis using patient attributes. K-Nearest Neighbors (K-NN) measure similarity between patient data points [29], aiding classification in diverse datasets. Naive Bayes classifiers efficiently handle large-scale data [30], utilizing probabilistic principles for classification. These algorithms collectively contribute to the precise and efficient detection of leukemia [31], [32], offering diverse approaches tailored to different types of patient data [33], and diagnostic needs.

A comparative analysis was conducted on machine learning techniques for the automatic classification of Acute Lymphoblastic Leukemia (ALL) cells. Utilizing the Acute Lymphoblastic Challenge Database (ALL-CDB), which contained 6500 digital microscopic pathology images from 118 subjects, geometric features were extracted and subjected to Principal Component Analysis (PCA) for feature selection. Seven machine learning methods, including Naive Bayes, k-Nearest Neighbor (k-NN), Linear Discriminant Analysis (LDA), Decision Tree (DT), Random Forest (RF), Support Vector Machine (SVM), and Multilayer Perceptron (MLP), were applied for classification. The MLP neural network exhibited the highest accuracy and F1-score of 97% for ALL cell classification. The study highlighted the challenges in distinguishing cancerous cells from healthy ones and emphasized the role of computer-aided systems in pathology, albeit with limitations in providing definitive diagnoses [34].

Employed deep learning and machine learning techniques to diagnose acute lymphoblastic leukemia (ALL) from microscopic images, using custom-tuned ResNet-50, VGG-16 networks, and a newly proposed convolutional network. With validation accuracies reaching up to 84.62%, the research highlights the efficacy of these computational approaches, particularly the proposed convolutional network, in accurately and efficiently diagnosing ALL, offering a promising tool for clinical application [35].

Deep learning models were utilized for accurate leukemia classification, focusing on categorizing leukemia subtypes through deep convolutional neural networks (CNN). The dataset included images from the ASH bank and ALL-IDB. By employing deep CNN-based techniques, precise classification of leukemia subtypes was achieved, with data augmentation and transfer learning enhancing model performance. Effective features for diagnosis were selected using machine learning techniques like SVM, KNN, GA, MLP, and Random Forest algorithms. The proposed model attained a 95.59% test accuracy in leukemic B-lymphoblast classification, surpassing individual network performances. This approach showcases the effectiveness of deep learning in enhancing diagnostic accuracy for leukemia subtyping [36].

Presented for diagnosing acute lymphoblastic leukemia (ALL) from microscopic blood smear images, employing quadratic discriminant analysis (QDA) to classify them as malignant or normal. The method involved efficient feature extraction aided by morphological processing, with QDA utilized for classification. Comparative analysis against widely used classifiers such as Support Vector Machine (SVM), k-Nearest Neighbor (K-NN), and Decision Tree (DT) revealed QDA's superior performance, achieving an accuracy of 94.4% compared to SVM (92.6%), K-NN (92.6%), and DT (91.7%). The study conducted experiments on 108 microscopic blood smear images from the ALL-IDB1 dataset, employing QDA, which operated on multivariate normal distribution and Bayesian discrimination principles. Image processing stages included histogram-based enhancement and thresholding-based segmentation. Notably, geometric features were extracted differently, focusing on mean values of connected components in segmented lymphoblasts rather than separate segmentation of cytoplasm and nucleus. Results underscored QDA's effectiveness for ALL diagnosis, demonstrating its potential superiority over other classifiers [37].

Analyzed over 4,000 patient data from the European Gaza Hospital to predict leukemia using data mining techniques, focusing on the relationship between blood properties, demographic factors, and leukemia. It utilized k-nearest neighbor (k-NN), decision tree (DT), and support vector machine (SVM) algorithms for classification, finding the DT algorithm most accurate at 77.30%. However, SVM showed competitive performance with a 76.82% accuracy and the highest F-Measure at 70%. External factors, such as residing in eastern regions possibly due to agricultural chemicals or conflict effects, were identified as increasing leukemia risk. The findings suggest further research with additional classifiers like Naive Bayes and neural networks to explore leukemia causes in high-risk areas, highlighting the variability and need for precise algorithm selection in medical diagnostics [38].

Compared various machine learning algorithms—Support Vector Machines, k-Nearest Neighbor, Neural Networks, Naïve Bayes, and Deep Learning—for leukemia classification. It evaluated their performance using a dataset that was divided into training (80%) and testing (20%) sets, consisting of images representing different leukemia subtypes and normal cases. A Convolutional Neural Network (CNN) model achieved a significant accuracy of 97.78% with a specific learning rate and number of epochs, highlighting the effectiveness of machine learning in diagnosing leukemia and assessing treatment impacts [39].

Developed a streamlined approach for diagnosing Acute Lymphoblastic Leukemia (ALL), focusing on two geometric features of ALL cells: the presence of cytoplasmic vacuoles and nucleus membrane regularity. These criteria were selected for their diagnostic accuracy across various classification algorithms. Evaluations using ROC curves and F1 scores revealed the method's effectiveness, despite its simplicity. Notably, the quadratic kernel SVM and ANN classifiers emerged as the most efficient, both achieving perfect F1 scores of 100%. This indicates the potential of this approach as an effective diagnostic aid for pathologists in identifying ALL subtypes [40].

Introduced a methodology that utilized an Optimized Convolutional Neural Network (OCNN) for classifying medical images into "normal" or "abnormal"

categories. It incorporated three steps: Image Preprocessing, Feature Extraction, and Classification. Fuzzy logic was applied to fine-tune the CNN's hyperparameters, significantly boosting its effectiveness. This approach led to the OCNN achieving an exceptional 99.99% accuracy rate on the C-NMC_Leukemia dataset, demonstrating the efficacy of fuzzy logic in enhancing the performance of CNN classifiers in medical image analysis [41].

Proposed a leukemia detection method, utilizing the Gini index–based Fuzzy Naive Bayes (GFNB) classifier, which integrates the Gini index and Fuzzy Naive Bayes classifier. Initially, the input multi-cell blood smear image underwent pre-processing, and blast cells were segmented using adaptive thresholding. Subsequently, the proposed classifier counted blast cells to determine the presence of leukemia for effective diagnosis. Experimental analysis using the ALL-IDB1 database confirmed that the proposed method outperformed existing methods in terms of accuracy, specificity, and sensitivity, which were found to be 0.9591, 0.9599, and 1, respectively. The experimental results demonstrated the reliability and accuracy of the proposed method, suggesting its potential to aid physicians in improving and expediting their diagnostic processes [42].

Focused on detecting leukemia using Machine Learning algorithms for medical diagnosis, specifically comparing the effectiveness of Decision Tree, Logistic Regression, and Random Forest algorithms. It was found that the Decision Tree algorithm achieved an 84.15% accuracy rate in predicting leukemia, with S. Tryglycerides as the root node, and was highlighted as a key algorithm along with Logistic Regression. Precision, Recall, and F1 Score metrics were employed for the evaluation of these algorithms. However, the paper was noted to lack discussion on the implications of false positives. The dataset used for this analysis was collected from Z H Shikder College and Hospital [43].

The study explored leukemia diagnosis using Machine Learning (ML) and Deep Learning (DL) approaches to enhance prediction accuracy, highlighting how ML models can identify compounds for clinical trials and tailor treatment plans. Three ML methods were applied, assessing their effectiveness through Recall, Precision, and Accuracy metrics, with SVM, RBFNN, and optimized methods found to be particularly suited for leukemia datasets. The research also delved into data mining to discover patterns and trends, alongside data cleaning and transformation processes to ensure data quality and suitability for analysis. It emphasized ML's role in optimizing hospital operations and supporting clinical decision-making, noting kernel selection as a significant constraint in SVM models. Despite these insights, the study did not specify any direct limitations within the given contexts [44].

Machine learning techniques were employed to assess Chronic Myeloid Leukemia (CML) progression. Various algorithms such as Support Vector Machines, Nearest Neighbor, Naive Bayes, and Deep Learning were utilized for classification and forecasting of CML phases, considering the role of the BCR-ABL1 fusion protein. A Sequential Forward Selection (SFS) approach was adopted to compare accuracy, with Support Vector Machines, K-Nearest Neighbors, Neural Networks, Naive Bayes, and Deep Learning algorithms evaluated. Results indicated high accuracy, with RF SFS, KNN SFS, SVC RBF, and SVC SFS achieving up to 97.66%. These techniques effectively forecasted CML disease evolution [45].

Provided for detecting acute lymphoblastic leukemia from microscopic views of white blood cells. Microscopic images underwent thorough pre-processing before classification. White blood cells were separated using morphological techniques, and their segmented regions were analyzed for textural, geometrical, and statistical properties. Four machine learning techniques, including random forest (RF), support vector machine (SVM), naive Bayes classifier (NB), and K nearest neighbor (KNN), were employed to assess algorithm performance. SVM proved effective in identifying leukemia malignancy, leading to consideration of EMC-SVM for leukocyte classification due to the diversity of blood smear images. The proposed method successfully distinguished white blood cells and accurately categorized each segmented cell, aiming to support medical activity in acute lymphocytic leukemia (ALL) detection. However, the study emphasized the need to increase the dataset size to enhance the classification model's performance and explore alternative validation methods beyond 10-fold cross-validation [46].

Utilized machine learning with classifiers to detect leukemia types, aiming to streamline diagnosis processes for both patients and physicians. The primary objective was to determine the most effective methods for leukemia detection. The WEKA application was employed to evaluate and analyze five classifiers (J48, KNN, SVM, Random Forest, and Naïve Bayes classifiers), revealing varied accuracies ranging from 83.33% to 98.61%, with the Naïve Bayes classifier achieving the highest accuracy. However, accuracy depended on factors such as sample shape, size, and the classification algorithm utilized. The study focused on improving leukemia detection accuracy through machine learning with classification algorithms, utilizing the Microarray Leukemia dataset and achieving notable accuracies across classifiers. Future research aims to assess the effectiveness of treatment provided to leukemia patients by employing appropriate machine learning classification algorithms for all leukemia types, potentially enabling parallel execution for improved response time and accuracy [47].

Table 1. Overview of the literature on detecting Leukemia based on CNN, SVM, DT, RF, K-NN, Naïve Bayes.

Authors	Year	Dataset	Algorithm	Limitations	Results
KOCATÜRK, CANDEMİR, and KOCABAŞ **[34]**	2022	Acute Lymphoblastic Challange Database (ALL-CDB) from Cancer Imaging Archive.,ASH Image Bank dataset for classifying leukemia subtypes.	Naive Bayes, K-NN, DT, RF, SVM	The study does not compare the performance with other leukemia types	Naive Bayes accuracy: 72% K-NN accuracy: 89% RF accuracy: 91% DT accuracy: 82% SVM accuracy: 73%
Rezayi et al. **[35]**	2021	Dataset from a CodaLab competition to classify leukemic cells.	CNN, RF, K-NN, SVM, LR	Feature engineering weakness in machine learning is a limitation. Possibility of errors in medical image interpretation is a limitation.	CNN accuracy: 85.79% RF accuracy: 81.72% K-NN accuracy: 77.89% SVM accuracy: 79.28% LR accuracy: 79.88%
K. SRIDHAR et al. **[36]**	2022	ALL-IDB1, ALL-IDB2	CNN, SVM, K-NN	Manual selection of crucial features is challenging in	CNN accuracy: 95.59% SVM accuracy: 89.82% K-NN accuracy 96.26%

				leukemia diagnosis.	
Chand and Vishwakarma **[37]**	2022	ALL-IDB1	QDA, SVM, K-NN, DT	The lack of comparison with more advanced machine learning algorithms.	QDA accuracy: 94.4% SVM accuracy: 92.6% K-NN accuracy: 92.6% DT accuracy 91.7%
Daqqa, Maghari, and Al Sarraj **[38]**	2020	Blood test data from European Gaza Hospital at Gaza Strip.	DT, K-NN, SVM	Noise data and boundary value data in medical disease data.	DT highest accuracy of 77.30%. K-NN accuracy of 72.15% SVM accuracy of 76.82%
Patil Babaso, Mishra, and Junnarkar **[39]**	2020	Blood cell images dataset for classification of leukemia subtypes.	SVM, K-NN, Naïve Bayes, CNN	Limited to classification of leukemia types using machine learning algorithms.	SVM accuracy: 92% K-NN accuracy: 80% Naïve Bayes: 80.88% CNN accuracy: 97.78%
Al-Tahhan et al. **[40]**	2020	Original ALL-IDB2 image database with 130 abnormal cell images.	K-NN, SVM	Existing approaches require a large number of features for diagnosis Few studies available with algorithms based on few features	K-NN accuracy of 97.4% SVM accuracy of 97.4%
Talaat and Gamel **[41]**	2024	C-NMC, Leukemia dataset used for classification model development. C-NMC.	CNN, K-NN, SVM, DT	Limited discussion on the impact of computational resources.	CNN accuracy: 99.99% accuracy with K-NN accuracy of 99.64% SVM accuracy of 99.93% DT accuracy of 95.69%
B. K. Das and Dutta **[42]**	2020	ALL-IDB1	Naive Bayes, CNN, K-NN, SVM,	The paper does not provide a comparison of the proposed method	Naïve Bayes accuracy: 0.9002 CNN accuracy: 0.9154 K-NN: 0.88 SVM accuracy: 0.7652
Mou et al. **[43]**	2020	Dataset collected from Z H Shikder College and Hospital.	DT, RF, LR	representativeness of the dataset used for analysis	DT accuracy 84.15%
Sasirekha, Sivaraman, and Sumitha **[44]**	2023	'Golub Leukemia Dataset' contains gene expression profiles of leukemia patients	SVM, RBFNN, SVM-RBFNN	Kernel selection is a primary constraint in SVM models.	SVM accuracy: 80%
Pandey and Pal **[45]**	2021	UCI machine learning repository Wisconsin dataset for training and testing.	SVM, K-NN, RF, Naive Bayes	Neural Networks have high computational overhead and substantial overfitting issues.	SVM accuracy: 97.66% K-NN accuracy: 97.66% RF accuracy: 97.66% Naive Bayes accuracy: 80.88%
More and Sugandhi **[46]**	2023	Dataset includes 1208 photos	SVM, K-NN, RF, Naive Bayes	To overcome this limitation, the paper proposes the use of a multi-class SVM classifier,	SVM accuracy: 98.85% K-NN accuracy: 97.10% RF accuracy: 97.60% Naive Bayes accuracy:

				which outperforms conventional techniques for leukocyte classification.	98.08%
Armya et al. **[47]**	2021	Clinically Collected Dataset, Microarray ALL, TCIA dataset used for analysis.	SVM, K-NN, RF, Naive Bayes	The paper lacks a specific section addressing limitations.	SVM accuracy: 95.83% K-NN accuracy: 87.5% RF accuracy: 88.88% Naive Bayes accuracy: 98.61%

B. K-Nearest Neighbor (K-NN), RF, Naïve Bayes for Detecting Leukemia:

In leukemia detection, K-Nearest Neighbors (K-NN) assesses patient data similarity, Random Forests (RF) use decision tree ensembles for robust classification, and Naïve Bayes assumes feature independence to efficiently categorize leukemia cases based on probabilistic principles.

Explored the use of thresholding-based segmentation and Hu moment feature extraction in the classification of Acute Lymphoblastic Leukemia (ALL) using the Gaussian Naive Bayes algorithm, has provided insightful results. It found varied performance across different folds, with an average precision of approximately 84.13% and fluctuating accuracy, recall, and F1-scores, indicating both potential and limitations of the approach. Despite consistently high precision in correctly identifying malignant cases, there was a need for further refinement to improve consistency and generalizability. These findings contribute to the field of medical image analysis, particularly in hematological malignancy classification, emphasizing the importance of integrating image processing and machine learning techniques in medical diagnostics [48].

Focused on using machine learning techniques to classify leukemic blood cells for aiding in the early detection of leukemia, given its initially ambiguous symptoms. The study introduced a method for predictive leukemia detection by extracting vital features from blood tests and employing various classifiers, with the AdaBoostM1 algorithm demonstrating superior performance over the Bayes Net classifier. Key factors such as age, infection status, and blood cell count significantly influenced leukemia detection. Utilizing the Random Forest classifier, the model achieved impressive accuracy metrics of 98.50%, sensitivity of 96.99%, specificity of 98.7%, and precision of 98.30%. The study employed a range of classifiers including Bayes net, naïve Bayes, decision stump, hoeffding tree, j48, LMT, random forest, random tree, and RepTree classifiers using WEKA, with the proposed Bayes Net classifier showing the highest accuracy in cross-validation, while Random Forest outperformed other tree classifiers, demonstrating exceptional performance in leukemia detection [49].

The scheme classified normal versus affected white blood cells in images, utilizing Discrete Orthogonal Stockwell Transform (DOST) for texture features and the Adaboost algorithm for classification. It achieved a superior accuracy of 99.66% compared to existing schemesThe process involved pre-processing, sub-imaging, feature extraction, feature reduction, and classification. The proposed method utilized the Adaboost algorithm with a Random Forest (RF) classifier for cell classification, incorporating features such as DOST-based texture extraction and Principal Component Analysis with Linear Discriminant Analysis (PCALDA)-

based dimensionality reduction. Simulation results demonstrated high accuracy through k-fold stratified cross-validation [50].

Developed a machine learning model that improved the diagnosis of Acute Myeloid Leukemia (AML) by accurately detecting and classifying immature leukocytes. Using a dataset from The Cancer Imaging Archive, the model applied image processing techniques and extracted 16 features, including two innovative nucleus color features. Trained with a random forest algorithm, it achieved detection and classification accuracies of 92.99% and 93.45%, respectively, outperforming existing methods. The model identified key features that significantly contributed to its accuracy, showcasing its potential as an AML diagnostic aid and laying groundwork for future investigations [51].

Prediction techniques were widely utilized in medical diagnosis, acknowledging the inherent uncertainty in medical science where no method was 100% accurate. Leukemia, impacting blood status, was identifiable through the Blood Cell Counter (CBC), although the accuracy of various classification algorithms varied. The classification technique presented aimed to assist doctors in confirming Leukemia diagnoses, recognizing the probabilistic nature of disease diagnosis. The Naïve Bayes algorithm demonstrated the highest percentage of accuracy at 95% compared to other techniques. Leveraging machine learning and big data models, the system's accuracy improved with more data. Despite this, opportunities for enhancement and the implementation of new methodologies within the proposed approach were explored. Future endeavors included the implementation of artificial intelligence supervised learning for classification [52].

Focused on utilizing the Random Forest (RF) classifier for recognizing Acute Lymphoblastic Leukemia (ALL) and lymphocyte cell subtypes. It introduced an automatic system for detecting and classifying ALL and lymphocytes, incorporating image enhancement, nuclei segmentation, feature extraction, and classification using the RF classifier. The proposed system accurately classified ALL and lymphocyte subtypes, outperforming other classifiers like Multi SVM and MLP. It served as a valuable diagnostic tool for hematologists, achieving 98% accuracy in cell classification with a selected subset of 13 features [53].

Developed to analyze the images of C-NMC 2019 and ALL-IDB2 datasets. Blood micrographs were initially enhanced and then processed using the active contour method to isolate WBC-only regions, which were subsequently analyzed by three CNN models (DenseNet121, ResNet50, and MobileNet). The primary strategy for analyzing ALL images from the datasets involved a hybrid technique combining CNN with RF and XGBoost classifiers. CNN models produced deep feature maps, which were refined using Principal Component Analysis (PCA) to select highly representative features for classification due to the early-stage similarity between infected and normal WBCs. Fusion of deep feature maps from different CNN models, followed by classification using RF and XGBoost classifiers, yielded promising results. The systems achieved high performance metrics, indicating their effectiveness in aiding hematologists in early leukemia detection. However, a notable limitation was the small dataset size, leading to potential overfitting, mitigated by data augmentation. Future work will focus on developing hybrid diagnostic systems combining CNN features with traditional feature extraction methods [54].

Provided identifying AML cells, particularly of the AML M1 and AML M2 types, through morphological imaging of WBCs using the Naïve Bayes' Classifier. Image-processing methods employed included YCbCr color space classification, image thresholding, morphological operations, chain code representation, and the use of bounding boxes, all based on the Naïve Bayes' Classifier. The test process was conducted on 30 images of each AML M1 and M2 cell types, resulting in a system accuracy of 73.33% for cell identification and 54.92% for cell type identification. The Naïve Bayes' Classifier method demonstrated potential in identifying dominant AML cell types based on WBC morphology. The study achieved an accuracy of 73.33% for AML image identification and 54.92% for cell type identification, with precision, sensitivity, and specificity measures provided. As part of future work, the authors recommended contrast normalization before segmentation to reduce noise during the segmentation process [55].

Table 2. Overview of the literature on detecting Leukemia based on RF, K-NN, Naïve Bayes.

Authors	Year	Dataset	Algorithm	Limitations	Results
Rismayanti et al. **[48]**	2023	Microscopic blood smear images labeled as benign or malignant (ALL)	Naive Bayes	Reliance on a specific dataset and Gaussian Naive Bayes assumptions.	Naive Bayes accuracy: 84.13%
Paliwal and Barua **[49]**	2022	Blood parameters dataset used for leukemia detection in machine learning models.	Naive Bayes, RF	The study lacks exploration of additional classifiers and datasets.	RF accuracy: 98.50%
Mishra, Majhi, and Sa **[50]**	2020	ALL-IDB1	K-NN, RF	The paper does not provide any information about the limitations of the preprocessing techniques	K-NN accuracy: 95.038% RF accuracy: 98.67%
Dasariraju, Huo, and McCalla **[51]**	2020	Dataset from The Cancer Imaging Archive with 18,365 labeled images.	RF	Small datasets may lead to overfitting in previous studies. Imbalance across different classes affects model performance.	RF accuracy: 93.45%
Efthakhar, Shovan, and Linkon **[52]**	2020	NCBI GEO dataset, Demo training dataset with 4 patients and 5 gene features.	Naive Bayes, K-NN	The study only uses the Naive Bayes classification algorithm, which assumes independence between features,	Naive Bayes accuracy: 95% K-NN accuracy: 80.2%
Mirmohammadi, Ameri, and Shalbaf **[53]**	2021	Blood samples from 7 normal subjects and 14 patients with ALL.	RF	Similarities in morphology between ALL and lymphocyte subtypes pose challenges.	RF accuracy: 98%

Ahmed et al. **[54]**	2023	C-NMC 2019, ALL-IDB2	CNN, RF	Low number of images in the dataset causing overfitting. Unbalanced dataset classes leading to accuracy bias	RF accuracy: 99.1%
Suryani et al. **[55]**	2021	WBC images of AML M1 and AML M2 types used.	Naive Bayes	The difficulty in accurately determining whether the cytoplasm of AML	Naive Bayes accuracy: 73.33%

C. K-Nearest Neighbor (K-NN), SVM for Detecting Leukemia:

In leukemia detection, K-Nearest Neighbors (K-NN) assesses data similarity, while Support Vector Machines (SVM) efficiently separate leukemia cases based on distinct features.

Focused on the classification of treatment outcomes for pediatric ALL patients, analyzing data from 241 patients at MAHAK hospital between 2012 and 2018. XGBoost and SVM algorithms were employed for classification, with hyperparameter tuning conducted using the grid search method. Tasks included data preprocessing, handling missing values, and normalization. In scenario one, XGBoost demonstrated superior performance with 88.5% accuracy, while SVM excelled in scenario two, achieving 94.90% accuracy. The analysis emphasized the importance of clinical and medical data in predicting treatment outcomes, with fever frequency identified as the most predictive factor [56].

Discussed the automated detection and classification of acute lymphocytic leukemia (ALL) using a Support Vector Machine (SVM). It outlined the process of lymphocyte extraction, feature extraction, and classification using PCA and SVM algorithms. The method achieved an accuracy of 96.00% and a sensitivity of 92.64% in classifying lymphocytes. Techniques such as color-based k-means clustering, GLCM, GLRLM algorithms, and PCA were employed for feature extraction and dimensional reduction. The SVM with an RBF kernel successfully classified white blood cells (WBCs) into healthy and ALL cells, achieving high accuracy. Additionally, Contrast Limited Adaptive Histogram Equalization (CLAHE) was utilized to enhance image quality and contrast, improving detection accuracy. Overall, the proposed method effectively detected and classified acute lymphocytic leukemia [57].

Developed an automatic early detection and classification system for diagnosing leukemia from blood images using machine learning and image processing algorithms. The study utilized 400 leukemic blood images and 50 normal blood images obtained from Jimma University Specialized Hospital, which were preprocessed with contrast enhancement. The system applied K-means image segmentation and feature extraction techniques. Multi-Class Support Vector Machine was employed for leukemia disease detection and classification based on the extracted feature parameters. The system achieved an accuracy of 94.62% in leukemia disease detection and classification, with a sensitivity of 94.17% and specificity of 100%. The diagnosis results were provided in an average of one minute. The potential of digital image analysis coupled with artificial intelligence

offered a less tedious and time-consuming alternative to manual methods for leukemia diagnosis. Additionally, the study recommended exploring direct diagnosing systems for leukemia without the need for staining processes in the future [58].

Compared the traditional Support Vector Machine (SVM) algorithm with the newer Extreme Learning Machine (ELM) for leukemia prediction using the ALL-IDB1 dataset of blood smear images. It found that ELM, with a 92.2448% accuracy rate, outperformed the SVM, which had an accuracy rate of 86.3636%, indicating ELM's superior efficacy in classifying leukemia from images [59].

Presented a method for cancer cell detection by extracting pivotal features from blood cell images and using various classifiers. The research determined that Gradient Boosting Decision Tree classifiers were more effective than Support Vector Machines (SVM). It also identified critical features such as adjacent nuclei presence and nuclear shape irregularity, significantly enhancing cancer detection. This technique, which does not require high computing power like a Graphics Processing Unit, registered an 85.6% F1 score on validation data. Moreover, it pinpointed essential image features that could aid medical professionals in analyzing stained images for leukemia diagnosis, with F1 scores for classifiers including SVM (linear) at 82.3%, SVM (RBF) at 82.6%, and LightGBM at 85.6% [60].

Table 3. Overview of the literature on detecting Leukemia based on K-NN, SVM.

Authors	Year	Dataset	Algorithm	Limitations	Results
Kashef, Khatibi, and Mehrvar **[56]**	2020	Clinical and medical data from paper-based records of 241 ALL patients.	SVM	Time-consuming process with a large volume of inpatient files.	SVM: 94.90% accuracy
P. K. Das, Jadoun, and Meher **[57]**	2020	ALL-IDB1, ALL-IDB2	K-NN, SVM	The paper does not mention the specific limitations or challenges faced during the implementation of the proposed method.	KNN accuracy: 91.20% SVM accuracy: 96%
Dese Gebremeskel et al. **[58]**	2018	70 ALL type leukemia images obtained from ALLDB online dataset.	SVM	Lack of direct leukemia diagnosing system without staining process	SVM accuracy: 94.62%
Chand and Vishwakarma **[59]**	2020	ALL-IDB1 and ALL-IDB2 datasets are used in the research.	SVM	The dataset used is limited to ALL-IDB1	SVM accuracy: 86.3636%
Mandal, Daivajna, and Rajagopalan **[60]**	2020	Dataset from Clark et al. in The Cancer Imaging Archive.	SVM	The study lacks comparison with other state-of-the-art classification algorithms.	SVM accuracy: 85.6%

D. Convolutional Neural Networks (CNN) for Detecting Leukemia:

In leukemia detection, Convolutional Neural Networks (CNNs) excel in analyzing medical images like blood smears, identifying abnormal cell morphology. Artificial Neural Networks (ANNs), including CNNs, process various data types such as patient demographics or genetic markers, to spot leukemia patterns.

Deep Neural Networks (DNNs), a subset of ANNs with multiple layers, enhance feature extraction for accurate leukemia detection. These neural architectures play crucial roles in automating leukemia diagnosis by analyzing diverse patient data effectively. Proposed to diagnose leukemia from microscopic blood sample images. The algorithm integrates squeeze and excitation learning to enhance feature discriminability, improving the representation of leukemia cells while suppressing less informative features. Extensive experiments, including data augmentation, were conducted on two publicly available datasets, ALL_IDB1 and ALL_IDB2, demonstrating promising results for computer-aided diagnosis of leukemia [61].

Developed a system that mimicked a hematologist's workflow, effectively identifying and discarding uncountable and damaged cells before classifying and counting the remaining ones for diagnosis. It achieved significant success in WBC classification, with an accuracy of 82.93%, precision of 86.07%, and an F1 score of 82.02%. In diagnosing acute lymphoid leukemia, the system showed an accuracy of 89%, sensitivity of 86%, and specificity of 95%. Additionally, it was effective in detecting bone marrow metastasis from lymphoma and neuroblastoma, maintaining an average accuracy of 82.93%. This pioneering study broadened the range of cell types considered in leukemia diagnosis and demonstrated high performance in actual clinical environments [62].

Attempted to develop a segmentation technique capable of addressing blood cell nucleus segmentation challenges using four distinct scenarios: K-means, FCM (Fuzzy C means), K-means with FFA (Firefly Algorithm), and FCM with FFA. We aimed to determine the most effective method for blood cell nucleus segmentation, which would be used for the Leukemia classification model. Subsequently, utilizing Convolutional Neural Network (CNN) as a classifier, we constructed a Leukemia cancer classification model from microscopic images. The proposed system's classification accuracy was evaluated using CNN on the ALL-IDB dataset, comparing it to the current state-of-the-art. Through experimental analysis, we found that the model's accuracy approached 99%, surpassing existing models in segmenting and classifying types of Leukemia cancer, particularly ALL. This research contributed to the development of automatic Leukemia cell nucleus segmentation and cancer classification models from microscopic images, addressing a challenging task in medical research. Future extensions of the proposed model could involve testing on larger datasets containing over one million microscopic images and exploring its applicability to other types of cancer [63].

Table 4. Overview of the literature on detecting Leukemia based on CNN.

Authors	Year	Dataset	Algorithm	Limitations	Results
Bukhari et al.	2022	ALL_IDB1, ALL-IDB2	CNN	The paper lacks	CNN accuracy: 97%

[61]				comparison with other state-of-the-art leukemia detection methods.	
Zhou et al. **[62]**	2021	AI-cell platform with 1,732 bone marrow images from children.	CNN	Model needs validation by prospective studies. Limited cell examples,	CNN accuracy: 82.93%
Sharma et al. **[63]**	2022	ALL-IDB dataset contains microscopic images for blood cell nucleus segmentation. Dataset includes ALL and AML samples from Bone Marrow and Blood.	CNN	The research focuses only on Acute Lymphoblastic Leukaemia (ALL) type. Limited discussion on other types of blood cancer like AML, CLL.	CNN accuracy: 99%

The pie chart illustrates the distribution of different classification algorithms used within a particular context. Support Vector Machines (SVM) are the most utilized, accounting for 26% of the classification algorithms applied. Random Forests (RF) follow closely, making up 21%. Logistic Regression (LR) is also significantly represented with 16%, while Naive Bayes (NB) algorithms account for 15%. Decision Trees (DTs) are employed 11% of the time. K-Nearest Neighbors (K-NN) and Convolutional Neural Networks (CNNs) have smaller shares, with 7% and 4% respectively. This distribution suggests a preference for SVM and RF for the tasks at hand, potentially due to their effectiveness in handling complex data patterns or their adaptability to various datasets. However, the less frequent use of CNNs might indicate either a context where image-based analysis is less relevant or the preference for more traditional ML approaches over deep learning in this scenario as shown in figure 3.

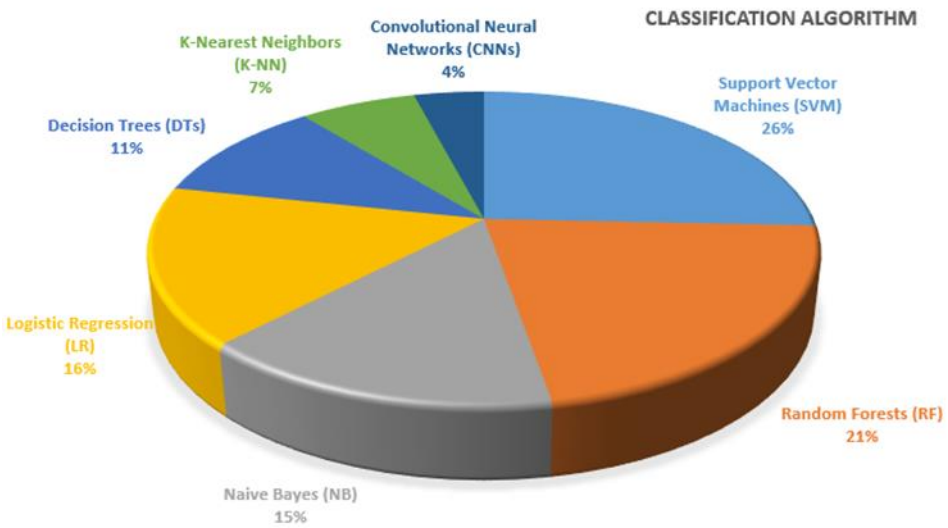

Figure 3. Distribution of Different Classification Algorithms

C. Result and Discussion

A comparative analysis was conducted on machine learning techniques for the automatic classification of Acute Lymphoblastic Leukemia (ALL) cells, utilizing the ALL-CDB database. Seven methods, including Naive Bayes, k-Nearest Neighbor (kNN), Linear Discriminant Analysis (LDA), Decision Tree (DT), Random Forest (RF), Support Vector Machine (SVM), and Multilayer Perceptron (MLP), were applied, with MLP achieving the highest accuracy of 97%. Additionally, deep learning models were employed for precise leukemia classification, achieving a 95.59% accuracy in subtyping. Data mining techniques were also utilized to predict leukemia, with the Decision Tree method being the most accurate. A Convolutional Neural Network (CNN) achieved a remarkable 97.78% accuracy, demonstrating the efficacy of machine learning in diagnosis. Geometric features of ALL cells were analyzed for diagnosis, with both SVM and Artificial Neural Network (ANN) achieving perfect F1 scores of 100%. Moreover, an Optimized CNN model achieved an impressive 99.99% accuracy, showcasing the effectiveness of fuzzy logic in enhancing model performance. The study highlights the potential of machine learning and deep learning techniques in the accurate classification and diagnosis of leukemia, providing a valuable tool for early detection and treatment planning. These findings underscore the critical role of advanced computational methods in modern medical diagnostics, where precision and accuracy are paramount. By leveraging diverse machine learning algorithms and deep learning architectures, researchers can improve diagnostic outcomes, ultimately leading to better patient care. This comparative analysis underscores the versatility and robustness of machine learning models in handling complex medical datasets and achieving high performance in classification tasks.

D. Conclusion

The accurate diagnosis of leukemia is essential for effective treatment planning and improved patient outcomes. The utilization of machine learning algorithms, ranging from traditional classifiers like Random Forest (RF) and K-Nearest Neighbors (K-NN) and Naïve Bayes (NB) and Decision Trees (DTs) to advanced techniques such as CNNs and Support Vector Machine (SVMs), has demonstrated remarkable efficacy in leukemia detection. These algorithms leverage diverse features extracted from blood smear images and patient data to classify leukemia cases with high accuracy. Moreover, recent studies have showcased the potential of deep learning models and hybrid approaches in enhancing diagnostic precision and subtype classification. Despite the progress made, challenges such as dataset variability and algorithm selection persist, necessitating further research to refine existing methodologies and develop more robust diagnostic tools for leukemia management. The integration of machine learning techniques holds immense promise in advancing leukemia diagnosis and contributing to personalized therapeutic interventions for improved patient care, thereby underscoring the pivotal role of data-driven approaches in oncology.

E. References

[1] A. Parthvi, K. Rawal, and D. K. Choubey, "A Comparative study using Machine Learning and Data Mining Approach for Leukemia," Proc. 2020 IEEE Int. Conf. Commun. Signal Process. ICCSP 2020, pp. 672–677, 2020, doi: 10.1109/ICCSP48568.2020.9182142.

[2] Y. Jusman et al., "Comparison between Support Vector Machine and K-Nearest Neighbor Algorithms for Leukemia Images Classification Using Shape Features," ICSEC 2021 - 25th Int. Comput. Sci. Eng. Conf., pp. 70–74, 2021, doi: 10.1109/ICSEC53205.2021.9684585.

[3] C. K. Tebbi, "Etiology of acute leukemia: A review," Cancers (Basel)., vol. 13, no. 9, pp. 1–19, 2021, doi: 10.3390/cancers13092256.

[4] H. Kantarjian et al., "Acute myeloid leukemia: current progress and future directions," Blood Cancer J., vol. 11, no. 2, 2021, doi: 10.1038/s41408-021-00425-3.

[5] H. Inaba and C. G. Mullighan, "Pediatric acute lymphoblastic leukemia," Haematologica, vol. 105, no. 11, pp. 2524–2539, 2020, doi: 10.3324/haematol.2020.247031.

[6] V. R. Minciacchi, R. Kumar, and D. S. Krause, "Chronic myeloid leukemia: A model disease of the past, present and future," Cells, vol. 10, no. 1, pp. 1–23, 2021, doi: 10.3390/cells10010117.

[7] M. Gavillet, J. Carr Klappert, O. Spertini, and S. Blum, "Acute leukemia in the time of COVID-19," Leuk. Res., vol. 92, no. March, 2020, doi: 10.1016/j.leukres.2020.106353.

[8] N. J. Short et al., "Advances in the treatment of acute myeloid leukemia: New drugs and new challenges," Cancer Discov., vol. 10, no. 4, pp. 506–525, 2020, doi: 10.1158/2159-8290.CD-19-1011.

[9] E. Jabbour and H. Kantarjian, "Chronic myeloid leukemia: 2020 update on diagnosis, therapy and monitoring," Am. J. Hematol., vol. 95, no. 6, pp. 691–709, 2020, doi: 10.1002/ajh.25792.

[10] S. J. Faderl and M. J. Keating, "Treatment of chronic lymphocytic leukemia.," Curr. Hematol. Rep., vol. 4, no. 1, pp. 31–38, 2005, doi: 10.1056/nejmra1908213.

[11] M. Siedlinski et al., "White Blood Cells and Blood Pressure: A Mendelian Randomization Study," Circulation, vol. 141, no. 16, pp. 1307–1317, 2020, doi: 10.1161/CIRCULATIONAHA.119.045102.

[12] P. Strati, N. Jain, and S. O'Brien, "Chronic Lymphocytic Leukemia: Diagnosis and Treatment," Mayo Clin. Proc., vol. 93, no. 5, pp. 651–664, 2018, doi: 10.1016/j.mayocp.2018.03.002.

[13] A. Pelcovits et al., "Diagnosis and management of AML in adults: 2022 recommendations from an international expert panel on behalf of the ELN," Blood Res., vol. 29, no. 3, pp. 6245–6259, 2020, [Online]. Available: https://doi.org/10.1016/j.leukres.2021.106727%0Ahttp://www.ncbi.nlm.nih.gov/pubmed/32236160

[14] I. Rahadi, M. Choodoung, and A. Choodoung, "Red blood cells and white blood cells detection by image processing," J. Phys. Conf. Ser., vol. 1539, no. 1, pp. 0–7, 2020, doi: 10.1088/1742-6596/1539/1/012025.

[15] Z. Mousavi Kouzehkanan et al., "Raabin-WBC: a large free access dataset of white blood cells from normal peripheral blood," bioRxiv, no. about 1145, p. 2021.05.02.442287, 2021, [Online]. Available: http://biorxiv.org/content/early/2021/05/06/2021.05.02.442287.1.abstract

[16] L. Alsulaiman and S. Al-Ahmadi, "Performance Evaluation of Machine Learning Techniques for DOS Detection in Wireless Sensor Network," Int. J. Netw. Secur. Its Appl., vol. 13, no. 2, pp. 21–29, 2021, doi: 10.5121/ijnsa.2021.13202.

[17] Z. A. Khan and A. Samad, "A Study of Machine Learning in Wireless Sensor Network," Int. J. Comput. Networks Appl., vol. 4, no. 4, pp. 105–112, 2017, doi: 10.22247/ijcna/2017/49122.

[18] B. Charbuty and A. Abdulazeez, "Classification Based on Decision Tree Algorithm for Machine Learning," J. Appl. Sci. Technol. Trends, vol. 2, no. 01, pp. 20–28, 2021, doi: 10.38094/jastt20165.

[19] P. K. Mallick, S. K. Mohapatra, G. S. Chae, and M. N. Mohanty, "Convergent learning–based model for leukemia classification from gene expression," Pers. Ubiquitous Comput., vol. 27, no. 3, pp. 1103–1110, 2023, doi: 10.1007/s00779-020-01467-3.

[20] I. Reis, M. Rotman, D. Poznanski, J. X. Prochaska, and L. Wolf, "Effectively using unsupervised machine learning in next generation astronomical surveys," Astron. Comput., vol. 34, p. 100437, 2021, doi: 10.1016/j.ascom.2020.100437.

[21] I. H. Sarker, "Machine Learning: Algorithms, Real-World Applications and Research Directions," SN Comput. Sci., vol. 2, no. 3, pp. 1–21, 2021, doi: 10.1007/s42979-021-00592-x.

[22] R. I. Agustin, A. Arif, and U. Sukorini, "Classification of immature white blood cells in acute lymphoblastic leukemia L1 using neural networks particle swarm optimization," Neural Comput. Appl., vol. 33, no. 17, pp. 10869–10880, 2021, doi: 10.1007/s00521-021-06245-7.

[23] S. M. Abas, A. M. Abdulazeez, and D. Q. Zeebaree, "A YOLO and convolutional neural network for the detection and classification of leukocytes in leukemia," Indones. J. Electr. Eng. Comput. Sci., vol. 25, no. 1, pp. 200–213, 2022, doi: 10.11591/ijeecs.v25.i1.pp200-213.

[24] V. Rupapara, F. Rustam, W. Aljedaani, H. F. Shahzad, E. Lee, and I. Ashraf, "Blood cancer prediction using leukemia microarray gene data and hybrid logistic vector trees model," Sci. Rep., vol. 12, no. 1, pp. 1–15, 2022, doi: 10.1038/s41598-022-04835-6.

[25] I. Ibrahim and A. Abdulazeez, "The Role of Machine Learning Algorithms for Diagnosing Diseases," J. Appl. Sci. Technol. Trends, vol. 2, no. 01, pp. 10–19, 2021, doi: 10.38094/jastt20179.

[26] S. M. S. Abdullah and A. M. Abdulazeez, "Facial Expression Recognition Based on Deep Learning Convolution Neural Network: A Review," J. Soft Comput. Data Min., vol. 2, no. 1, pp. 53–65, 2021, doi: 10.30880/jscdm.2021.02.01.006.

[27] F. S. K. Sakthiraj, "Autonomous Leukemia Detection Scheme Based on Hybrid Convolutional Neural Network Model Using Learning Algorithm,"

Wirel. Pers. Commun., vol. 126, no. 3, pp. 2191–2206, 2022, doi: 10.1007/s11277-021-08798-1.

[28] V. Acharya and P. Kumar, "Detection of acute lymphoblastic leukemia using image segmentation and data mining algorithms," Med. Biol. Eng. Comput., vol. 57, no. 8, pp. 1783–1811, 2019, doi: 10.1007/s11517-019-01984-1.

[29] R. Khandekar, P. Shastry, S. Jaishankar, O. Faust, and N. Sampathila, "Automated blast cell detection for Acute Lymphoblastic Leukemia diagnosis," Biomed. Signal Process. Control, vol. 68, no. April, p. 102690, 2021, doi: 10.1016/j.bspc.2021.102690.

[30] N. Ghane, A. Vard, A. Talebi, and P. Nematollahy, "Classification of Chronic Myeloid Leukemia Cell," pp. 382–404, 2019.

[31] K. Muthumayil, S. Manikandan, S. Srinivasan, J. Escorcia-Gutierrez, M. Gamarra, and R. F. Mansour, "Diagnosis of leukemia disease based on enhanced virtual neural network," Comput. Mater. Contin., vol. 69, no. 2, pp. 2031–2044, 2021, doi: 10.32604/cmc.2021.017116.

[32] N. S. Fatonah, H. Tjandrasa, and C. Fatichah, "Identification of acute lymphoblastic leukemia subtypes in touching cells based on enhanced edge detection," Int. J. Intell. Eng. Syst., vol. 13, no. 4, pp. 204–215, 2020, doi: 10.22266/IJIES2020.0831.18.

[33] J. Rawat, J. Virmani, A. Singh, H. S. Bhadauria, I. Kumar, and J. S. Devgan, "FAB classification of acute leukemia using an ensemble of neural networks," Evol. Intell., vol. 15, no. 1, pp. 99–117, 2022, doi: 10.1007/s12065-020-00491-9.

[34] C. KOCATÜRK, C. CANDEMİR, and İ. KOCABAŞ, "Akut Lenfositik Löseminin Makine Öğrenimi Yöntemleriyle Otomatik Tespitine İlişkin Karşılaştırmalı Bir Çalışma," Deu Muhendis. Fak. Fen ve Muhendis., vol. 24, no. 72, pp. 1021–1032, 2022, doi: 10.21205/deufmd.2022247229.

[35] S. Rezayi, N. Mohammadzadeh, H. Bouraghi, S. Saeedi, and A. Mohammadpour, "Timely Diagnosis of Acute Lymphoblastic Leukemia Using Artificial Intelligence-Oriented Deep Learning Methods," Comput. Intell. Neurosci., vol. 2021, 2021, doi: 10.1155/2021/5478157.

[36] K. SRIDHAR, Ajay Reddy Yeruva, Renjith P N, Asmita Dixit, Aatif Jamshed, and Ravi Rastogi, "Enhanced Machine learning algorithms Lightweight Ensemble Classification of Normal versus Leukemic Cel," J. Pharm. Negat. Results, vol. 13, no. 9, pp. 496–505, 2022, doi: 10.47750/pnr.2022.13.s09.056.

[37] S. Chand and V. P. Vishwakarma, "Application of Quadratic Discriminant Analysis Algorithm for the Classification of Acute Leukemia Using Microscopic Image Data," Adv. Appl. Math. Sci., vol. 21, no. 5, pp. 2737–2750, 2022.

[38] K. A. S. A. Daqqa, A. Y. A. Maghari, and W. F. M. Al Sarraj, "Prediction and diagnosis of leukemia using classification algorithms," ICIT 2017 - 8th Int. Conf. Inf. Technol. Proc., pp. 638–643, 2017, doi: 10.1109/ICITECH.2017.8079919.

[39] S. Patil Babaso, S. K. Mishra, and A. Junnarkar, "Leukemia Diagnosis Based on Machine Learning Algorithms," 2020 IEEE Int. Conf. Innov. Technol. INOCON 2020, pp. 1–5, 2020, doi: 10.1109/INOCON50539.2020.9298321.

[40] F. E. Al-Tahhan, M. E. Fares, A. A. Sakr, and D. A. Aladle, "Accurate automatic detection of acute lymphatic leukemia using a refined simple classification,"

Microsc. Res. Tech., vol. 83, no. 10, pp. 1178–1189, 2020, doi: 10.1002/jemt.23509.

[41] F. M. Talaat and S. A. Gamel, "Machine learning in detection and classification of leukemia using C-NMC_Leukemia," Multimed. Tools Appl., vol. 83, no. 3, pp. 8063–8076, 2024, doi: 10.1007/s11042-023-15923-8.

[42] B. K. Das and H. S. Dutta, "GFNB: Gini index–based Fuzzy Naive Bayes and blast cell segmentation for leukemia detection using multi-cell blood smear images," Med. Biol. Eng. Comput., vol. 58, no. 11, pp. 2789–2803, 2020, doi: 10.1007/s11517-020-02249-y.

[43] A. Das Mou, M. W. Hasan, P. K. Saha, N. A. R. Priom, and A. Saha, "Prediction and rule generation for leukemia using decision tree and association rule mining," Proc. 2020 11th Int. Conf. Electr. Comput. Eng. ICECE 2020, pp. 133–136, 2020, doi: 10.1109/ICECE51571.2020.9393069.

[44] K. Sasirekha, M. Sivaraman, and J. Sumitha, "Identification and Classification of Leukemia Using Machine Learning Approaches," Int. Conf. Self Sustain. Artif. Intell. Syst. ICSSAS 2023 - Proc., no. Icssas, pp. 282–288, 2023, doi: 10.1109/ICSSAS57918.2023.10331767.

[45] M. K. Pandey and S. Pal, "Evaluation of Chronic Myelogenous Leukemia (Cml) As the Chronic Phase of Disease Using Machine Learning," Int. J. Mech. …, vol. 6, no. Cml, pp. 198–206, 2021, [Online]. Available: https://kalaharijournals.com/resources/SP-Vol.6_32.pdf

[46] P. More and R. Sugandhi, "Automated and Enhanced Leucocyte Detection and Classification for Leukemia Detection Using Multi-Class SVM Classifier," Eng. Proc., vol. 37, no. 1, pp. 1–7, 2023, doi: 10.3390/ECP2023-14710.

[47] R. E. A. Armya, A. M. Abdulazeez, A. B. Sallow, and D. Q. Zeebaree, "Leukemia Diagnosis using Machine Learning Classifiers Based on Correlation Attribute Eval Feature Selection," Asian J. Res. Comput. Sci., vol. 9, no. 3, pp. 52–65, 2021, doi: 10.9734/ajrcos/2021/v9i330225.

[48] N. Rismayanti, A. Naswin, U. Zaky, M. Zakariyah, and D. Amalia Purnamasari, "Evaluating Thresholding-Based Segmentation and Humoment Feature Extraction in Acute Lymphoblastic Leukemia Classification using Gaussian Naive Bayes," Int. J. Artif. Intell. Med. Issues, vol. 1, no. 2, pp. 74–83, 2023, [Online]. Available: https://doi.org/10.56705/ijaimi.v1i2.9974Datasetlink:https://www.kaggle.com/datasets/mehradaria/leukemia/

[49] M. H. Paliwal and K. Barua, "Machine learning based analytical system for predictive detection of Leukemia using WEKA," vol. 71, no. 4, pp. 8880–8894, 2022.

[50] S. Mishra, B. Majhi, and P. K. Sa, "Texture feature based classification on microscopic blood smear for acute lymphoblastic leukemia detection," Biomed. Signal Process. Control, vol. 47, pp. 303–311, 2019, doi: 10.1016/j.bspc.2018.08.012.

[51] S. Dasariraju, M. Huo, and S. McCalla, "Detection and classification of immature leukocytes for diagnosis of acute myeloid leukemia using random forest algorithm," Bioengineering, vol. 7, no. 4, pp. 1–12, 2020, doi: 10.3390/bioengineering7040120.

[52] U. A. Efthakhar, B. Shovan, and C. Linkon, "A Statistical Approach To Classify The Leukemia Patients From Generic Gene Features," 2020 Int. Conf. Comput. Commun. Informatics, ICCCI 2020, pp. 20–25, 2020.

[53] P. Mirmohammadi, M. Ameri, and A. Shalbaf, "Recognition of acute lymphoblastic leukemia and lymphocytes cell subtypes in microscopic images using random forest classifier," Phys. Eng. Sci. Med., vol. 44, no. 2, pp. 433–441, 2021, doi: 10.1007/s13246-021-00993-5.

[54] I. A. Ahmed, E. M. Senan, H. S. A. Shatnawi, Z. M. Alkhraisha, and M. M. A. Al-Azzam, "Hybrid Techniques for the Diagnosis of Acute Lymphoblastic Leukemia Based on Fusion of CNN Features," Diagnostics, vol. 13, no. 6, 2023, doi: 10.3390/diagnostics13061026.

[55] E. Suryani, Wiharto, A. P. Putra, and W. Widiarto, "Detection of Acute Myeloid Leukemia based on White Blood Cell Morphological Imaging using Naïve Bayesian Algorithm," Int. J. Adv. Comput. Sci. Appl., vol. 12, no. 10, pp. 244–251, 2021, doi: 10.14569/IJACSA.2021.0121027.

[56] A. Kashef, T. Khatibi, and A. Mehrvar, "Treatment outcome classification of pediatric Acute Lymphoblastic Leukemia patients with clinical and medical data using machine learning: A case study at MAHAK hospital," Informatics Med. Unlocked, vol. 20, p. 100399, 2020, doi: 10.1016/j.imu.2020.100399.

[57] P. K. Das, P. Jadoun, and S. Meher, "Detection and Classification of Acute Lymphocytic Leukemia," Proc. 2020 IEEE-HYDCON Int. Conf. Eng. 4th Ind. Revolution, HYDCON 2020, 2020, doi: 10.1109/HYDCON48903.2020.9242745.

[58] K. Dese Gebremeskel, T. Chung Kwa, K. Hakkins Raj, G. Ayana Zewdie, T. Yemaneh Shenkute, and W. Addisu Maleko, "Automatic Early Detection and Classification of Leukemia from Microscopic Blood Image," Abyssinia J. Sci. Technol., vol. 3, no. 1, pp. 1–10, 2018.

[59] S. Chand and V. P. Vishwakarma, "Leukemia Diagnosis using Computational Intelligence," IEEE Int. Conf. Issues Challenges Intell. Comput. Tech. ICICT 2019, 2019, doi: 10.1109/ICICT46931.2019.8977709.

[60] S. Mandal, V. Daivajna, and V. Rajagopalan, "Machine learning based system for automatic detection of leukemia cancer cell," 2019 IEEE 16th India Counc. Int. Conf. INDICON 2019 - Symp. Proc., pp. 1–4, 2019, doi: 10.1109/INDICON47234.2019.9029034.

[61] M. Bukhari, S. Yasmin, S. Sammad, and A. A. Abd El-Latif, "A Deep Learning Framework for Leukemia Cancer Detection in Microscopic Blood Samples Using Squeeze and Excitation Learning," Math. Probl. Eng., vol. 2022, 2022, doi: 10.1155/2022/2801227.

[62] M. Zhou et al., "Development and Evaluation of a Leukemia Diagnosis System Using Deep Learning in Real Clinical Scenarios," Front. Pediatr., vol. 9, no. June, pp. 1–10, 2021, doi: 10.3389/fped.2021.693676.

[63] A. Sharma, D. Prashar, A. A. Khan, F. A. Khan, and S. Poochaya, "Automatic Leukaemia Segmentation Approach for Blood Cancer Classification Using Microscopic Images," Comput. Mater. Contin., vol. 73, no. 2, pp. 3629–3648, 2022, doi: 10.32604/cmc.2022.030879.