Archives available at journals.mriindia.com

International Journal on Advanced Computer Engineering and Communication Technology

ISSN: 2278-5140

Volume 15 Issue 01, 2026

Early Detection of Leukemia Using AI and Blood Test

[1]Jonisha S., [2]Nigil Chryso Praises N. C., [3]Rajisha C., [4]Pakkiya Shri S., [5]Susmitha Rajlin R. S., [6]Rejisha S., [7]Sreeja S. K.

[1,2,3,4,5,6,7] *Arunachala College of Engineering for Women*

Peer Review Information

Submission: 20 April 2026
Revision: 15 May 2026
Acceptance: 27 May 2026

Keywords

Leukemia, Early Detection, Artificial Intelligence, Machine Learning, Complete Blood Count (CBC), Blood Test Analysis, Symptoms-Based Prediction, Low-Cost Healthcare, Risk Prediction, Healthcare Technology

Abstract

The research offers a hybrid method that uses low-cost blood testing, patient symptoms, and artificial intelligence to diagnose leukemia early. The Complete Blood Count (CBC) results, which include the white blood cell count, red blood cell count, platelet count, and hemoglobin levels, as well as symptoms like exhaustion, fever, infections, and bleeding, are gathered by the system. A machine learning model created in Python using tools like Scikit-learn and Pandas is used to preprocess and analyze the data. The approach divides leukemia risk into three categories: low, medium, and high. The system can suggest additional medical testing for confirmation based on the prognosis. The suggested technique is appropriate for early screening, particularly in rural and low-resource areas, because it is quick, inexpensive, and easy to use. For ease of use, it can be developed as a straightforward online or mobile application. The technology aids in early detection and lowers the likelihood of late-stage identification, but it cannot take the place of expert medical diagnosis. All things considered, this strategy raises awareness, facilitates prompt diagnosis, and enhances treatment results.

Introduction

Leukemia is a dangerous kind of blood cancer that affects the bone marrow and blood cell development. Normal white blood cells, red blood cells, and platelets that aid in the fight against infections, transport oxygen, and regulate bleeding are produced by the bone marrow in a healthy body. Leukemia, on the other hand, causes aberrant white blood cells to proliferate and malfunction. The immune system is weakened and general bodily functions are impacted when these aberrant cells proliferate out of control and displace healthy blood cells. There are several forms of leukemia, including acute and chronic leukemia. While chronic leukemia grows gradually over time, acute leukemia progresses quickly and necessitates immediate medical intervention. Although the exact causes of leukemia are not always known, risk factors include weakened immune systems, exposure to dangerous chemicals, radiation, and genetic abnormalities. Fatigue, recurrent infections, fever, inexplicable weight loss, easy bruising, and bleeding are typical symptoms. Leukemia is frequently discovered at a later stage, making treatment more challenging, because many of these symptoms are modest and resemble those of other common disorders. Leukemia treatment results and patient survival are greatly enhanced by early identification. Blood tests include a complete blood count (CBC), peripheral blood smear, and bone marrow examination are examples of traditional diagnostic techniques. Even though these techniques work well, persons in rural or low-resource areas might not always be able to buy or obtain them. Delays in diagnosis can also result in major problems and a lower possibility of a successful course of therapy. Artificial Intelligence (AI) and Machine Learning (ML) have become potent tools in the healthcare industry as a result of technological advancements. Large volumes of medical data may be swiftly and precisely analyzed by these

technologies, which aids in early disease prediction and decision-making. It is conceivable to create a system that can help with leukemia early diagnosis more quickly and effectively by fusing AI with basic medical data, such as patient symptoms and inexpensive blood testing. In order to predict the risk of leukemia, this study suggests a hybrid early detection system that combines patient symptoms, Complete Blood Count (CBC) results, and machine learning algorithms. Because of its straightforward, affordable, and user-friendly design, the system can be widely used, particularly in places with restricted access to cutting-edge medical facilities. The system's output offers a risk level classification (low, medium, or high), which can assist in determining which people need additional medical assessment.The main objective of this project is to improve early diagnosis, reduce healthcare costs, and increase awareness about leukemia. By enabling quick and preliminary screening, the proposed system aims to support doctors and healthcare providers in making timely decisions and ultimately improve patient outcomes

Literature Survey

Leukemia identification and prediction utilizing medical data and artificial intelligence methods has been the subject of numerous research investigations. For an accurate diagnosis, traditional approaches mostly rely on laboratory testing such bone marrow biopsy, peripheral blood smear, and Complete Blood Count (CBC). Although these techniques yield dependable findings, they are expensive, time-consuming, and might not be readily available in rural or low-resource environments. Recent developments in machine learning and artificial intelligence have enhanced leukemia early detection. To interpret blood test data, numerous researchers have created models utilizing techniques like Decision Trees, Support Vector Machines (SVM), and Neural Networks. These algorithms may accurately predict the presence of leukemia by recognizing trends in platelet counts, red blood cell counts, and white blood cell counts. Some studies have also focused on image-based detection using microscopic blood cell images. Deep Learning techniques, especially Convolutional Neural Networks (CNN), are used to automatically detect abnormal cells from blood smear images. Although these methods provide high accuracy, they require high-quality imaging equipment and large datasets, making them less practical for low-cost applications. Other research works have explored symptom-based prediction systems, where patient symptoms are used along with medical data to improve prediction results. These hybrid approaches show better performance compared to single-data methods because they combine multiple factors affecting the disease. However, the majority of current systems either just concentrate on one kind of data, such photos or lab findings, or rely on pricey diagnostic instruments. A straightforward, inexpensive, and readily available approach that integrates symptoms and fundamental blood test results for early detection is still required. In order to provide a quick and affordable solution for leukemia early diagnosis, particularly in settings with limited resources, this study suggests a hybrid strategy that incorporates patient symptoms, CBC results, and machine learning approaches.

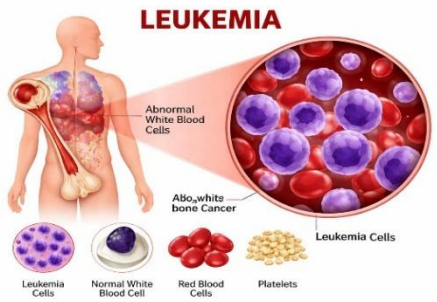

Figure 1. Leukemia and Abnormal White Blood Cell Proliferation

Problem Statement

Leukemia is a dangerous blood malignancy that interferes with the body's ability to produce and use blood cells normally. Early leukemia detection is crucial since it can greatly enhance treatment results and raise survival rates. However, the illness is frequently identified later on because of mild and nonspecific early symptoms like fever, infections, and exhaustion. Even if they are accurate, traditional diagnostic techniques are expensive, time-consuming, and difficult for everyone to use, particularly in rural and low-resource areas. This project proposes a hybrid early detection system that makes use of patient symptoms, artificial intelligence, and inexpensive blood tests like complete blood counts (CBCs). In addition to symptoms, the system gathers blood characteristics such as hemoglobin levels, platelet counts, red blood cell counts, and white blood cell counts. A machine learning model created with Python and pertinent libraries is then used to handle and analyze this data. The algorithm forecasts the leukemia risk level and divides it into low, medium, and high categories based on the analysis. This system's primary benefit is that it offers a quick, affordable, and user-friendly early screening approach. It can be made available to a

wider audience by implementing it as a straightforward application. It serves as a helpful tool for early detection and assists in identifying people who need additional medical testing, even if it cannot take the place of a professional medical diagnosis. All things considered, this study emphasizes how crucial it is to integrate technology and healthcare to solve practical issues. The suggested solution seeks to decrease diagnosis delays, raise awareness, and facilitate prompt medical response by utilizing AI and basic medical data. In addition to improving patient outcomes, this method may eventually be used to identify additional illnesses.

Objectives

Leukemia is a potentially fatal blood cancer that damages the bone marrow and interferes with normal blood cell synthesis. Leukemia is frequently discovered at a later stage due to its non-specific early symptoms, such as fatigue, fever, and recurrent infections, which lowers treatment efficacy and survival rates. Improving patient outcomes requires early diagnosis, but conventional diagnostic techniques like bone marrow biopsy and sophisticated laboratory testing can be costly, time-consuming, and not readily available to all societal segments. The goal of this research is to create a hybrid early detection system that combines low-cost blood tests like Complete Blood Count (CBC) with patient symptoms and artificial intelligence. In addition to symptoms, the system gathers blood characteristics such as hemoglobin levels, platelet counts, red blood cell counts, and white blood cell counts. A machine learning model constructed with Python and tools like Scikit-learn and Pandas is then used to preprocess and analyze this data. The algorithm forecasts the leukemia risk level and divides it into low, medium, and high risk groups based on the analysis. The suggested approach has a number of benefits, such as quicker analysis, lower costs, and user-friendliness. Even in remote and resource-poor locations, it can be implemented as a straightforward application. It functions as an efficient first screening tool that aids in identifying people who need additional medical evaluation, even if it cannot take the place of a professional medical diagnosis. This promotes prompt treatment and shortens diagnosis delays.

Proposed Methodology

A hybrid approach incorporating patient symptoms, blood test data, and machine learning algorithms is used in the suggested system for early leukemia detection. First, patient information is gathered, including symptoms (fatigue, fever, infections, bleeding), as well as Complete Blood Count (CBC) values (white blood cell count, red blood cell count, platelet count, and hemoglobin levels). After that, the gathered data is preprocessed to eliminate errors and transform it into an analysis-ready format. To increase the system's accuracy, key features are chosen. This data is used to train a machine learning model to identify leukemia-related patterns. The model is used to forecast the degree of risk for fresh input data after it has been trained. The outcome is categorized by the system as low, medium, or high risk. The system can suggest additional medical testing for confirmation based on the prognosis. Python and machine learning libraries are used throughout the system's implementation, making it quick, easy, and economical. This approach facilitates prompt diagnosis and early screening.

Figure 2. Proposed Machine Learning Framework for Leukemia Risk Prediction

Working Principle

The suggested method for early leukemia detection relies on a machine learning algorithm to analyze patient symptoms and simple blood test results. The system first gathers input data, such as symptoms (fatigue, fever, infections, bleeding) and Complete Blood Count (CBC) values, which include hemoglobin levels, white blood cell counts, red blood cell counts, and platelet counts. After that, this data is preprocessed to eliminate errors and transform it into an analysis-ready format. A trained machine learning model receives the processed data and uses previously learnt data to find patterns linked to leukemia. After analyzing the input, the model forecasts the leukemia risk. Lastly, the outcome is shown by the system as low, medium, or high risk. The method recommends additional medical testing for

confirmation if the risk is significant. As a result, the technology facilitates quick, inexpensive, and early leukemia screening. The first step in the leukemia early detection method is gathering patient symptoms and CBC blood test results. Important features are then chosen for analysis after the data has been preprocessed. This data is used to train a machine learning model to identify leukemia-related patterns. The system evaluates fresh patient data and determines whether the risk level is low, medium, or high. It makes suitable recommendations based on the outcome, such as suggesting additional medical testing for high-risk cases. The system may be updated with fresh data to increase accuracy, and the output is presented in an easy-to-understand style.

Applications

A dangerous and sometimes fatal blood cancer, leukemia damages the bone marrow and interferes with the regular creation of blood cells. Leukemia is frequently discovered at a later stage, which lowers the likelihood of a successful course of therapy, because of its moderate and non-specific early symptoms, such as fatigue, fever, and recurrent infections. Therefore, improving patient outcomes and survival rates requires early identification. However, conventional diagnostic techniques like bone marrow tests and sophisticated laboratory procedures are costly, time-consuming, and difficult for residents of rural and low-resource areas to get. Artificial intelligence, patient symptoms, and inexpensive blood tests like Complete Blood Count (CBC) are all included in this project's hybrid early detection system. In addition to symptoms, the system gathers blood characteristics such as hemoglobin levels, platelet counts, red blood cell counts, and white blood cell counts. A machine learning model created with Python using tools like Scikit-learn and Pandas is used to preprocess and analyze this data. The leukemia risk level is classified into low, medium, or high categories by the trained algorithm, which also finds patterns. The suggested technique is intended to be straightforward, quick, economical, and easy to use. A variety of users, including those who live in remote places, can access it because it can be deployed as a web or mobile application. The technology reduces diagnosis delays by identifying those who need more medical evaluation by offering a preliminary risk prediction. The system serves as an efficient screening tool that aids physicians and other healthcare providers in making decisions, even while it does not take the role of expert medical diagnosis. Additionally, it contributes to raising awareness about leukemia and the significance of early identification. All things considered, this study shows how combining artificial intelligence with fundamental healthcare data can offer a workable solution to actual medical issues. To sum up, the suggested hybrid method helps to improve patient outcomes, lower healthcare expenses, and improve early diagnosis. This method is an important development in the field of healthcare technology since it can be expanded to identify additional illnesses.

Performance Analysis

The effectiveness of the suggested approach for early leukemia identification is assessed by how well it can forecast the risk level using patient symptoms and CBC blood test results. To find trends linked to leukemia cases, the system employs machine learning algorithms that have been trained on accessible datasets. Standard assessment metrics including accuracy, precision, recall, and F1-score are used to gauge how successfully the model detects real situations and steers clear of false predictions. The model's performance in terms of true positives, true negatives, false positives, and false negatives is examined using a confusion matrix. In order to guarantee that real leukemia cases are not overlooked during early detection, high recall is particularly crucial in this approach. Additionally, the system's processing time is assessed, demonstrating that it produces results quickly and is hence appropriate for real-time or almost real-time applications. Additionally, because the system uses inexpensive blood tests like Complete Blood Count (CBC) rather than costly diagnostic procedures, it is cost-effective. The size and quality of the dataset used to train the model determine the overall performance. The system can produce dependable and consistent outcomes with the right data and customization. In summary, the suggested approach performs well in terms of accuracy, speed, and cost-effectiveness, making it a workable option for leukemia early screening. It can help medical professionals make better judgments and facilitate prompt diagnosis.

Advantages

The suggested leukemia early detection system has a number of significant benefits that make it a workable and efficient solution in the medical field. Early detection, which is essential for increasing patient survival rates and treatment results, is one of the main advantages. The technique helps minimize problems and facilitates prompt medical intervention by detecting the disease at an early stage. Because the approach depends on widely accessible and

basic blood tests like Complete Blood Count (CBC), it eliminates the urgent need for costly diagnostic treatments, which is another important benefit. This makes it ideal for usage in low-resource, rural areas where access to cutting-edge medical facilities may be difficult. Additionally, the system uses machine learning techniques to deliver fast results, enabling rapid analysis and cutting down on the time needed for diagnosis. Because it only needs basic inputs like symptoms and blood test results, the system is also simple to use and intuitive. It can be developed as a mobile or web application, so a variety of users can access it. Additionally, by assisting physicians in identifying high-risk patients and setting priorities for additional testing and treatment, it serves as a decision support tool. Scalability is another significant benefit, since the system may be expanded to identify additional illnesses by changing the model and input data. By adding fresh data, it can also get better over time, becoming more accurate and dependable. All things considered, the suggested approach blends effectiveness, affordability, and accessibility, making it a useful instrument for early screening and improved healthcare administration.

Future Scope

Future advancements and practical application of the suggested leukemia early detection method are highly promising. Increasing the system's accuracy by utilizing bigger and more varied medical datasets is one of the main areas for improvement. To improve performance and provide more accurate predictions, sophisticated methods like deep learning and neural networks can be used. Additionally, by integrating the system with laboratory and hospital databases, patient records may be automatically analyzed, saving human work and boosting productivity. The system will be more accessible if a web-based or mobile application is created, enabling users to conveniently check their health status at any time and from any location. Explainable AI is another significant advancement that boosts user and healthcare professional trust by offering concise and intelligible explanations for forecasts. The accuracy of detection can also be improved by incorporating image-based analysis, such as blood smear image processing. In conclusion, this system has a broad and bright future ahead of it, with chances to increase functionality, enhance accuracy, and include cutting-edge technologies. These advancements have the potential to increase the system's strength, dependability, and utility in delivering effective and easily accessible healthcare solutions.

Conclusion

In conclusion, by fusing artificial intelligence with patient symptoms and inexpensive blood test data, the suggested approach for early leukemia identification offers a workable and efficient option. This approach aids in the early discovery of the illness, which is essential for bettering treatment outcomes and raising survival rates. The approach is effective and time-saving since leukemia risk levels may be predicted quickly and accurately thanks to the application of machine learning algorithms. The system continues to be affordable and accessible, particularly in rural and low-resource areas, by depending on straightforward and easily accessible data like Complete Blood Count (CBC). The system is a useful screening and decision support tool for patients and healthcare providers, but it does not take the place of expert medical diagnosis. It facilitates prompt medical intervention and lessens diagnosis delays. All things considered, this research shows how crucial it is to combine technology and healthcare to address practical issues. The suggested system may enhance early diagnosis, lower medical expenses, and improve patient outcomes.

Reference

"Cancer." Accessed: Jun. 26, 2025 [Online]. Available:https://www.who.int/healthtopics/cancer#tab=tab_1.

S. Wait, et al., Towards sustainable cancer care: reducing inefficiencies, improving outcomes—a policy report from the All.Can initiative, J. Cancer Policy. 13
(Sep. 2017) 47–64,
https://doi.org/10.1016/J.JCPO.2017.05.004.

International Agency for Research on Cancer, "Cancer Site ranking," Global Cancer Observatory, 2020. Online]. Available:
https://gco.iarc.fr/today/onlineanalysis-table.
Accessed: Oct. 3, 2025].

National Cancer Institute, Statistics at a glance: 2022 top 5 most frequent cancers – number of new cases, SEER Cancer Stat. Rev. (2022).
Online]. Available:
https://seer.cancer.gov/statistics-at-a-glance/.
Accessed: Oct. 3, 2025].

B.S. Chhikara and K. Parang, "Chemical biology LETTERS Global Cancer Statistics 2022: the trends projection analysis." [Online]. Available: https://pubs.
thesciencein.org/cbl.

International Agency for Research on Cancer, "Cancer Site ranking," Global Cancer Observatory, 2020. Online]. Available: https://gco.iarc.fr/today/onlineanalysis-table. Accessed: Oct. 3, 2025].

K. Singh, A. Grover, and K. Dhanasekaran, "Unveiling the cancer epidemic in India: a glimpse into GLOBOCAN 2022 and past patterns," 2025 [Online]. Available: www.thelancet.com.

A.K. Deshantri, et al., Nanomedicines for the treatment of hematological malignancies, J. Control Release 287 (Oct. 2018) 194–215, https://doi.org/10.1016/j.