JCBI
Journal of Computing & Biomedical Informatics
ISSN: 2710 - 1606
Volume 03 Issue 02
2022

Research Article
https://doi.org/10.56979/302/2022/73

A Systematic Literature Review on Leukemia Prediction Using Machine Learning

Saroosh Malik [1], Arslan Iftikhar [2 *], Faheem Haider Tauqeer[3], Muhammad Adil[4], and Sheeraz Ahmed[3,4]

[1]Department of Computer Science, University of Gujrat, Gujrat, Pakistan.
[2]Department of Computer Science, TIMES Institute Multan, Multan, Pakistan.
[3]Career Dynamics Research Center, Peshawar, Pakistan.
[4]Department of Computer Science, Iqra National University, Peshawar, Pakistan.
[*]Corresponding Author: Arslan Iftikhar. Email: itsmeiftikhar@gmail.com

Received: July 11, 2022 **Accepted:** September 05, 2022 **Published:** September 27, 2022

Abstract: Blood cancer is one of the most dangerous diseases in kids because it spreads throughout the body, damages healthy cells, and causes uncontrolled white blood cell growth. If it is not promptly treated, blood cancer can kill a person. Leukemia, commonly known as "White Blood Cancer," is the most common type of blood cancer and it has a devastating effect on many people. In essence, it develops in the marrow of bones. Uncontrolled white blood cell proliferation in leukemia harms the body and creates malignant cells. However, throughout the past few years, numerous computer-aided technologies have been employed to manage, identify, treat, and detect leukaemia. Many researchers presented algorithms, methods, and approaches to treat this disease, and after applying them with traditional dataset samples and some of their own created ones, they produced many positive outcomes. Because imaged data sets are utilized in medical examinations, big data plays a significant part in the data collecting for datasets. Large datasets produced positive results for machine learning algorithms such as convolutional neural networks, which are regarded as the best method for classifying images. Hematologic malignancies, which include blood-related diseases like anemia, are cancers that are connected to the blood. Leukemia and multiple myeloma, Hodgkin's lymphoma, non-Hodgkin's lymphoma and encompasses acute AML(Acute Myeloid leukemia), ALL (Acute Lymphoblastic Leukemia), CML(Chronic Myeloid Leukemia), and CLL(Chronic Lymphocytic Leukemia) and affect adolescents, young grownups and the aged, and their prevalence increases with age. The main objective of this research is to present a systematic literature review associated with survey questions on leukemia detection by using machine learning algorithms, techniques, and methods recently. In this study, we outline some past and present uses of machine learning in leukemia prediction, identify manual diagnostic issues, and then discuss the historical background of leukemia discovery. We have also covered leukemia recognition, research motivation, research limits, the composition of white blood cells, and leukemia detection techniques. In order to do our review methodically, the methodology was written with the classification of papers according to their quality.

Keywords: Leukemia; Machine Learning; Cancer; Blood; White Blood Cells.

1. Introduction

As aberrant white blood cells within the bone alter the DNA of the blood cells, blood cancer begins in the bone marrow. It has a few types according to some research namely, leukemia, lymphoma (Hodgkin lymphoma, follicular lymphoma and Burkitt lymphoma, lymphocytic) [1], myelogenous ((myeloid, myeloma, myelodysplastic syndromes (MDS) and myeloproliferative neoplasms (MPN)) [2]. The creation of an excessively high number of cells occurs as a result of leukaemia (blood cancer), which starts in the bone

marrow. The most prevalent forms of leukaemia are acute lymphoblastic leukaemia (ALL), acute myeloid leukaemia (AML), chronic lymphocytic leukaemia (CLL), and chronic myeloid leukaemia (CML) [3]. MEDLINE and the IEEE Xplore Digital Library were searched electronically using mesh terminology and Boolean logic. In addition to the top Google Scholar results, the electronic search also included a manual search of related study references. The digitization of pathological slides should continue, and larger libraries for different pathologies and diseases are needed. These libraries can serve as reliable databases that are used to create and test new models. In addition to quantitatively boosting sample numbers, libraries can allow sets to be more diverse and not constrained to a specific demography. The development of models should give way to their use in real clinical situations in AHI/ML research. As a result, AHI/ML research ought to move in the direction of incorporating these models into routine clinical care [4]. Minimal residual disease (MRD), as measured by multi-parameter flow cytometry (FCM), is a reliable and independent indicator of prognosis in B-cell acute lymphoblastic leukaemia (B-ALL). However, for precise flow cytometric MRD detection, operator skill and experience are essential. A reliable, automated, objective approach of FCM-MRD quantification would be most helpful in overcoming technological variability and analytical subjectivity.

As shown in B-ALL by our recommended GMM-based technique, automated FCM-MRD assessment appears to be a fair and consistent assessment tool for use across numerous laboratories [5]. A quicker sickness diagnosis can also significantly enhance patient treatment. The large volume of data that cannot be used for two different causes is one of the limits. The first factor was the nature of the disease, which causes a wide range of CBC value outlier data. Additional contributing causes were the measurement error, the noisy data utilised to record the results, the loss of results, or the failure to disclose results (missing data).

As a result, several samples were dropped in order to reduce the accuracy and generalizability of the study. The second problem of the proposed approach was the absence of clinical signs and symptoms. The cause is that the medical personnel often does not thoroughly and accurately record the clinical indications and symptoms. When the patient is referred to the hospital and when the initial symptoms appear, which might also result in samples, are not always obvious. Differentiating between the two types of leukaemia, including ALL and AML, is a crucial aspect of the suggested approach.

Leukocytes are different from other blood cells like red blood cells and platelets, so they are typically divided into two groups based on cell structure (granulocytes and granulocytes) and cell heredity (lymphoid and myeloid). Neutrophils, basophils, eosinophils, lymphocytes, and monocytes are the five subcategories that are further broken down into these five types of leukocytes. Thymus (T cells), bone marrow (B cells), and natural killer (NK or K cells) cells are the three cell categories into which lymphocytes are further subdivided.

Another technique for diagnosing leukemia involves cytogenetic analysis, which looks for abnormalities in certain chromosomes. The typical approach for determining whether someone has leukemia is a complete blood count (CBC). One such technique for analyzing blood samples is immunohistochemistry. Leukemia can also be diagnosed using techniques from interventional radiology, such as biopsy, catheter drainage, and percutaneous aspiration, as well as methods like Long Distance Inverse Polymerase Chain Reaction (LDI-PCR), Molecular Cytogenetics, and Array-based Comparative Genomic Hybridization (aCGH). Manual detection methods have been demonstrated to be less dependable, efficient, and accurate than computerised diagnosis techniques. They also get around the drawbacks of manual diagnostics. In this field, numerous studies using image processing techniques for leukaemia detection, blood smear analysis, blood counts, and WBC segmentation have been published. The two acute leukaemia types, AML and All, were classified using a technique that used twelve manually derived elements from photographs. Once more, investigations were done using 1500 photo datasets for classification using K-NN, and the accuracy was 86%. Using a segmentation and classification framework backed by mobile clouds, leukocytes were split into subclasses. This method used the k-means clustering algorithm to execute morphological operations on a variety of data, including statistical, texture, and geometric features, in order to remove unnecessary components. After that, a multi-class ensemble SVM was used for classification. Here, 98.6% accuracy was reached using 1030 images of WBCs as the evaluation dataset [6].

This paper's main contribution is a brief explanation of the current and previously employed algorithms, methodologies, and strategies along with an analysis of how well they perform on the pertinent

datasets. To the best of our knowledge, there is no SLR that links the current and prior technologies with the best algorithm, method, and approach to employ for the diagnosis of leukemia and its types and stages utilizing machine learning cumulatively.

By leveraging medical imaging datasets to create patterns and improve usability linked to other types of leukemia, we aimed to research the literature and provide recent leukemia rapids highlighting the breadth of machine learning and its utilization in the medical profession. We have developed a taxonomy of leukemia-causing blood cancers, including its subtypes, in the form of trees. This SLR is intended to assist physicians and researchers with selecting the most accurate method for leukemia detection. Below figure 1 shows the leukemia types with their subtypes lymphoid and myeloid.

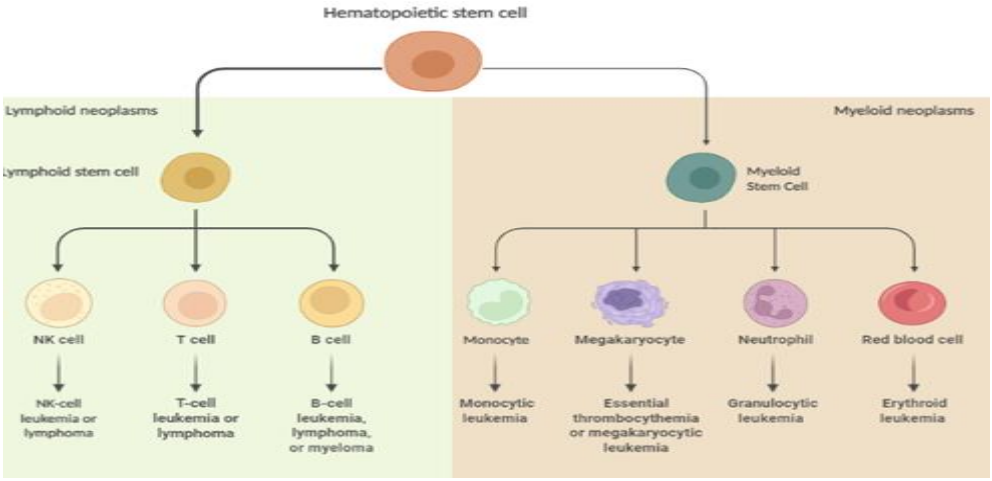

Figure 1. Leukemia types with their subtypes lymphoid and myeloid

This paper's main goal is to present the current and prior leukemia detection technologies, procedures, and algorithms. In order to give the reader an understanding of leukemia cancer and its brief history of origin, we have briefly outlined what leukemia is discussed in section 1 and how its type came to exist shows in figure 2. We reviewed the research on the use of various machine learning algorithms for leukemia detection and diagnosis in the II section of the paper. Finally, we came to the conclusion of our research that the best technology to use is one that is helpful for the detection of leukemia cancer and how it is helpful for patients to secure their lives. To that end, we explored how a systematic literature review was conducted in the III section in the hopes of concluding the paper.

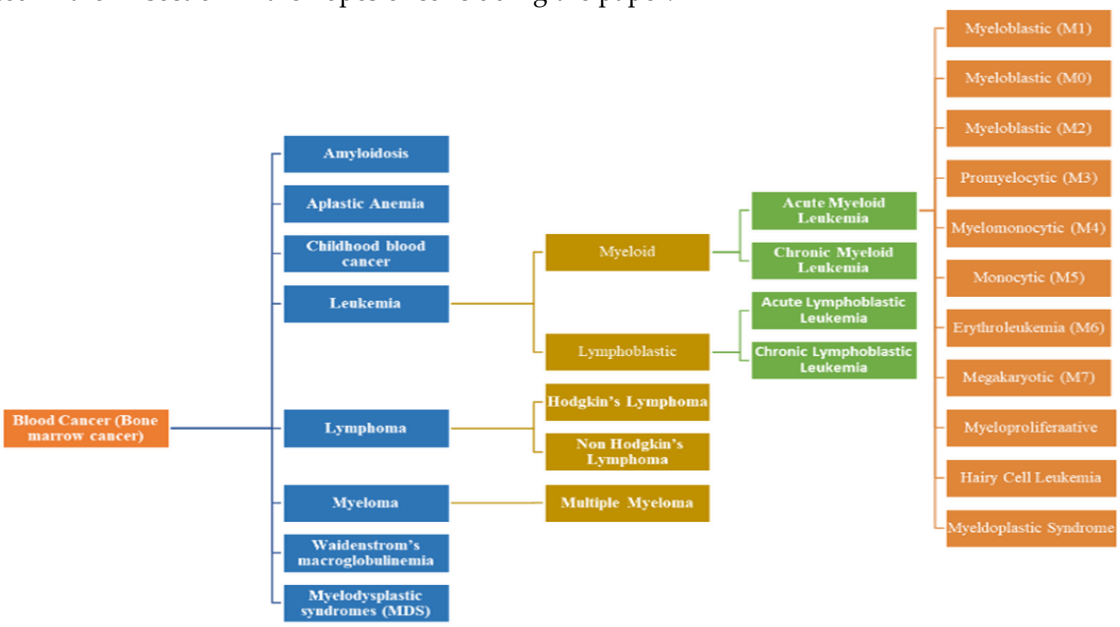

Figure 2. Hierarchy of blood or bone marrow cancer

2. Background

Blood circulates throughout the body of a typical human person, enabling the body's organs to perform duties with the aid of oxygen, nutrition, hormones, and antibodies. Red blood cells provide oxygen to the lungs and muscles, white blood cells fight off germs and viruses, but too many white blood cells in the body can cause infections, platelets thicken blood to stop bleeding, and plasma makes up the majority of blood (responsible for fluid equilibrium and body temperature). The blood weight in our bodies is 8% that of a typical human. Therefore, if any of the blood's components aren't functioning properly, it makes a person more susceptible to diseases, especially if their white blood cells are aberrant. The functions of blood cells that fight infections are disrupted as the unchecked expansion of white blood cells rises. Hematologic and blood cancers may also result from this unchecked growth of white blood cells.

This brief historical overview documents the earliest descriptions of the illness as well as the key developments that occurred during its development up until the turn of the 20th century. Despite the fact that until the middle of the 20th century the majority of leukemia therapies were ineffective (Thomas, 2013). The first account of notable leukemia treatment, "A Chemotherapy," was published in 1948. Today, about 80% of children and teenagers with ALL are treated, which is the best achievement thanks to the supportive work of doctors and researchers in the IT and medical fields who use appropriate algorithms, tools, and methods. In 1965, only about 1% or less of children with ALL were expected to have venerable survivors [7]. Figure 3, shows the normal and affected cells of body.

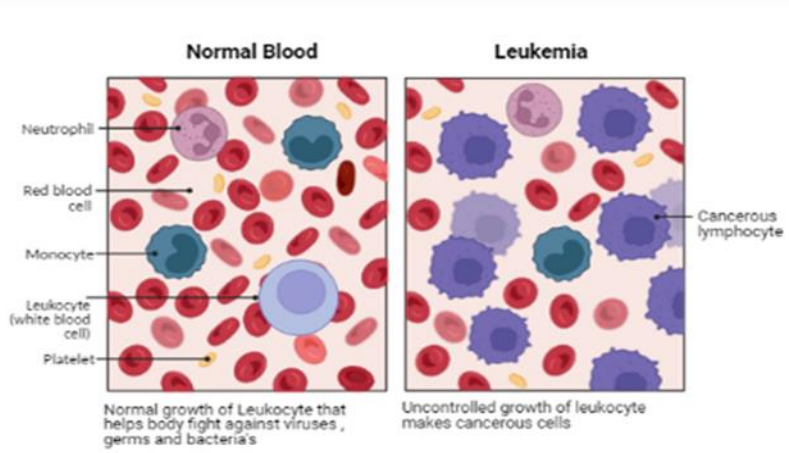

Figure 3. Normal to affected cells of blood

Faster disease diagnosis can also significantly improve patient outcomes [8]. Cancer is a lethal disease that is frequently brought on by the accumulation of hereditary disorders and many pathological alterations. Cancerous cells are abnormal growths that can appear anywhere in the body and are life-threatening. To determine what might be helpful for its treatment, cancer, also known as a tumor, must be promptly and accurately discovered in its early stages. Even while each method has its own unique issues, the main causes of mortality are convoluted histories, inadequate diagnoses, and inadequate treatments, Figure 4 shows the main types of leukemia. Most researchers tested their suggested strategies on tiny sets or chose not to use benchmark 2. For this purpose, the current state of art techniques is compared on benchmark datasets and limitations of existing techniques are highlighted.

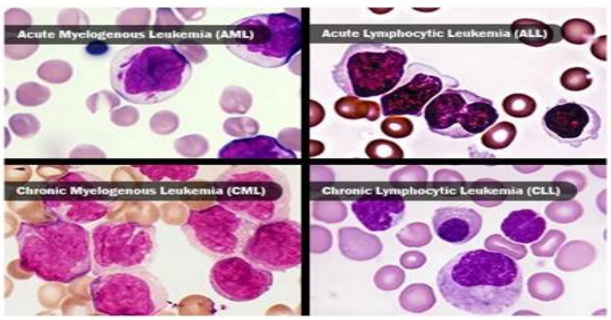

Figure 4. Main types of leukemia

The main challenges in the cancer detection and cure process are redesigning the research pipeline, comprehending cancer growth phenomena, developing preclinical models, precisely handling complex cancers, early treatment, innovative methods of designing and delivering clinical trials, and improving the accuracy that will be useful for doctors as a second and early opinion [9].

Frameworks for computer vision, machine learning (ML), and deep learning (DL) that more accurately recognize leukemia blood cells. A major problem with these traditional methods is that they may provide satisfactory solutions to some problems while decreasing accuracy for others. The adoption of cutting-edge DL networks, which have increased levels of accuracy in segmentation and classification, is a significant advancement in the automated diagnosis of leukemia. The idea of transfer is the most recent invention in this regard. Learning that improves the accuracy of a deep learning network for another problem, such as detecting leukaemia. ResNet, AlexNet, and other networks are among these. The introduction of ML and DL approaches improved the accuracy of automated leukaemia diagnosis, albeit at the expense of significant resources used for computer power and dataset collection. Naive Bayes and K-NN were used to categorize malignant and benign cells using geometric, color, statistical, and textural data. Here, 60 blood image samples were used for examination, yielding a classification accuracy of 92.8%.

3. Literature Review

White blood cells in the bone marrow are affected by the blood malignancy of leukemia. The percentage of white blood cells in peripheral blood must be counted in order to identify acute leukemia. If this type of cancer is not found in its early stages, it can be fatal.

In reality, acute leukemia is found using manual microscopic inspection techniques. But because of human limitations like weariness, stress, and inexperience, these manual approaches are imprecise, prone to mistakes, and time-consuming. In order to replace the manual approaches, a number of image processing algorithms have been developed. We cover a number of computer-aided techniques for diagnosing acute leukemia, including image acquisition, pre-processing, segmentation, feature extraction, and classification. Leukemia can be quickly and readily diagnosed using image processing techniques, increasing the likelihood that patients will survive and receive the right care.

In terms of accuracy and speed, image processing and machine learning algorithms may be able to substitute blood analysis professionals in the detection and categorization of leukemia [10]. Manual diagnostic techniques need more time, are less accurate, and frequently result in mistakes because of various human factors including stress, weariness, and so on [11]. One of the early strategies for leukemia detection is segmentation with image processing. There are various algorithms, techniques, and methodologies created to identify blood malignancy leukemia. Automated hematology analysis has fundamental challenges with image segmentation, which necessitates accurate execution.

Three algorithms—k means, Marker controlled watershed, and HSV color-based segmentation—as well as the SVM classifier, were employed by V P. Jagadev and H. G. Virani to analyse data and identify the different forms of leukemia [3]. Pan, L. etal used machine learning applications to identify the risk factor of acute lymphoblastic leukemia in children by involving socio-demographic, clinical, immunological, and cytogenetic variables with the cross-validation segmentation nested by 10 folds [12]. Sahlol, A.T etal. used ALL-IDB2 by applying the social spider algorithm (SSOA) for the detection of abnormal white blood cells, they first convert the image RGB into color space (CMYK) and do segmentation using Zack Algorithm, then some features (shape, color, texture) were selected and the best one used for SSOA resulting accuracy of 99.23% of segmentation, 100% sensitivity and 97.1% of specificity [13].

Rehman etal. Proposed an automated segmentation method to identify leukemia types and deep learning techniques used for the classification of ALL are considered a novelty. They segmented the entire cell's nucleus in order to accurately classify L2 and L3 blast cells by feature extraction, but it still requires some changes in order to improve the accuracy of the classifier findings and the ability to analyze overlapped cells [14]. The term "minimal residual diseases" refers to conditions where dangerous cells are still present in a patient's body after diagnosis and treatment for malignant disease. By using a machine learning tool for automated MFC analysis on MRD detection in AML and MDS, multi-color flow cytometry analysis ensures that ML tool clinical validation has a significant advantage due to its ability to integrate with other diagnostic and treatment methods [15]. Hematologists have been manually diagnosing leukemia for the past few years by examining the patients' blood samples and bone marrow under a microscope,

but this method has a 30–40% mistake rate and lowers the reliability of the diagnosis because it incorporates human variability. Anything that involves people has some risk of failing. As a result of ignoring these problems testing deep convolutional neural networks (CNNs), a method of machine learning algorithm Abdulsalami et al. evaluated, which are used in computer-aided systems that employ microscopic pictures to identify leukemia [16]. The advancement of pathological diagnosis is possible thanks to machine learning (ML), especially given the rising popularity of digitalizing microscopic pictures. Leukemia diagnosis is time-consuming and difficult in many settings around the world, and there is an increasing trend to use ML approaches for its diagnosis [17]. Leukemia is the name for blood and bone marrow malignancy. Immature white blood cells' unchecked reproduction is what causes it. It impairs the body's capacity to resist infection.

White blood cells (WBC) are typically impacted by leukemia. Acute myelogenous leukemia (AML), acute lymphocytic leukemia (ALL), chronic myelogenous leukemia (CML), and chronic lymphocytic leukemia are the four main kinds of leukemia (CLL) [3] mentioned that no one describes and distinguishes the types of blood cancer and bone marrow cancer through any technique, doctors take too much time to identify the type of leukemia by testing the blood smear of patients. To distinguish between the various forms of leukemia, they employ image processing and machine learning methods. The picture segmentation techniques that have been employed are marker-controlled watershed segmentation, HSV color-based segmentation, and K-means clustering. The purpose of this essay is to define leukemia and determine whether it is AML, ALL, CML, or CLL. Successfully carried out with MATLAB. To make the identification procedure more accurate and exact, additional features were also retrieved, including Cell size, Mean, Variance, Standard Deviation, Entropy, Correlation, Skewness, Contrast, Kurtosis, Smoothness, and Homogeneity. The three picture segmentation techniques that were used were the K-means Clustering algorithm, the HSV color-based segmentation algorithm, and the Marker-Controlled Watershed segmentation algorithm [18].

One percent of all blood cells are leukocytes, which are made in the bone marrow. Blood cancer develops as a result of the unchecked proliferation of these white blood cells. They thereby compete with and obstruct the development of healthy blood cells; Figure 05 below shows the main classification of leukemia. Traditionally, the process was completed manually by a knowledgeable expert over the course of a long period of time.

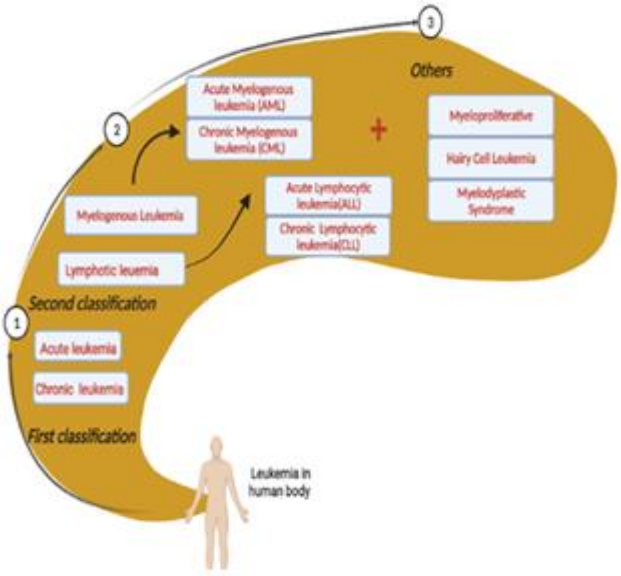

Figure 5. Classification of leukemia

The suggested method has issues with consistency brought on by manual classification work, the need for a qualified professional, and inaccuracies brought on by cells that are difficult to distinguish when seen under a microscope. The suggested model uses deep learning methods, namely convolutional neural networks, to completely eliminate the possibility of errors in the human process. The model retrieves the best

features from the pre-processed photos after being trained on images of cells. Following that, the model is trained using an optimal Dense Convolutional Neural Network architecture (hence referred to as DCNN), and the type of cancer present in the cells is then predicted. [19]. To further assist the physician in daily practice in the diagnosis, prognostication, and therapy allocation of patients with AML, and eventually improve patient outcomes, the goal is to develop integrative tools that can evaluate and interpret data from diverse diagnostic modalities [20].

The suggested structure helps the victims of pandemics like COVID-19. New blood image samples can be added to the dataset in the future, and new augmentation methods can be used to improve performance. Additionally, the suggested IoMT-based framework is capable of detecting the subcategories of each type of leukemia. The suggested models can also be used to identify further blood anomalies. [21]. Here, we draw the conclusion that a variety of modern machine learning methods can classify leukemia sickness. However, it is preferable to employ deep learning architectures for classifications when we have a huge dataset of images [22].

Image Acquisition ALL-IDB2 (https://homes.di.unimi.it/scotti/all/) is used as the source for the input photos. Processing before (There are 260 images. There are 130 photos of leukemia cells and 130 photographs of normal cells. The pre-processing phase of the detection module involves resizing the images to 56 by 256 pixels, shows in figure 6.

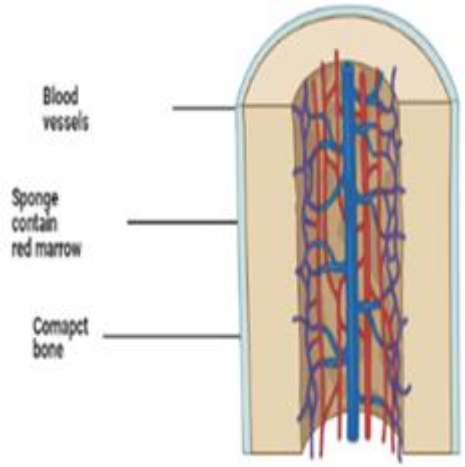

Figure 6. Inner look of bone with vessels

3.1 Performance Evaluation:

Chronological Sine Cosine Algorithm, Features Extraction, and Classification (Dice Coefficient IoU Accuracy) (SCA). To assess the model's functionality, performance metrics like accuracy, loss, precision, recall, and f1 score are computed. [23] The outcomes show that when the same SVM is applied to images segmented using multiple techniques, it produces diverse outcomes. As a result, it immediately attests to the segmentation methods' efficacy as well as their accuracy. The categorization will be more accurate the better the segmentation. Only a few segmentation techniques were used in the experiment. With a larger selection of segmentation methods, the same experiment can be carried out. The presented research illustrates that accurate segmentation is a prerequisite for feature extraction and selection. It was discovered that the manually chosen threshold of 80 to separate the intensity values yields the best result, or 96.89 when compared to other segmentation. Segmentation based on stretching the histogram Manual Thresholding, HSV-based Segmentation, and Otsu's Segmentation are the other three methods [24].

Python has been used to create a model to identify leukemia from microscopic images. In this procedure, infected White blood cells are obtained using image processing techniques like segmentation, dilation, and erosion. Following that, features like shape and color features are extracted from the segmented

picture using the VGG16 architecture, and the images are categorized based on the acquired features. If the input image is malignant or not, the model's output will indicate that. Identifying, diagnosing, controlling, treating, monitoring, and evaluating diseases in as much detail as possible is the major goal of image analysis. 55% plasma and 45% RBC make up blood. Less than 1% of WBC and platelets are currently present. Many algorithms are available and have been used to segment and categorize medical images by a variety of people. The image can be precisely split from the white and black pixels by employing the thresholding procedure. There are many algorithms, including Watershed techniques and K means clustering. There was also the usage of GFCVR (Gaussian feature convolutional visual recognition). KNN classifiers and SVM classifiers were frequently utilized techniques. In order to segment color images of microscopic blood, we utilized the K-mean clustering approach. Next, we extracted shape and color features from the image using VGG architecture, and we used that information to determine if the input image was normal or leukemic [25]. The accuracy, precision, and F1 score for CNN's classification of WBCs were 82.93%, 86.07%, and 82.02%, respectively. And the accuracy, sensitivity, and specificity for the diagnosis of acute lymphoid leukemia were 89%, 86%, and 95%, respectively. With an average accuracy of 82.93%, the method also does well at identifying lymphoma and neuroblastoma bone marrow metastases. This is the first study to use a larger range of cell types to diagnose leukemia, and it performed reasonably well in actual clinical settings. [26]. the diagnosis of many blood-related disorders depends heavily on the image analysis of blood smears. Early detection of leukemia and the initial smears can result in an accurate diagnosis and prompt treatment initiation. In order to diagnose early-onset leukemia and identify subtypes with the least amount of error and the quickest turnaround time possible, blood smear image analysis using machine learning techniques can be used. The use of novel ML algorithms, particularly DL, in CAD systems, whole-slide imaging (WSI), and even apps and software at hematology laboratories to aid pathologists and oncologists in better detecting leukemia can be a potential future route for research. During the 2018 American Society of Hematology meeting, Hollein et al. looked into 43 applications of AI in MFC for the detection of B cell lymphoma and leukemia. A model was created utilizing neural networks using data from 38416 patients and control groups. The system's accuracy in distinguishing between normal and abnormal cells was 97%. Nevertheless, 74% of B cell lymphoma and leukemia diagnoses were accurate. It is advised that the employment of ML algorithms for the analysis of blood smear images move from the modelling stage to the implementation stage in the near future [27].

4. Research Methodology

Author of the selected study describe how we compiled our pertinent information and data from numerous studies in this section. for combining all methods and technologies in order to find the most effective and efficient ones. Before choosing the following plan for this evaluation, we explored a variety of search techniques and received hundreds of results. For the format of our paper, we first came up with a few questions. We merely performed including and excluding pertinent study papers once the inquiries were finalised, and then we defined our search strategy, using which we attained the results and data. Finally, we studied their findings and abstracts to extract the data. However, we also prepared a separate sheet with different parameters to observe for each publication, making it easier to extract the data and enabling us to filter out the useless data. The identified and drawn process of systematic mapping is shown in Figure 7.

Figure 7. Methods for research

4.1. Research Objectives (RO)

The primary objectives of the selected research studies shows in table 01, for leukemia classification techniques in ML.

Table 1. Research objectives.

	DESCRIPTION
RO1	The initial objective is to identify the machine learning algorithms used for detecting leukemia in the earlier and present eras.
RO2	Another objective is to identify the type of datasets used to detect leukemia by machine learning.
RO3	The main emphasis is to describe why early diagnosis is important.
RO4	Lastly, to determine the best data sample size for detecting leukemia i.e. Big data or small data, without affecting the level of accuracy.

4.2. Research Questions (RQ)

Our primary focused question is a genuine game-changer for our article because it contains the key information about our subject. We also learn every detail about our subject, which is divided into divisions, from these questions. The table 02 below lists the research questions that this study addresses.

Table 2. Research Questions

	DESCRIPTION
RQ1	Which machine learning algorithms are mostly used for detecting leukemia in the earlier and present eras?
RQ2	What types of datasets are used to detect leukemia by machine learning?
RQ3	Why early detection of leukemia is important for patients?
RQ4	Is big data affects the accuracy level of data sets or a small data samples best to use for detecting leukemia?

Question 1 - Which machine learning algorithms are mostly used for detecting leukemia in the earlier and present eras?

In the literature review, all the algorithms, techniques, and methods are elaborated on which authors proposed and used them in their research papers. They also suggest their new proposed algorithm with accuracy levels.

Question 2 - What types of datasets are used to detect leukemia by machine learning?

The dataset in machine learning is essentially just a collection of data snippets that a computer can keep as a stand-alone element for systematic and estimation reasons. Many data sets for medical use are publicly accessible on official medical websites, and researchers occasionally create their own sample data using the suggested machine learning techniques. The National Library of Medicine Data Sharing Resources, HealthData.gov Datasets, Interuniversity Consortium for Political and Social Research Health Data Guide, and National Institute of Child Health and Human Development (NICHD) Datasets & Research Resources are some of the resources used to collect healthcare data. Image data sets are frequently used by researchers to identify machine learning methods. Because numerous algorithms used in machine learning create patterns for identifying and forecasting malignant disorders. The architecture that how machine learning algorithm work with data sets is drawn in figure 8.

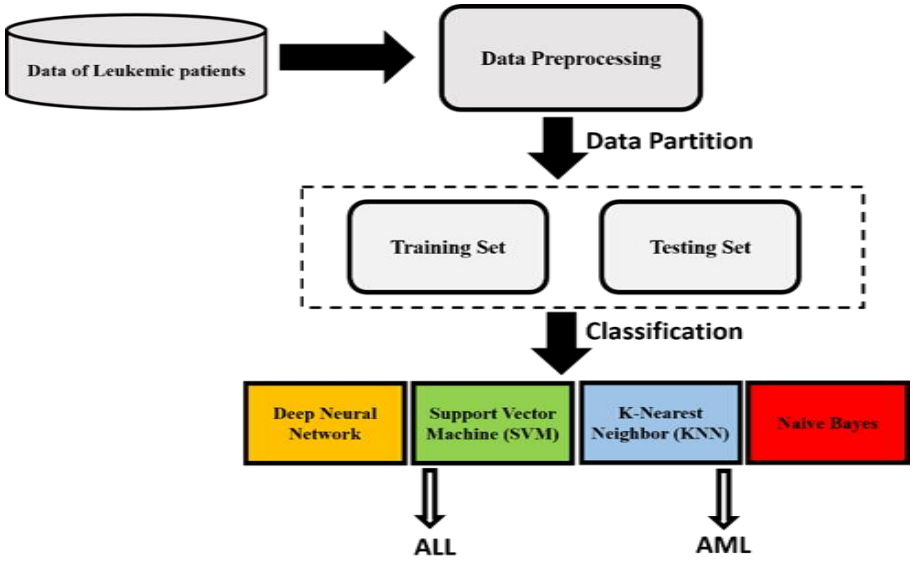

Figure 8. ML Architecture

Question 3 - Why early detection of leukemia is important for patients?

According to popular belief, leukaemia affects children more often than not and is the cause of one in three new cases of childhood cancer each year. The prognosis and survival of people with leukaemia are improved by early diagnosis. Delays in diagnosis can significantly affect a patient's chances of survival, how well they respond to treatment, and even their quality of life. This is why it's critical to identify leukaemia at an early stage. Numerous highly effective algorithms are used by researchers for the detection and diagnosis of leukaemia. Numerous research have shown that AI algorithms could be quite useful in assisting haematologists in diagnosing morphological leukaemia in real clinical settings. In order to train the CNN and evaluate its performance in diagnosing leukaemia by comparing it to that of morphologically knowledgeable hema-tologists, additional data will be collected in the future, and a larger leukaemia database will be developed. According to the National Cancer Institute, leukaemia survivorship is around 65%. Here are several graphs that display the anticipated number of new cases through 2021. According to the well-known website world life expectancy, the most recent WHO data available in 2020 showed that the number of leukaemia deaths in Pakistan was 5,254, or 0.36% of all deaths.

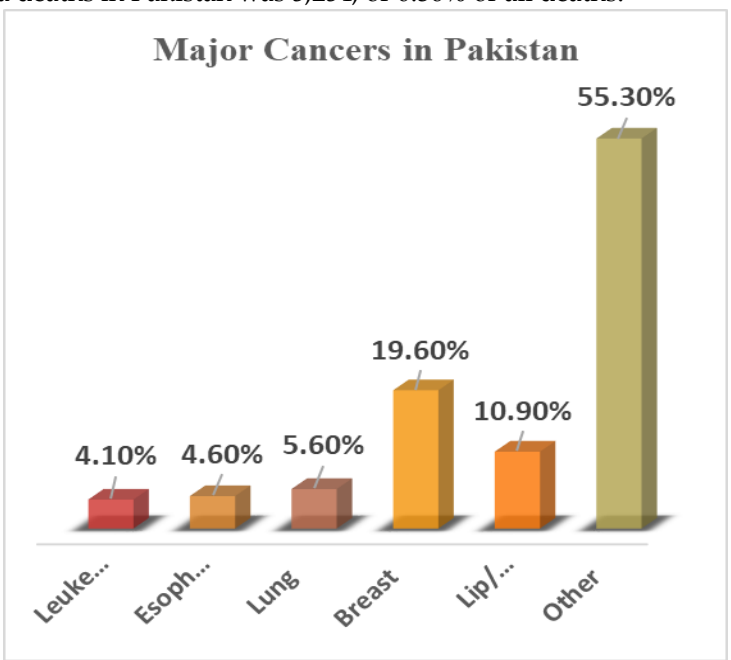

Figure 9. Cancer data

Question 4 - Is big data affect the accuracy level of data sets or a small data samples best to use for detecting leukemia?

Survival with leukemia depends on early identification. For this reason, millions of records are retained in order to clearly distinguish the most correct record. The massive statistics produced by numerical computer-aided technologies used to assemble patient records and assist in the management of hospital records, or else an excessive amount of data that is bulky and complex for out-of-date technology. In accordance with the data collected, many researchers employed small produced data samples to forecast leukemia in patient bodies and reached 80–90% prediction accuracy. The accuracy level may be impacted by big data in the sense that record collection errors may occur, algorithms may not be suitable for the data type, or procedures or methods may not be processed properly. In order to achieve better results, modern medicine (ML) increasingly encourages patient involvement in the healing process [28]. The internet of medical things has sped up the joining machines' categorization of records with greater accuracy. Many multinational organizations use big data analytics to take advantage of the potential of data. The statistical model has the advantage of using data-driven models to analyze the data set and then generate patterns for outcome predictions.

There are privations of numerous important research questions like what further feature collection methods must be used to choose a healthier set of features for the consequence prophecy, how the projected algorithm might be improved to diminish the time and astronomical intricacy, and what other obtainable datasets should be composed for the enactment assessments.

4.3. Search Scheme

We will outline our process for collecting data for our article from both online and offline sources in this part, however we will only mention electronic sources. Five separate databases—the ACM Digital Library, IEEE Explorer, Springer, Wiley Online Library, Science Direct, Research Gate, and JMIR Publications—were used in the course of our study. Then, after gathering some of the most desirable papers from the results, we changed our search query to include some synonyms and tried again. After collecting papers using this method, we tried experimenting with keywords, and by doing so, we were able to obtain the 25 most pertinent articles from 2015 to 2021. These sites were selected because they are thought to be the most trustworthy for information about leukaemia, blood cancer, and bone marrow cancer detection. All of the keyword categories that were part of the study were thoroughly researched.

4.4. Search String

An effective and appropriate analysis has been carried out by expressing a keyword-based string to review and gather available works in the field using numerous well-known digital research sites. In order to confirm the authenticity of the search string regarding the relevance of its effects, the principal perceptions have been evaluated in light of the research questions in order to obtain pertinent keywords and terms used in the particular field of learning. The confirmed keywords and other terms were combined using the logical operators "AND" and "OR," which create a search string. While the "AND" operator concatenates the relations to identify the search alternatives and to limit the query to acquire relevant search results, the "OR" operator allows for more search options. There are two fragments in the confirmed search string. The terms "Bone Marrow Cancer (BMC)" OR "Blood Cancer (BC)" AND "Diagnosis", "Treatment", "Detection" AND ("Using Machine Learning" or "Deep Learning" or "Image Processing") are used to restrict the results relevant to these concepts. resources since they are thought to be the most trustworthy sources of knowledge regarding blockchain protocols. All of the keyword categories that were a part of the study were thoroughly researched.

4.5 Literature Resources

When looking for research papers on my topic, search terms are described. The confirmed search string consists of two components. The string's first character is used to limit the consequences associated with the terms "blood cancer" or "bone marrow cancer," while its second character refers to leukaemia. Table 3 lists the facts of thorough backgrounds, effective examination categorizations, and findings.

Table 3. Particulars of repositories with search strings:

Repository	Search Strings
PLOS	(("leukaemia"[All Fields] OR "leukemia"[MeSH Terms] OR "leukemia"[All Fields] OR "leukaemias"[All Fields] OR "leukemias"[All Fields] OR "leukemia s"[All Fields]) AND ("detect"[All Fields] OR "detectabilities"[All Fields] OR "detectability"[All Fields] OR "detectable"[All Fields] OR "detectables"[All Fields] OR "detectably"[All Fields] OR "detected"[All Fields] OR "detectible"[All Fields] OR "detecting"[All Fields] OR "detection"[All Fields] OR "detections"[All Fields] OR "detects"[All Fields]) AND ("machine learning"[MeSH Terms] OR ("machine"[All Fields] AND "learning"[All Fields]) OR "machine learning"[All Fields])) AND (fft[Filter])″)
Wiley online library	(("leukaemia"[All Fields] OR "leukemia"[MeSH Terms] OR "leukemia"[All Fields] OR "leukaemias"[All Fields] OR "leukemias"[All Fields] OR "leukemia s"[All Fields]) AND ("detect"[All Fields] OR "detectabilities"[All Fields] OR "detectability"[All Fields] OR "detectable"[All Fields] OR "detectables"[All Fields] OR "detectably"[All Fields] OR "detected"[All Fields] OR "detectible"[All Fields] OR "detecting"[All Fields] OR "detection"[All Fields] OR "detections"[All Fields] OR "detects"[All Fields]) AND ("machine learning"[MeSH Terms] OR ("machine"[All Fields] AND "learning"[All Fields]) OR "machine learning"[All Fields])) AND (fft[Filter])
PMC	(("leukaemia"[All Fields] OR "leukemia"[MeSH Terms] OR "leukemia"[All Fields] OR "leukaemias"[All Fields] OR "leukemias"[All Fields] OR "leukemia s"[All Fields]) AND ("detect"[All Fields] OR "detectabilities"[All Fields] OR "detectability"[All Fields] OR "detectable"[All Fields] OR "detectables"[All Fields] OR "detectably"[All Fields] OR "detected"[All Fields] OR "detectible"[All Fields] OR "detecting"[All Fields] OR "detection"[All Fields] OR "detections"[All Fields] OR "detects"[All Fields]) AND ("machine learning"[MeSH Terms] OR ("machine"[All Fields] AND "learning"[All Fields]) OR "machine learning"[All Fields])) AND (fft[Filter])
Springer Link	(("leukaemia"[All Fields] OR "leukemia"[MeSH Terms] OR "leukemia"[All Fields] OR "leukaemias"[All Fields] OR "leukemias"[All Fields] OR "leukemia s"[All Fields]) AND ("detect"[All Fields] OR "detectabilities"[All Fields] OR "detectability"[All Fields] OR "detectable"[All Fields] OR "detectables"[All Fields] OR "detectably"[All Fields] OR "detected"[All Fields] OR "detectible"[All Fields] OR "detecting"[All Fields] OR "detection"[All Fields] OR "detections"[All Fields] OR "detects"[All Fields]) AND ("machine learning"[MeSH Terms] OR ("machine"[All Fields] AND "learning"[All Fields]) OR "machine learning"[All Fields])) AND (fft[Filter])
Science Direct	

	(("leukaemia"[All Fields] OR "leukemia"[MeSH Terms] OR "leukemia"[All Fields] OR "leukaemias"[All Fields] OR "leukemias"[All Fields] OR "leukemia s"[All Fields]) AND ("detect"[All Fields] OR "detectabilities"[All Fields] OR "detectability"[All Fields] OR "detectable"[All Fields] OR "detectables"[All Fields] OR "detectably"[All Fields] OR "detected"[All Fields] OR "detectible"[All Fields] OR "detecting"[All Fields] OR "detection"[All Fields] OR "detections"[All Fields] OR "detects"[All Fields]) AND ("machine learning"[MeSH Terms] OR ("machine"[All Fields] AND "learning"[All Fields]) OR "machine learning"[All Fields])) AND (fft[Filter])
IEEE Xplore	(("leukaemia"[All Fields] OR "leukemia"[MeSH Terms] OR "leukemia"[All Fields] OR "leukaemias"[All Fields] OR "leukemias"[All Fields] OR "leukemia s"[All Fields]) AND ("detect"[All Fields] OR "detectabilities"[All Fields] OR "detectability"[All Fields] OR "detectable"[All Fields] OR "detectables"[All Fields] OR "detectably"[All Fields] OR "detected"[All Fields] OR "detectible"[All Fields] OR "detecting"[All Fields] OR "detection"[All Fields] OR "detections"[All Fields] OR "detects"[All Fields]) AND ("machine learning"[MeSH Terms] OR ("machine"[All Fields] AND "learning"[All Fields]) OR "machine learning"[All Fields])) AND (fft[Filter])
JMIR Publications	(("leukaemia"[All Fields] OR "leukemia"[MeSH Terms] OR "leukemia"[All Fields] OR "leukaemias"[All Fields] OR "leukemias"[All Fields] OR "leukemia s"[All Fields]) AND ("detect"[All Fields] OR "detectabilities"[All Fields] OR "detectability"[All Fields] OR "detectable"[All Fields] OR "detectables"[All Fields] OR "detectably"[All Fields] OR "detected"[All Fields] OR "detectible"[All Fields] OR "detecting"[All Fields] OR "detection"[All Fields] OR "detections"[All Fields] OR "detects"[All Fields]) AND ("machine learning"[MeSH Terms] OR ("machine"[All Fields] AND "learning"[All Fields]) OR "machine learning"[All Fields])) AND (fft[Filter])
ACM Digital Library	(("leukaemia"[All Fields] OR "leukemia"[MeSH Terms] OR "leukemia"[All Fields] OR "leukaemias"[All Fields] OR "leukemias"[All Fields] OR "leukemia s"[All Fields]) AND ("detect"[All Fields] OR "detectabilities"[All Fields] OR "detectability"[All Fields] OR "detectable"[All Fields] OR "detectables"[All Fields] OR "detectably"[All Fields] OR "detected"[All Fields] OR "detectible"[All Fields] OR "detecting"[All Fields] OR "detection"[All Fields] OR "detections"[All Fields] OR "detects"[All Fields]) AND ("machine learning"[MeSH Terms] OR ("machine"[All Fields] AND "learning"[All Fields]) OR "machine learning"[All Fields])) AND (fft[Filter])
Research gate	(("leukaemia"[All Fields] OR "leukemia"[MeSH Terms] OR "leukemia"[All Fields] OR "leukaemias"[All Fields] OR "leukemias"[All Fields] OR "leukemia s"[All Fields]) AND ("detect"[All Fields] OR "detectabilities"[All Fields] OR "detectability"[All Fields] OR "detectable"[All Fields] OR "detectables"[All Fields]

	OR "detectably"[All Fields] OR "detected"[All Fields] OR "detectible"[All Fields] OR "detecting"[All Fields] OR "detection"[All Fields] OR "detections"[All Fields] OR "detects"[All Fields]) AND ("machine learning"[MeSH Terms] OR ("machine"[All Fields] AND "learning"[All Fields]) OR "machine learning"[All Fields])) AND (fft[Filter])

4.6 Inclusion and Exclusion Criteria

Finding and choosing the articles that are most appropriate to the main goals of this research study is to determine the most recent trends. Table 4 displays the inclusion and exclusion standards applied in this investigation.

Table 4. Methods for assessing the quality

Sr #	Evaluation Questions	Expected Answers
1	The topic and the importance of the topic are precisely elaborated?	Yes
2	The literature reviewed on the topic was described in detail?	Intermediate
3	Innovative method, tool, or way of study is applied or developed for the study and is well explained?	Yes
4	Is the research well-structured and depicted through understanding for the reader?	Yes
5	Will the research deliver new knowledge important to the academic sphere?	Intermediate
6	Does the research support the conclusion and the conclusion is relevant and effective?	Yes

After the search was complete, we started looking at the titles, abstracts, summaries, or conclusions to see what should be included and what should be excluded. In order to ensure that they can include or omit data, we built an excel file with a number of parameters for data extraction. Using an Excel sheet, we were able to gather practically all of the crucial information for our paper. The evaluation criteria in Table 4 determine the paper's score.

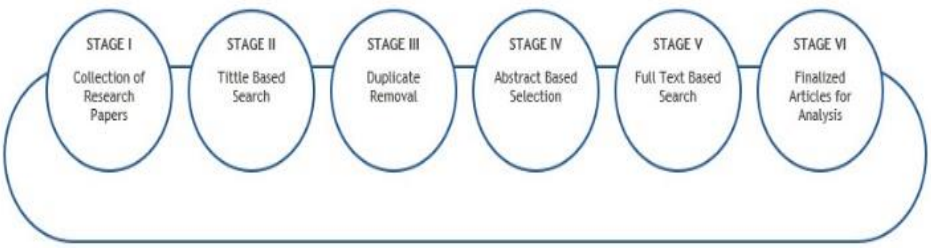

Figure 10. Selection Procedure

5. Data Analysis

In this area, we combine the results and give a clear evaluation of each individual research study. The chosen articles were assessed for their ability to appropriately address the research concerns. The findings of the various search series are shown in the first section, the valuation scores are explained in the second section, and thorough considerations are covered in the last section to answer the research questions.

Figure 10 depicts the selection process, which starts with the collection of all papers, is followed by a search for the root titles of the articles, the elimination of duplicate papers, the selection of papers based on strong abstracts, a full text-based search, and finally the analysis of the papers to produce final research papers.

5.1. Search Scheme

164,940 articles were found using the primary search method across several digital databases. The selection method mentioned in the piece before was effective in this collection. The steps of the selection process are also neatly labelled in Figure 10, and Table 5 articulates the stage levels variety results. Dualistic authors completed the title-generated selection in phase I as P-I, with the results appearing in a collection of 6722 articles. Phase P-II then proceeded to separate the replica articles and the extraneous articles according to the insertion and deletion standards established in phase I.

Table 5. Publisher-Based Stage Wise Selection

Database/Source	Primary Search	P-I	P-II	p-III	p-IV
NIH	81,157	46,244	32,363	90	7
PLOS	22680	5	25	9	5
Wiley Online Library	25	20	15	13	11
Springer Link	11434	26	16	5	1
Science Direct	4	17	12	5	2
IEEE Xplore	7970	6661	159	5	2
ACM Digital Library	122,852	13	7	6	4
Total	164,965	6742	234	43	26

6. Data Quality and Valuation Score

This section presents the conclusions and a thorough analysis of all pertinent research papers. The scoring methodology is precisely described in Table 6 and goes above and beyond the scores that were anticipated in each study's specifically planned analysis of the inner and outer measures. The scores acquired for the exterior and interior criteria are designated as E-Score and I-Score, respectively.

Table 6. Results

Ref.	Source	Paper Type	Data Type	Year	Rank	Score
[3]	IEEE	Conference paper	Imaged Data	2017	Q1	9

[4]	Wiley online library	Journal	Imaged Data	2019	Q2	8
[5]	Wiley Online library	Journal	Imaged Data	2019	Q1	9
[8]	SAGE Journals	Journal	Imaged Data	2020	Q2	8
[9]	Elsevier	Review article	Imaged Data	2020	Q1	9
[10]	Research Gate	Research Article	Imaged Data	2019	not yet awarded	6
[11]	IEEE	Conference Paper	Imaged Data	2020	Q1	9
[12]	Scientific Reports	Research Article	Imaged Data	2017	Q1	9
[13]	Springer	Research Paper	Imaged Data	2019	Q3	7
[14]	Wiley Online Library	Research Article	Imaged Data	2018	Q1	9
[15]	Elsevier	Journal	Image Data	2018	Q1	9
[16]	Elsevier	Journal	Imaged Data	2018	Q1	9
[17]	National Librar	Journal	Imaged Data	2019	Q1	9

	y of Medicine					
[18]	Academia	Research Article	Imaged Data	2020	Q1	8
[19]	IEEE	Journal	Imaged Data	2020	Q1	9
[20]	Blood advances	Review Article	Imaged Data	2020	Q1	9
[21]	Hindawi	Journal	Imaged Data	2020	Q2	8
[22]	IEEE	Conference	Imaged Data	2020	Q1	9
[23]	Springer	Research Article	Imaged Data	2021	not yet awarded	7
[24]	European Union Digital Library	Conference Paper	Imaged Data	2021	not yet awarded	6
[25]	IOP Science	Conference	Imaged Data	2021	Q4	5
[26]	National Library of Science	Journal	Imaged Data	2021	Q1	9
[27]	Research Gate	Research Article	Imaged Data	2022	not yet awarded	7

7. Conclusion

Machine learning (ML) has been shown to be a versatile, accurate, and reliable tool in the diagnostic and therapeutic evaluation of AML, with a number of difficulties for future research. Blood cancer is one of the most deadly illnesses that people may contract. It spreads throughout the body, kills healthy cells,

and induces an uncontrolled development of white blood cells that, if ignored, can lead to death. Leukemia is the most prevalent type of blood cancer and it has a devastating effect on many people. It develops from the bone marrow, much as how unchecked white blood cell proliferation harms the body and produces malignant cells. However, over the past few years, a lot of computer-aided technologies have been employed to regulate, diagnose, treat, and detect leukemia. Many researchers offered algorithms, methods, and approaches to treat this condition, and they had successful outcomes when applied to typical datasets used as samples. The other researchers, however, have created their own for their suggested research because huge data is crucial for the datasets data collecting. Since convolutional neural networks are thought to be the best algorithms for image categorization because they are utilized in medical examinations, large datasets produced positive findings for machine learning techniques. Hematological malignancies are cancers that have a connection to blood and include blood-related illnesses including Leukemia, multiple myeloma, Hodgkin's lymphoma, and non-lymphoma. Adolescents, young adults, and the elderly are all affected by leukaemia, and the risk increases with age. It encompasses acute lymphoblastic leukaemia (ALL), chronic myelogenous leukaemia (CML), and acute myeloid leukaemia (AML) (CLL). We have discussed a few instances in which machine learning algorithms, methods, and approaches have been used to diagnose leukaemia in the past and in the present in this work, which presents a systematic literature review of the research on machine learning algorithms, methodologies, and strategies for leukaemia diagnosis. To determine which approach is most appropriate for use, we have also chosen the most popular machine learning algorithm utilized up to this point and created comparative research with a systematic review. Finally, we have outlined our technique for selecting high-quality papers for our review. Scientists and researchers have created methods and strategies to diagnose and treat leukemia and other blood malignancies in the future. Although there have been advancements, additional machine learning systems still need to be developed due to the limitations of the available

Conflicts of Interest: Throughout this investigation, the authors have declared no conflicts of interest.

References

1. Jagadev, P., & Virani, H. G. (2017, May). Detection of leukemia and its types using image processing and machine learning. In *2017 International Conference on Trends in Electronics and Informatics (ICEI)* (pp. 522-526). IEEE..
2. Pan, L., Liu, G., Lin, F., Zhong, S., Xia, H., Sun, X., & Liang, H. (2017). Machine learning applications for prediction of relapse in childhood acute lymphoblastic leukemia. *Scientific reports*, *7*(1), 1-9.
3. Sahlol, A. T., Abdeldaim, A. M., & Hassanien, A. E. (2019). Automatic acute lymphoblastic leukemia classification model using social spider optimization algorithm. *Soft Computing*, *23*(15), 6345-6360.
4. Salah, H. T., Muhsen, I. N., Salama, M. E., Owaidah, T., & Hashmi, S. K. (2019). Machine learning applications in the diagnosis of leukemia: Current trends and future directions. *International journal of laboratory hematology*, *41*(6), 717-725.
5. Reiter, M., Diem, M., Schumich, A., Maurer-Granofszky, M., Karawajew, L., Rossi, J. G., ... & International Berlin-Frankfurt-Münster (iBFM)-FLOW-network and the AutoFLOW consortium. (2019). Automated flow cytometric MRD assessment in childhood acute B-lymphoblastic leukemia using supervised machine learning. *Cytometry Part A*, *95*(9), 966-975.
6. Shah, A., Naqvi, S. S., Naveed, K., Salem, N., Khan, M. A., & Alimgir, K. S. (2021). Automated diagnosis of leukemia: a comprehensive review. *IEEE Access*.
7. Thomas, X. (2013). First contributors in the history of leukemia. *World J Haematol 2: 62*, *70*.
8. Fathi, E., Rezaee, M. J., Tavakkoli-Moghaddam, R., Alizadeh, A., & Montazer, A. (2020). Design of an integrated model for diagnosis and classification of pediatric acute leukemia using machine learning. *Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine*, *234*(10), 1051-1069.
9. Saba, T. (2020). Recent advancement in cancer detection using machine learning: Systematic survey of decades, comparisons and challenges. *Journal of Infection and Public Health*, *13*(9), 1274-1289.
10. Aljaboriy, S., Sjarif, N., & Chuprat, S. (2019). Segmentation and detection of acute leukemia using image processing and machine learning techniques: a review. *AUS*, *26*, 511-531.
11. Parthvi, A., Rawal, K., & Choubey, D. K. (2020, July). A Comparative study using Machine Learning and Data Mining Approach for Leukemia. In *2020 International Conference on Communication and Signal Processing (ICCSP)* (pp. 0672-0677). IEEE.
12. Pan, L., Liu, G., Lin, F., Zhong, S., Xia, H., Sun, X., & Liang, H. (2017). Machine learning applications for prediction of relapse in childhood acute lymphoblastic leukemia. *Scientific reports*, *7*(1), 1-9.
13. Sahlol, A. T., Abdeldaim, A. M., & Hassanien, A. E. (2019). Automatic acute lymphoblastic leukemia classification model using social spider optimization algorithm. *Soft Computing*, *23*(15), 6345-6360.
14. Rehman, A., Abbas, N., Saba, T., Rahman, S. I. U., Mehmood, Z., & Kolivand, H. (2018). Classification of acute lymphoblastic leukemia using deep learning. *Microscopy Research and Technique*, *81*(11), 1310-1317.
15. Ko, B. S., Wang, Y. F., Li, J. L., Li, C. C., Weng, P. F., Hsu, S. C., ... & Tang, J. L. (2018). Clinically validated machine learning algorithm for detecting residual diseases with multicolor flow cytometry analysis in acute myeloid leukemia and myelodysplastic syndrome. *EBioMedicine*, *37*, 91-100.
16. Abdelsalam, E. M. N., Hussain, K. F., Omar, N. M., & Ali, Q. T. (2018). Computer Aided Leukemia Detection using Microscopic Blood Image Based Machine Learning" Convolutional Neural Network". *Clinical Lymphoma, Myeloma and Leukemia*, *18*, S297.
17. Li, H., Xu, C., Xin, B., Zheng, C., Zhao, Y., Hao, K., ... & Zhou, Y. (2019). 18F-FDG PET/CT radiomic analysis with machine learning for identifying bone marrow involvement in the patients with suspected relapsed acute leukemia. *Theranostics*, *9*(16), 4730.
18. Arunachalam, S. (2020). Applications of Machine learning and Image processing techniques in the detection of leukemia. *International Journal of Scientific Research in Computer Science and Engineering*, *8*(2).
19. Kumar, D., Jain, N., Khurana, A., Mittal, S., Satapathy, S. C., Senkerik, R., & Hemanth, J. D. (2020). Automatic detection of white blood cancer from bone marrow microscopic images using convolutional neural networks. *IEEE Access*, *8*, 142521-142531.
20. Eckardt, J. N., Bornhäuser, M., Wendt, K., & Middeke, J. M. (2020). Application of machine learning in the management of acute myeloid leukemia: current practice and future prospects. *Blood Advances*, *4*(23), 6077-6085.
21. Bibi, N., Sikandar, M., Ud Din, I., Almogren, A., & Ali, S. (2020). IoMT-based automated detection and classification of leukemia using deep learning. *Journal of healthcare engineering*, *2020*.
22. Ratley, A., Minj, J., & Patre, P. (2020, January). Leukemia disease detection and classification using machine learning approaches: a review. In *2020 First International Conference on Power, Control and Computing Technologies (ICPC2T)* (pp. 161-165). IEEE.
23. Alagu, S. (2021). Automatic Detection of Acute Lymphoblastic Leukemia Using UNET Based Segmentation and Statistical Analysis of Fused Deep Features. *Applied Artificial Intelligence*, *35*(15), 1952-1969.
24. Chand, S., & Vishwakarma, V. P. (2021). Comparison of Segmentation Algorithms for Leukemia Classification.
25. Jayavikash, K. P., Gopinath, P., Nandheeswaran, U., Sivabalan, R., & Bharathi, T. D. (2021, May). Detection of Leukemia Using Machine Learning Algorithms. In *Journal of Physics: Conference Series* (Vol. 1916, No. 1, p. 012220). IOP Publishing.
26. Zhou, M., Wu, K., Yu, L., Xu, M., Yang, J., Shen, Q., ... & Zhao, L. (2021). Development and evaluation of a leukemia diagnosis system using deep learning in real clinical scenarios. *Frontiers in Pediatrics*, 616.
27. Ghaderzadeh, M., Asadi, F., Hosseini, A., Bashash, D., Abolghasemi, H., & Roshanpour, A. (2021). Machine learning in detection and classification of leukemia using smear blood images: a systematic review. *Scientific Programming*, *2021*.

28. Markus, A., Roche, A., Ngan, C. K., Cheung, Y. T., & Prifti, K. (2022). A Prognostic Machine Learning Framework and Algorithm for Predicting Long-term Behavioural Outcomes in Cancer Survivors. In *HEALTHINF* (pp. 671-679).