International Journal of Scientific Research in Science, Engineering and Technology
Print ISSN: 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com)
doi : https://doi.org/10.32628/IJSRSET

Leukemia Detection Using Ensemble Model in Machine Learning

[*1]Arunthathi.S, [#2]Ramyaa Sri.S, [#2]Deepika.S, [#2]Baavana.S
[*1]Assistant Professor, Department of Biomedical Engineering, Sri Manakula Vinayagar Engineering College, Puducherry, Tamil Nadu, India
[#2]UG Scholar, Department of Biomedical Engineering, Sri Manakula Vinayagar Engineering College, Puducherry, Tamil Nadu, India

ARTICLEINFO

Article History:

Accepted: 01 March 2023
Published: 05 April 2023

Publication Issue
Volume 10, Issue 2
March-April-2023

Page Number
255-263

ABSTRACT

Leukemia is a kind of blood cancer characterised by the unregulated and abnormal synthesis of white blood cells (leukocytes) in the blood by the bone marrow. Acute lymphocytic leukaemia (ALL), acute myeloid leukaemia (AML), chronic lymphocytic leukaemia (CLL), and chronic myeloid leukaemia are the four primary kinds of leukaemia (CML). The non-specific character of Leukemia symptoms leads to incorrect diagnosis. Additionally, in microscopic investigation, leukemic cells are observed to be extremely similar to normal cells, making identification more challenging. The proposed system aims to develop leukemia detection at early stage using Machine Learning. The Thresholding technique is an existing method which is used to identify cancer cells at early stages. The method of machine learning produces considerable results in identifying the forms of leukaemia by applying the Ensemble classifier algorithm and/or regression. The Machine Learning technique is turned into development tools for diagnosing leukaemia in this suggested system, which would aid healthcare institutions in remote regions with less medical professionals, particularly in screening. It may also be changed to become open source software, which itself is available to use and share. Hence, early detection of leukaemia results in optimal therapy for the patient, ultimately lowering the mortality rate from leukaemia.

Keywords : Machine Learning, Microscopic Analysis, Ensemble Model, Application Software.

I. INTRODUCTION

The objective of medical imaging is to create pictures of various regions of our bodies for the purposes of diagnosis and therapy. This includes the use of certain techniques and processes. During the course of the last decade or two, imaging in medicine has developed into an essential component of medical practise. There has been a significant advancement in the technology used for capturing, analysing, and storing medical pictures. Identifying, diagnosing, controlling, regulating, monitoring, and evaluating illnesses is the primary goal of image analysis. The mortality rate for cancer is among the highest in the world. The malignancy

known as leukaemia is one that affects the cells that produce blood [1]. A high number of aberrant WBCs produced by the bone marrow can cause this type of cancer to develop. Among the blood's main components are plasma, red blood cells, white blood cells, and platelets. Blood consists of 45% RBCs and 55% plasma. WBC and platelets make up less than 1% of this sample. In the event that the patient has leukaemia, WBCs will be generated in greater numbers, and this will cause disruptions in the circulating blood [2]. When a blood sample is taken and examined by a haematologist, it is possible for the haematologist to detect the presence of white blood cells that are abnormal or in excess. After that, the haematologist would conduct an inspection of the microscopic pictures using the naked eye [3]. The undertaking would be stressful and may take a significant amount of time. Moreover, obtaining consistent findings from the perceptible examination is a very challenging endeavour [4]. An automated method in image processing has the potential to eliminate mistakes and other issues that are caused by visual examination. In order to discern normal cells from leukaemia cells, the system that is being developed will focus mostly on microscopic pictures. The earlier an illness is diagnosed, the more chance there is of delivering prompt and suitable treatment for it. It is possible to handle a large number of photos using an automated system, which also helps to cut down on scanning time and improves accuracy. White blood cells (WBC) are examined and classified according to their size, shape, and colour [5]. This is done so that leukaemia can be properly examined and categorised.

II. RELATED WORKS

A comprehensive analysis of the most recent progress made in this area of knowledge [6]. In this paper, a variety of artificial intelligence (AI)-based ALL detection strategies are investigated in a methodical way, with both their merits and their demerits being discussed.

Acute leukaemia is a kind of leukaemia that is characterised by an accumulation of undeveloped WBC, which multiply in the blood system in order to replace blood cells that are healthy [7]. Acute leukaemia is a form of malignant sickness.

Techniques that use deep learning are becoming more important in the field of medical sciences. These methods assist medical professionals in the early diagnosis of illnesses, which reduces both their workload and the likelihood of making mistakes [8].

It is critical to make a prompt diagnosis and adhere to the recommended treatment plan while dealing with leukaemia [9]. The examination of gene expression may provide a diagnosis of cancer that is specific to the subtype. Acute lymphoblastic leukaemia, abbreviated as ALL, and acute myeloid leukaemia, abbreviated as AML, are the two subtypes of leukaemia (AML).

A significant challenge in the field of disease diagnostics [10] is the accurate distinction of malignant leukocytes at low cost and in the early stages of the illness. There is a lack of molecular biological equipment, and the approaches that are accessible at laboratory diagnostic facilities are laborious and time consuming. This is despite the fact that leukaemia is quite common.

There is no cure for cancer. An early diagnosis of malignancy is crucial to successful treatment. Medical Picture Processing is an extremely important component in the whole process of illness diagnosis [11].

CML, also known as chronic myeloid leukaemia, is a condition that affects clonal stem cells and accounts for 15% of adulthood leukaemia cases. In this section, we address the use of computational techniques to forecast CML employing blood number of viable cells prior to diagnosis [12].

The development of leukemia occurs when there is an abnormal increase in the number of white blood cells in the bone marrow. The progression of leukaemia may be broken down into two categories: acute leukaemia and chronic leukaemia. Acute leukaemia progresses at a rapid rate, while the progression of chronic leukaemia is more gradual.

Diseases have grown into the primary cause of morbidity and death among people with chronic lymphocytic leukaemia (CLL) as a direct result of the cytotoxic therapy for CLL, which causes the immune system to become dysfunctional. Nevertheless, there are no available models for infection prediction [14].

Patients typically get sophisticated medical services, such as the detection of leukaemia, as a result of hospitals adopting methods including artificial intelligence for the purpose of identifying illnesses [15].

III. PROPOSED METHODOLOGY

A proper and timely diagnosis is necessary for both the rehabilitation and treatment of patients suffering from leukaemia, which is a kind of cancer that has the potential to be lethal. When it comes to evaluating, diagnosing, and making prognostications about symptoms, the traditional approaches have been replaced by automated computer technologies. Patients may get assistance in diagnosing terminal conditions, such as leukaemia, via the use of modern technologies. Since leukaemia is such a dangerous illness, it is essential to catch it at an early stage when it first manifests. In many cases, the capacity for generalisation possessed by an ensemble is much superior to that possessed by basic learners. The support vector machine (SVM) and the k-nearest neighbour classifier (KNN), both operating under a wrapper-based feature selection model, are the fundamental classifiers that are used in the ensemble learning approach that is suggested in this article.

The purpose of this study is to apply the Machine Learning approach to analyse leukaemia smear pictures in order to enhance accuracy, decrease the amount of time needed for diagnosis, and deliver diagnostic services that are quicker, cheaper, and safer. The approach that is presented in this research is to first extract the lymphocytes from the blood after first separating the other clotting factors from the leukocytes. Many fractal properties, form features, and other textural features are retrieved from the lymphocytes that have been removed.

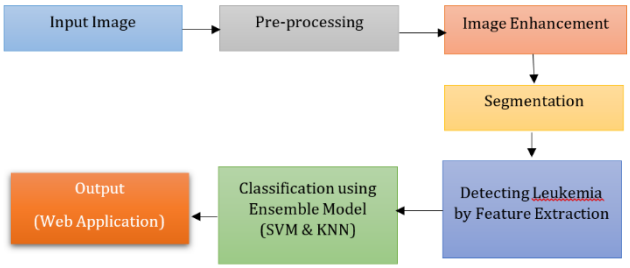


Figure 1. Overall Workflow of Proposed Method

3.1 Pre-processing

A fundamental level of abstraction is represented by the operations performed on pictures called image pre-processing. Entropy measures the amount of information in a picture, so these actions do not increase it; rather, they decrease it. Preprocessing is used to improve picture data by eliminating unwanted distortions or enhancing certain visual elements for further processing and analysis. Procedures are designed to accomplish this goal.

3.2 Image Enhancement

Image enhancement involves processing the original picture to increase the size of the data set. This may help the model function more effectively, at least to some degree. If the initial data set is not very large, it will not be enough for the retraining of the system model, which will have a negative impact on the model's overall performance.

3.2.1 Histogram Equalization

Histogram Image contrast may be improved using a computer method called equalisation, which is part of the field of image processing. This is accomplished by efficiently spreading out the intensity values that occur most often, which results in the image's intensity range being stretched out to a greater extent. Since the useful data of an image is represented by values of near contrast, this approach will often improve the general contrast of the picture. Because of this, regions that previously had a lower local contrast are now able to attain a greater contrast.

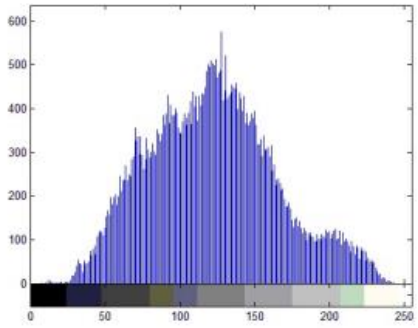


Figure 2. Histogram Image

Histogram Equalization, known as "HE," is a widely used method for enhancing the contrast of a picture. In the past, this method of enhancing picture contrast has been used successfully on photographs depicting both people and natural settings. The results were satisfactory. HE stands for "height equality," and it refers to a procedure that modifies the distribution of grayscale values in a picture such that they are all the same. The objective here is to get a spread that is homogeneous throughout the histogram such that each value on the grey scale has a roughly equivalent amount of pixels. Flattening the histogram is accomplished by replacing the gray level of a pixel (r) with a new grey scale (s) using a transformation function such as T. The following formula, which may be represented in mathematical notation: s = T (r). r may be reconstructed using s by performing an inverse transformation, as shown in the following equation:

$$R = T^{-1}(s)$$

Where 0<s<1. The equation used to calculate the histogram equalization can be written as follow:

$$k = r(\frac{Ci\ (2^r - 1)}{wXh}$$

Where:

Ci = the total number of ith grayscales in the input images

r = rounding to the closest whole number

K = Histogram equalisation grey level value

w = Image width

h = Image height

This technique is also helpful for pictures in which the foreground and backdrop are the same brightness or the same darkness. In particular, this technology offers a clearer and more detailed perspective of the skeletal structure shown in the imaging in the field of biomedicine, making it possible to examine the skeleton more closely.

3.3 Segmentation

In automated image analysis and processing, image segmentation involves a technique that is often used to divide an image into various portions or areas. This partitioning is typically done on the basis of the properties of the pixels included within the picture. The process of segmenting an image may include separating the foreground from the background, or it may require grouping together sections of pixel resolution that are comparable in colour or form.

3.4 Feature Extraction

The process of extracting features transforms a vast collection of duplicate information into a collection of features with decreased dimensions. During the process of feature extraction, the data that was obtained from the picture is converted and labelled so that it may be applied to a specific set of features. These features are then going to be utilised for further categorization. In this step, the characteristics that were produced from the entities that were separated from areas of the picture were extracted and identified. This stage was employed.

3.5 Classification

A collection of previously unclassified data may be assigned and classed via the application of classification. Both unsupervised as well as supervised classifications are examples of different kinds of classifiers. Supervised classifications are the more common of the two. The collection of potential outcomes or classes is already known in advance while doing supervised categorization. During the unsupervised classification process, the collection of classes that will be used is not known in advance.

The Support Vector Machine (SVM) is a method for supervised as well as linear machine learning that is most often used for the purpose of solving classification issues. In the procedure for establishing a function in real space, let's begin with the presumption that we are

already conversant with the ideas of domain, range, and co-domain. When we specify x in an actual space, we are able to comprehend its domain. Moreover, when we map a functional for y = f(x), we are able to comprehend its range as well as its co-domain.

In addition to this, mapping the point into an intricate feature space x,

$$\Phi(x) \in RM$$

It is possible to define the converted feature space as follows, with each input feature having been translated to a converted basis vector x:

$$\phi(x) : R^D \mapsto R^M$$

The phrase "finding nearest neighbours" may be described as "the process of locating the nearest approach to the intermediate node from the supplied data set." This is one way of thinking about the notion. The algorithm remembers all of the existing instances (the test data), and then assigns classes to new cases based on which of its K neighbours has the majority vote. The first thing to do when putting KNN into practise is to convert data points into the corresponding mathematical values (vectors). In order for the method to function, it must first determine the gap in numerical value that exists between each pair of points. It begins by calculating the proximity among each information source as well as the test data, after which it determines the chance of each data point having a similarity to the testing dataset. The points that have the greatest probability are used to determine which groups belong to which categories. The Euclidean, Minkowski, or Hamming distance may be used as the function for measuring distance. The separation of two points on a plane that have the dimensions (x, y) and (a, b), respectively, may be calculated using the Euclidian formula, which states that the distance between the two points is

$$dist((x,y),(a,b)) = \sqrt{(x-a)^2 + (y-b)^2}$$

When the distance has been computed, the input x is then allocated to the category that has the highest probability:

$$P(y = j|X = x) = \frac{1}{K} \sum_{i \in \mathcal{A}} I(y^{(i)} = j)$$

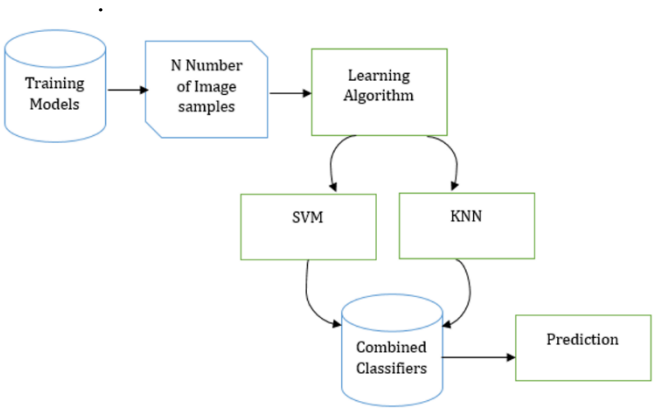


Figure 3. Combined Classifier Model

Algorithm for Ensemble Model

Input: 2D Images with class labels

Output: Leukemia Classification

function test_ensemble(SVM, KNN, iterations, features, labels)

initialize:rate_of_learining=random.math();

for reate_of_learning in iterations

Error=0;

for x in features

if (labels*features*images)<1

then update: rate_of_learning*features*iterations (images)

else

update: rate_of_learning

end if

end

end

• Calculate a histogram based on the image pixels of the picture that was supplied. The frequency of each pixel f[x,y] is sorted by the histogram and placed into one of L bins that are evenly spaced apart.

When L equals 2^8 and the dimension of the picture is MN

• Determine the function of cumulative distribution using the formula.

In order to create the output picture, you must first apply the cumulative distributive function to the input image and then scale it.

IV. EVALUATION OF ENSEMBLE MODEL

We classified leukaemia patients by using a benchmark dataset that had a binary class. First ever value of the class was either acute lymphocytic leukaemia (ALL) or acute myeloid leukaemia (AML). The second value of the class was either chronic lymphocytic leukaemia (CLL) or chronic myeloid leukaemia (CML) (CML). In addition, the tools provided by Dot Net were used in order to pick and categorise features. The array dataset used as part of this research had 7,129 characteristics, and the total number of samples was 72. Also, the samples were split into 38 sets for use as learning data samples, and 34 sets were used as test data sample. The gene expression data that can be accessed via the website http://kaggle.com/ are the dataset that this paper relies upon.

After identifying the dataset, the first thing to do is the preprocessing of the data. The data were made clean and ready for prediction modelling by going through the following procedures as part of the pre-processing procedure: remove the anomaly and extreme values, update the incomplete data (all missing data in a dataset will be replaced with the averages and median first from training data), and eliminate characteristics that share a single value for further than 99% of the patients. For the online application, we employed the suggested ensemble technique, which combined the results of four separate classifiers into a single result using ensemble learning.

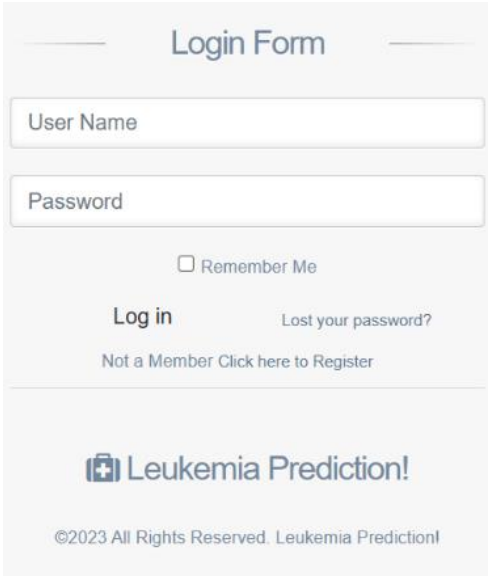
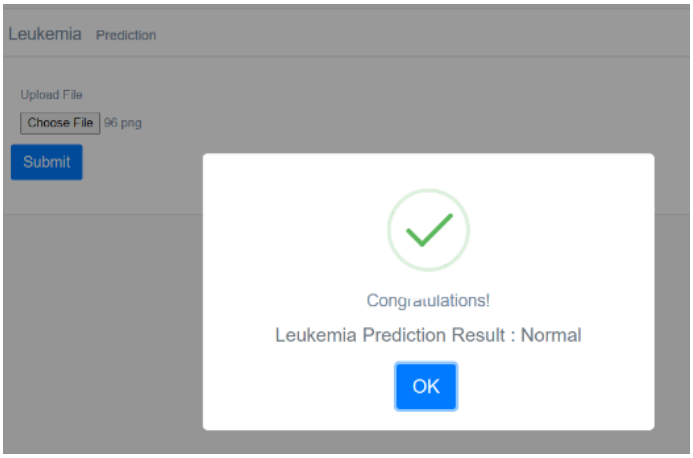
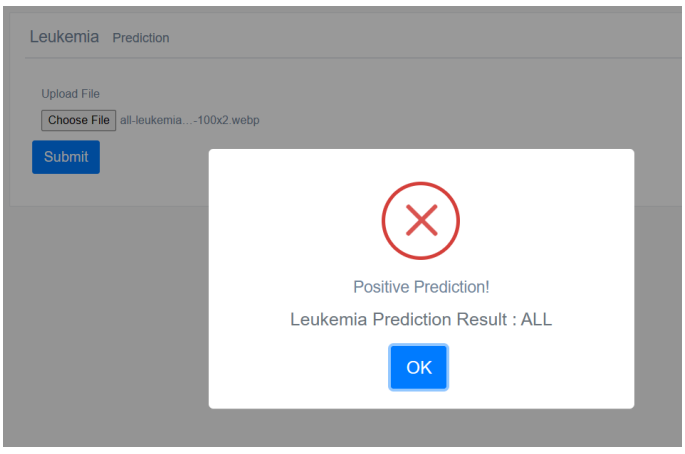


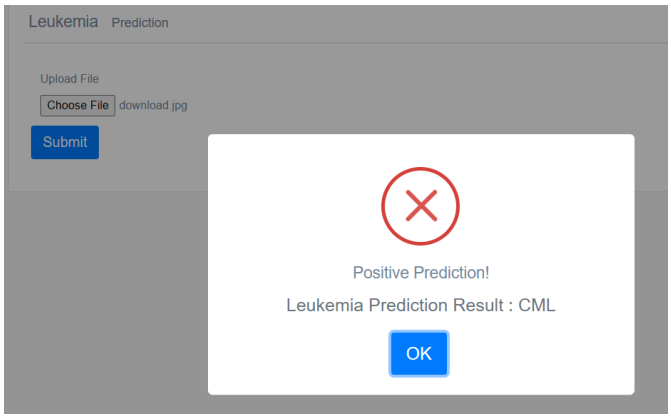
Figure 3. Login Form



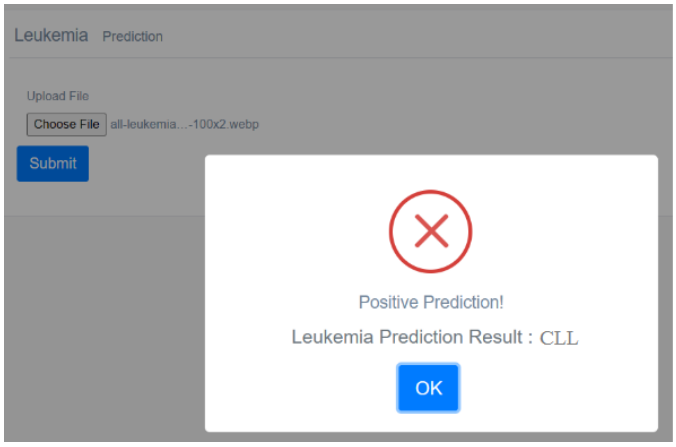
(a)



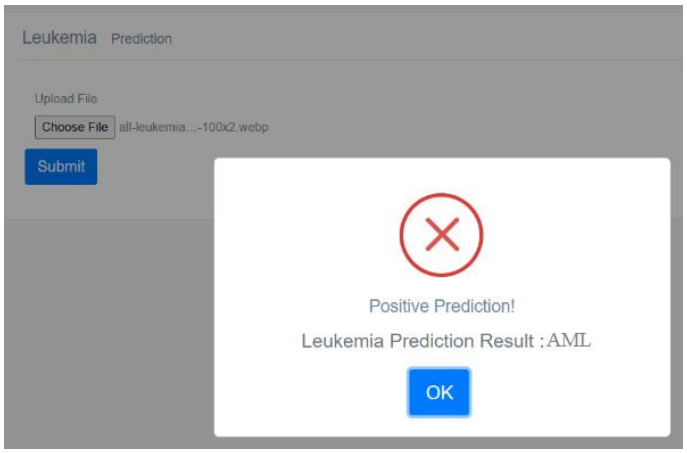
(b)



(c)



(d)



(e)

Figure 4. Classification of Leukemia

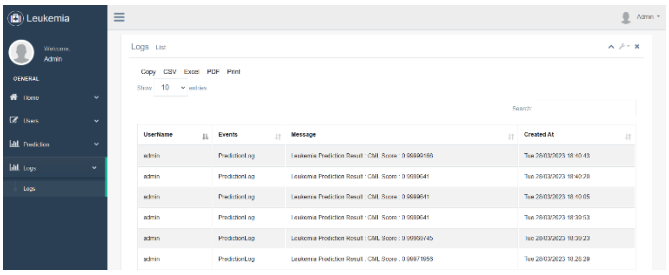


Figure 5. User Logs

Table 1: Feature Extraction Values

Contrast = 2.495546
Correlation = 0.978744
Energy = 0.373831
Homogeneity = 0.811750
Mean = 162.489632
Standard_Deviation = 124.584658
Entropy = 4.684154
RMS = 13.718339
Variance = 11434.417326
Smoothness = 1.000000
Kurtosis = 1.351636
Skewness = -0.534152

Finding the ideal number of features that are considered adequate to learn from the classifier in order to increase accuracy as well as enhance the model's performance requires finding the optimum number of features.

Table 2: Performance Evaluation

Algorithms	Precision	Recall	Accuracy
SVM	82	86	89
KNN	81	83	86.5
Proposed Ensemble Model	84.3	86.2	96.2

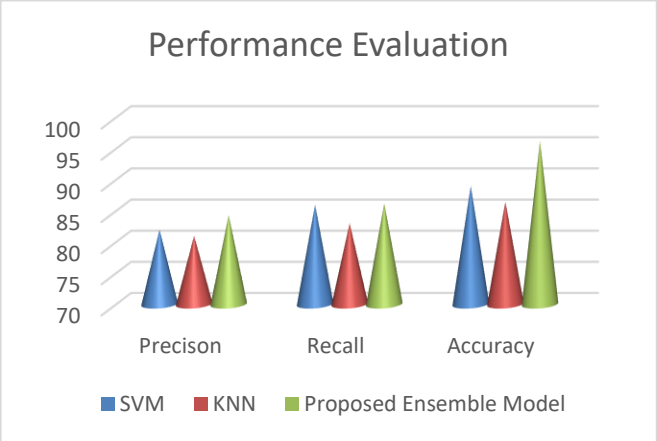


Figure 6. Performance Comparison

Table 2 and Figure 6 provide the results of a comparison that examines how well individual classifiers, including SVM and k-NN, perform. Following that, the efficiency of each classification layer will be measured, and lastly, the suggested ensemble approach will be evaluated to determine which has the highest accuracy.

V. CONCLUSION

We categorised leukaemia patients using a benchmark dataset with a binary class, the first value of which is Acute lymphocytic leukaemia (ALL), Acute myeloid leukaemia (AML), Chronic lymphocytic leukaemia (CLL), Chronic myeloid leukaemia (CML), and Chronic myeloid leukaemia (CML) (CML). The purpose of this study is to apply the Machine Learning approach to analyse leukaemia smear pictures in order to increase accuracy, decrease diagnosis time, and deliver quicker, simpler, as well as better diagnostic services. The approach presented in this study is to remove the other components of blood from the leukocytes then extract the lymphocytes. Fractal characteristics, form features, and other textural properties are recovered from the extracted lymphocytes. The Machine Learning technique produces considerable results in identifying the forms of leukaemia by applying the Ensemble classification algorithm and/or regression. The Machine Learning technique is turned into development tools for diagnosing leukaemia in this suggested system, which would aid healthcare institutions in remote regions with less medical professionals, particularly in screening. It may also be changed to become open source software that is free for individuals to utilize and distribute.

VI. REFERENCES

[1]. Sos Agaian and Anthony T. Chronopoulos, "Automated Screening System for Acute Myeloid Leukaemia Detection in Blood Microscopic Images", IEEE SYSTEMS JOURNAL, vol. 8, no. 3, SEPTEMBER 2014.

[2]. Emad A. Mohammed, Mostafa M. A. Mohamed, Christopher Naugler and Behrouz H. Far, "Chronic lymphocytic leukaemia cell segmentation from microscopic blood images using watershed algorithm and optimal thresholding", 26th IEEE Canadian Conference of Electrical and Computer, 2013.

[3]. Fauziah Kasmin, Anton Satria Prabuwono, Azizi Abdullah, "Detection of Leukemia in Human Blood Sample Based on Microscopic Images: A Study", Journal of Theoretical and Applied Information Technology, Vol.46, No.2, pp.579-586, 2012.

[4]. Mostafa Mohamed, Behrouz Far, AmrGuaily, "An Efficient Technique for White Blood Cells Nuclei Automatic Segmentation", IEEE International Conference on Systems, Man and Cybernetics, COEX, Seoul, Korea, October 14-17, 2012.

[5]. S.N.Deepa and B.Aruna Devi, "A survey on artificial intelligence approaches for medical image classification"; IJST, vol 4, No.11 (Nov 2011).

[6]. P. K. Das and S. Meher, "An efficient deep convolutional neural network based detection and classification of acute lymphoblastic leukemia," Expert Syst. Appl., vol. 183, Nov. 2021, Art. no. 115311

[7]. P. K. Das, A. Pradhan, and S. Meher, "Detection of acute lymphoblastic leukemia using machine

learning techniques," in Machine Learning, Deep Learning and Computational Intelligence for Wireless Communication. Singapore: Springer, 2021, pp. 425–437.

[8]. P. P. Banik, R. Saha, and K.-D. Kim, "An automatic nucleus segmentation and CNN model based classification method of white blood cell", Expert Syst. Appl., vol. 149, Jul. 2020, Art. no. 113211.

[9]. 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, Jan. 2019.

[10]. R. B. Hegde, K. Prasad, H. Hebbar, B. M. K. Singh, and I. Sandhya, "Automated decision support system for detection of leukemia from peripheral blood smear images," Journal of Digital Imaging, vol. 33, pp. 361–374, 2019

[11]. L. H. Vogado, R. M. S. Veras, F. H. D. Araujo, R. R. V. Silva, and K. R. T. Aires, "Leukemia diagnosis in blood slides using transfer learning in CNNs and SVM for classification," Eng. Appl. Artif. Intell., vol. 72, pp. 415–422, Jun. 2018.

[12]. J. Zhao, M. Zhang, Z. Zhou, J. Chu, and F. Cao, "Automatic detection and classification of leukocytes using convolutional neural networks," Medical & Biological Engineering & Computing, vol. 55, no. 8, pp. 1287–1301, 2017

[13]. F. Xing and L. Yang, "Robust nucleus/cell detection and segmentation in digital pathology and microscopy images: a comprehensive review," IEEE Reviews in Biomedical Engineering, vol. 9, pp. 234–263, 2016.

[14]. R. G. Bagasjvara, I. Candradewi, S. Hartati, and A. Harjoko, "Automated detection and classification techniques of acute leukemia using image processing: A review," in Proc. 2nd Int. Conf. Sci. Technology-Computer (ICST), Oct. 2016, pp. 35–43.

[15]. Himali P. Vaghela, Hardik Modi, Manoj Pandya and M.B. Potdar, "Leukaemia Detection using Digital Image Processing Techniques", International Journal of Applied Information Systems (IJAIS), November 2015.

Cite this article as :

Arunthathi. S, Ramyaa Sri. S, Deepika.S, Baavana.S, "Leukemia Detection Using Ensemble Model in Machine Learning", International Journal of Scientific Research in Science, Engineering and Technology (IJSRSET), Online ISSN : 2394-4099, Print ISSN : 2395-1990, Volume 10 Issue 2, pp. 255-263, March-April 2023.

Journal URL : https://ijsrset.com/IJSRSET2310240