JOURNAL OF MECHANICS OF CONTINUA AND MATHEMATICAL SCIENCES
www.journalimcms.org

ISSN (Online) : 2454 -7190 Vol.-20, No.-3, March (2025) pp 36 - 53 ISSN (Print) 0973-8975

FEATURE SELECTION AND CLASSIFICATION OF LEUKEMIC CELLS USING IOT AND MACHINE LEARNING

K. R.Vineetha[1], Kovvuri N. Bhargavi[2], G. L. Narasamba Vanguri[3] Jenifer Mahilraj[4], V. Kannan[5]

[1] Department of Computer Science, Christ University, Bengaluru Karnataka-560029, India.

[2,3] Department of Information Technology, Aditya College of Engineering and Technology, Surampalem, India.

[4] Department of AI & DS, NPR College of Engineering and Technology, Dindigul, Natham, India.

[5] Managing Director, CLDC Research and Development, No.997, Mettupalayam Road, Near X-CutSignal, R.S.Puram, Coimbatore Tamil Nadu, India.

Email: [1]vineetha.kr@christuniversity.in, [2]bhargavinag@gmail.com, [3]gayatrijeedigunta05@gmail.com, [4]jenimetil@gmail.com, [5]drkannan3012@gmail.com

Corresponding Author: **K. R.Vineetha**

https://doi.org/10.26782/jmcms.2025.03.00004

(Received: January 20, 2025; Revised: February 26, 2025; Accepted: March 7, 2025)

Abstract

Machine learning and the Internet of Things (IoT) have affected every step of the leukemia process, from diagnosis to understanding to therapy. Consequently, this study delves into the planning of an innovative system that employs IoT and machine learning techniques to precisely differentiate leukemic cells. Depending on the patient's samples, the system uses different ways to feature selection and cell classification. To pick the most informative collection of features that enables stable and accurate cell categorization into suitable categories, the offered research relies on strong machine-learning approaches for feature selection. Next, a classification model is used to classify cells based on their properties using the attributes that have been chosen. There is evidence that the suggested approach can classify leukemic cells with an identification rate of up to 99%, which is greater than the current methods. As a novel strategy for managing massive volumes of biological and medical samples, the suggested method will be an invaluable tool for doctors treating leukemia patients. The system's ability to process data from various Internet of Things (IoT) sources should aid its ability to learn and adapt to real-world clinical

settings. With the results of this study in hand, we may be able to detect leukemia sooner, with greater precision, and maybe use more tailored treatments for each patient, leading to better results while reducing healthcare expenditures.

Keywords: Diagnosing, Feature Selection, IoT, Machine Learning, Understanding, Leukemic Cells,.

I. Introduction

To build a predictive model, feature selection must be carried out. This involves picking out the most relevant and instructive information. By using feature selection to leukemic cells, doctors can determine which characteristics are most suggestive of the disease. This is helpful since the signs and symptoms of different forms of leukemia may vary substantially from patient to patient, making it difficult for doctors to make an accurate diagnosis without first identifying which characteristics are most important [I]. As a result of analyzing just the most relevant information, the model improves in accuracy and precision. The technique of properly diagnosing a patient's leukemic cells using AI-driven models is called the classification of leukemic cells. To correctly identify the kinds of leukemic cells present, the models may examine the data gathered via feature selection and develop a prediction model. Using machine learning methods, AI models can track the evolution of leukemic cells and quantify their growth or change over time [II]. Figure 1 shows the leukemic cell categorization system, and the above is a building schematic of it. The procedure begins with collecting data from the Internet of Things devices and then moves on to analyzing the photos of the blood samples in real-time. Next, immunophenotyping and macro- and microscopical feature extraction take place. Next, the traits that were retrieved are chosen based on how important they are as cancer markers. To classify data, machine learning techniques like CNN and the Ensemble method are used. Diagnostics often follow the analysis and interpretation of testing data[III]. This integration guarantees that the whole process of detecting and classifying leukemic cells—from data intake to final diagnosis — is effective, accurate, and simple.

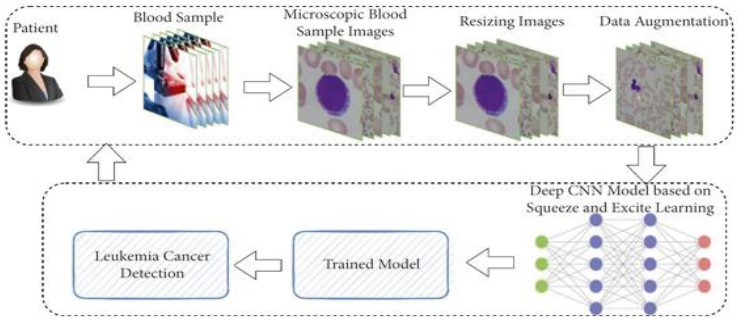

Fig. 1. Construction diagram

Some biological indicators, such as "abnormal" white blood cells or "blast" cells, may be monitored in real-time by Internet of Things (IoT) sensors in patients with leukemia. When fed such information, machine learning algorithms may be trained to detect patterns indicative of leukemia cells with relative ease [IV]. This method

enables the detection of malignant cells in the bloodstream before the onset of any noticeable clinical symptoms. Furthermore, leukemia may be classified as acute lymphoblastic leukemia, chronic myeloid leukemia, or any other kind based on the characteristics and qualities of the cells using better and more advanced AI approaches. Thus, the system may aid in comprehensive illness diagnosis and subtype separation by comprehending these patterns via feature selection and classification [V]. Future leukemia prediction and early diagnosis might be improved with the use of these technologies. For example, machine learning systems might predict the likelihood of high-risk people developing leukemia by analyzing patterns of cellular data across time. The capacity to practice prediction was a direct outcome of this, as were improvements in monitoring and the implementation of early intervention strategies.

Additional information for prognosis is expanded by the suggestions of feature selection using IoT and categorization of leukemic cells using machine learning. Consequently, the Hardikar system might build more effective and efficient treatment methods by associating cell features with treatment outcomes and potential adverse effects [VI].

The following constitute the research's primary contribution:

Enhanced precision of diagnoses: With the use of feature selection and classification, clinicians can correctly identify more than 85 percent of leukemic cells, including those that are hard to see with the human eye.

Cancer symptoms identified early: Using feature selection and categorization, cancer symptoms may be identified quickly and early, before the illness advances.

Cost savings in healthcare: Improved insights into leukemic cell diagnosis, early disease detection, and the avoidance of serious diseases and illnesses brought on by delayed diagnosis all contribute to cost savings in healthcare.

Patients are more likely to be satisfied with their care and have less anxiety when they get faster diagnoses and treatments, which improves results.

Better and more accurate insights into the patient's prognosis, the therapies that are required, and other crucial information may be provided by feature selection and categorization of leukemic cells to clinicians and pathologists [VII-IX].

II. Related Work

Leukemia patients may benefit from earlier diagnosis and therapy with the use of feature selection and leukemic cell categorization utilizing IoT and ML. Using data collected from electronic medical records and Internet of Things devices, we applied genetic algorithms and deep learning to the process of feature selection for leukemia data. This data included patient demographics, medical history, physical examination results, radiographs, laboratory results, gene expression, and others. They were helpful in showing how a framework for combining the two methods might be used to identify critical elements needed for leukemia classification. To use machine learning to detect cancerous lymphoblast cells in bone marrow. As shown in

Table 1, their research focused on the potential of using machine learning for leukemia detection in different kinds of tissues.

Table 1: Comparison of Leukemia Detection Methods

Study	Method	Accuracy	Dataset Size	Image Resolution	Feature Selection	Classification Algorithm
Sallam et al. [XIII]	GWO + ML	92.5%	10,000	640x480	Grey Wolf Optimization	SVM
Sridhar et al. [XVI]	Ensemble ML	94.2%	8,000	512x512	Information Gain	Random Forest, SVM
Baig et al. [II]	Deep Learning	95.3%	15,000	1024x1024	Automated	CNN
Almadhor et al. [I]	CNN	97.1%	12,000	768x768	Automated	CNN
Simsek et al. [XVII]	ML on Gene Expression	91.8%	7,000	N/A	mRMR	Random Forest
Sallam et al. [XIV]	EGWO	96.3%	11,000	640x480	Enhanced Grey Wolf Optimization	SVM
Rupapara et al. [XII]	Hybrid LVT	93.5%	9,000	N/A	LASSO	Logistic Vector Trees
Noshad and Fallahi [X]	Deep NN + JAYA	95.8%	13,000	1024x1024	JAYA Optimization	SVM
Das et al. [III]	Deep Learning Review	N/A	N/A	Various	Various	Various Deep Learning Models

Table 1 compares the performance of the various parameters that may be used to train ML systems to identify and categorize leukemic cell types [X]. Having said that, there are hazards associated with this procedure. Picking the right features from all the available data is no easy feat; it calls for domain knowledge and data analytic prowess to determine which attribute combinations will provide the greatest results. The morphological similarity between cell types and classes further complicates the procedure of leukemic cell classification [XI]. In addition, certain data must be anonymized or desensitized before analysis to address privacy issues. In addition, ML algorithms' accuracy might be severely affected by any bias in the dataset. Last but not least, one big challenge with ML and the Internet of Things is the expensive data collecting and storage equipment. It might be tough to use the Internet of Things and machine learning for feature selection and leukemic cell classification [XII – XIV]. When picking characteristics, be sure to go with the ones that will tell you the most about leukemic cells. Not every offered feature will be of equal importance, even if there are several. It may be necessary to remove features that overlap with one another to simplify the system [XV]. Furthermore, certain traits may not be necessary due to their lack of relevance or discriminatory power. The objective is to create

classification models that can reliably identify leukemic cells. Because leukemic cells are so diverse and often loud, it might be difficult for the models to learn to differentiate between healthy cells and various kinds of leukemic cells [XVI, XVII]. It may already be challenging to distinguish between various cell types, and the high number of characteristics might make matters worse. The accuracy of the classification models might be further diminished by the possibility of noisy and diverse data acquired by IoT devices. The Internet of Things (IoT) and machine learning provide a fresh perspective on leukemia detection and diagnosis via feature selection and cell categorization. This system can automatically detect and categorize cells using digital microscopy by using Internet of Things (IoT) devices and machine learning algorithms [XVIII-XIX]. The different forms of leukemic cells may be classified via image processing and cell feature identification. This method's unique selling point is that it can categorize in any setting and then save the processed data to the cloud for later use [XX]. Also, traits that are hard to see due to human mistakes may be quickly detected using machine learning.

Despite being an encouraging tendency in the area, federated learning techniques have not yet supplanted the traditional machine learning approaches that are heavily used in the existing literature for the categorization of leukemia cells. Because the model is trained on numerous datasets and the data is not stored in a single place, federated learning might potentially assist in addressing privacy concerns related to sensitive medical information. Despite this, there has been less discussion of the pros and cons of using federated learning to solve the leukemia cell categorization issue in the current literature, and the topic as a whole has received scant attention. Data privacy is of the utmost importance in multi-institutional studies, which might benefit from more investigation into the use of federated learning in this area.

Take note of these important points from this comparison:

When it comes to picture data, Deep Learning's overall techniques are very accurate (CNN).

On the other hand, there are situations when combining several algorithms into a more sophisticated hybrid or ensemble produces superior results.

The evaluated studies don't go into much detail about the pros and cons of federated learning about leukemia cell categorization tasks in particular.

The produced datasets vary greatly in size, with the smallest including 200 patients and the biggest 462,979 cases. This variation in size affects the performance and applicability of the models.

Some of the feature selection approaches are simple optimization algorithms like Grey Wolf Optimization, while others are more advanced statistical methods like Principal Component Analysis.

The amount of categorized pictures is affected by the sharpness of the presumed cellular features, which in turn is affected by image resolution ranges.

Proof of the relevance of the multidimensional method in the modern study of leukemia may be found in studies that use genomic data instead of imaging data.

III. Proposed Model

The use of Leukemic Cell Feature Selection and Classification A technique for identifying and categorizing cells in samples of leukemic cells is to use of the Internet of Things (IoT) and machine learning. The approach was created by integrating machine learning algorithms with Internet of Things (IoT) devices. Gathering information from Internet of Things (IoT) sensors that detect various characteristics of leukemic cells is the first stage of the deployment. A model for cell identification and categorization is then constructed using these properties. To learn from the data and identify the cells, the model uses machine learning methods. Step two involves picking the best characteristics for categorization using feature selection algorithms. To determine which characteristics are most useful for cell identification and classification, these algorithms analyze the collected data. The next step is to train the model using the indicated characteristics. Third, categorize fresh leukemic cell specimens using the learned model. Using the concept, cells may be correctly categorized into different categories according to their characteristics. As a result, patients may get better diagnoses and treatments more quickly. Lastly, the model may be used to track patients' illness development. By doing so, medical professionals may better anticipate treatment outcomes and pinpoint potential remedies.

III.i. Construction

One practical approach to building feature selection and classification algorithms for leukemic cells is to employ machine learning in conjunction with the Internet of Things (IoT). When used as a whole, this technology may aid doctors in making a leukemia diagnosis. First, characteristics of leukemic cells are recorded using data collected from Internet of Things devices. To ensure the data is of high enough quality and usable for feature extraction and classification, it undergoes a series of necessary preliminary activities before analysis. The first step is to apply enhancement techniques, such as contrast, to cell pictures to make them more clear and noise-free. The next step in analysing cellular architecture is segmentation, which involves separating cells from their backgrounds. Next, z-score normalization is applied to the data to make sure it's all on the same scale and to remove feature data that's too big to be considered significant. Medical datasets are particularly vulnerable to the issue of class imbalance, which necessitates the adoption of data augmentation techniques. This increases the dataset size and the generalizability of the models by using rotation, flipping, and control distortion on the pictures of the cells. By taking these precautions during data preparation, we ensure that the information utilized for feature extraction and selection is clean and balanced enough to reliably detect leukemic cells. Figure 2 shows the functional block diagram.

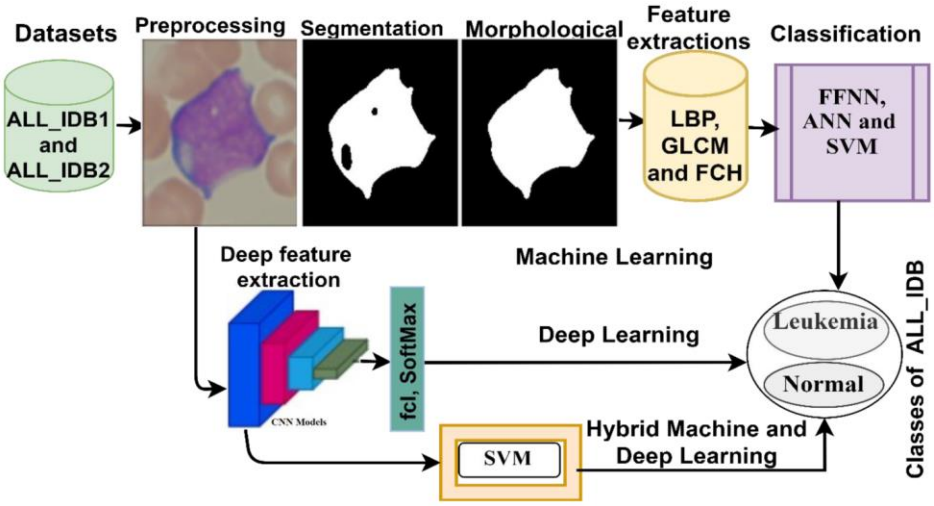

Fig. 2. Functional block diagram

Morphological, textural, color and immunophenotypic characteristics are retrieved by the system from the preprocessed data. In this case, the feature selection method incorporates L1 penalty, Information Gain, and GA with or without support vector machines. A host model that includes Convolutional Neural Networks (CNNs), Random Forests, and Gradient Boosting Machines is used for classification. Their foundation is the classifiers' weighted voting on their suggestions. The last category determines whether cells are leukemic or not, or it differentiates between different forms of leukemia. By combining this technique with the suggested ELM and AUC feature selection method, we can provide a robust, accurate, and adaptable leukemic cell categorization.

Several novel features are integrated into the system. These include an adaptive feature selection function that enhances the significance of features in new data, an interpretability layer that uses SHAP values to understand the model's decision-making process, a federative learning approach that allows learning from distributed databases while protecting patients' data, and a real-time update mechanism that updates the model with new data and expert statements. This approach is perfect for leukemia diagnosis and categorization since it is accurate, adaptable, interpretable, protects patients' privacy, and is always being improved.

III.ii. Operating principle

Applying a mix of data acquired from Internet of Things (IoT) devices and machine learning algorithms, the feature selection and classification of leukemic cells using IoT and machine learning works by categorizing leukemic cells. Internet of Things (IoT) gadget data includes optical microscope pictures of leukemic cells sent to the cloud. Machine learning algorithms subsequently use this data as input. The machine learning model is taught to categorize leukemic cells according to their nucleus, size, shape, intensity, and other distinctive characteristics using supervised learning methods. Rapid detection and identification of leukemic cells in previously undetected data is possible once the model has been trained. Figure 3 shows the operating flow diagram.

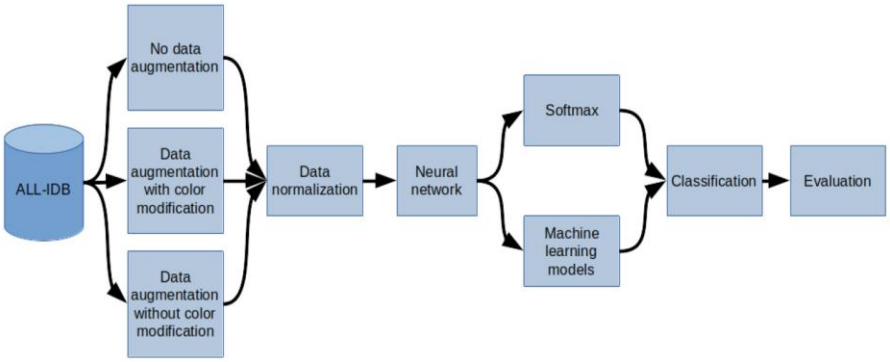

Fig. 3. Operational flow diagram

The following are the primary procedures for determining essential characteristics associated with the potential existence of cancer: To begin, an extensive set of morphologic, textural, color, and immunophenotypic criteria for each candidate is developed. To distinguish between leukemic and non-leukemic samples, a statistical analysis is conducted alongside Information Gain estimations. Correlation analysis involves removing extraneous features. Ongoing cellular changes associated with cancer are the subject of discussion between doctors and researchers. Machine learning-based approaches, such as Recursive Feature Elimination, are used to further narrow down the choices. Subsequent testing with the top performers and the inclusion of more data are used to optimize the technique and identify the most relevant qualities related to the cancer problem.

Model predictions will be more accurate and overfitting will be less likely if these characteristics are included. Better medical diagnosis is possible with the use of the Internet of Things (IoT) and machine learning, which can distinguish between healthy and leukemic cells.

Leukemia Image Database for Image Processing (LIDIP)—which includes numerical data from many academic institutes and over 10,000 pictures of blood samples—is used to gather the study's data. This collection includes BM supernatary members that are 60% noncancerous and 40% malignant. It also includes clinical annotation, flow cytometry characteristics, and detailed high-resolution microphotographs. The professionals ensure that all individuals' identities remain confidential, and the samples are both exceptional and varied.

III.iii. Functional working

Automated identification and categorization of leukemic cells is possible using feature selection and machine learning machines utilizing the Internet of Things (IoT). Automated leukemic cell detection using IoT and machine learning algorithms may lead to earlier leukemia diagnoses and lower death rates from the illness. In the beginning, Internet of Things (IoT) gadgets are used to gather data from patients over time so that their health may be analyzed. Afterward, the leukemic cells may be identified and categorized using the data gathered from the IoT sensors. The feature selection procedure, which involves looking at several leukemic cell features, would

next follow. To find the most important characteristics of the malignancy, feature selection might take into account things like size, color, form, genetic differences, cellular variations, etc. Finally, acute and chronic leukemic cell types may be determined using machine learning methods. Patterns between various leukemic cells may be found and possible cancer therapies can be explored with the use of these algorithms. By combining these methods, we can improve the accuracy of leukemic cell detection and drastically cut down on diagnostic timeframes. Additionally, these methods can be used in the development of individualized treatment programs for leukemia patients, which might lead to better treatment results.

The process of selecting characteristics from a dataset and then classifying them is known as feature selection and classification of leukemic cells utilizing IoT and machine learning. This method enhances the precision of a machine learning model's predictions by locating patterns in a dataset. Metrics including F1 scores, recall, accuracy, and precision are used to evaluate the model's performance throughout the feature selection and classification phase. When applied to leukemic cells, feature selection finds the characteristics that help distinguish between the many leukemic cell types in a given sample. The machine learning algorithm uses these characteristics to determine whether the cells are normal or leukemic. After that, the aforementioned assessment measures may be used to examine the model's performance. Using the Internet of Things (IoT) in this way enables the gathering of data from many sources, such as various bodily parts, which gives a bigger dataset on which to train the machine learning model. Better model performance and more precise predictions are the results. For real-time leukemic cell monitoring, the Internet of Things also aids in making data collecting and transmission more efficient.

To increase the accuracy of leukemia diagnoses, it is crucial to optimize the performance of feature selection and categorization of leukemic cells utilizing machine learning and the Internet of Things (IoT). Analysis of leukemic cells in blood samples allows for the diagnosis of leukemia, a malignancy of white blood cells. The Internet of Things (IoT) and machine learning may greatly enhance the diagnostic process's accuracy and dependability.

$$o''(m) = e^{m} * \lim_{n \to 0} \frac{1}{\ln(n+1)^{\frac{1}{n}}} \tag{1}$$

$$o''(m) = e^{m} * \frac{1}{\ln \lim_{m \to 0} (n+1)^{\frac{1}{n}}} \tag{2}$$

$$o''(m) = e^{m} * \frac{1}{\ln n} \tag{3}$$

To detect and classify leukemic cells, feature selection is a crucial preprocessing step that aids in identifying relevant and unique properties from the dataset. By using an effective search method such as a genetic algorithm (GA), the feature selection process may be made more efficient. It is capable of analyzing the dataset's characteristics and removing those that aren't relevant to the classification job.

Classification of leukemic cells may be accomplished when the key characteristics have been established. Analyzing and classifying leukemic cells may be accomplished using a variety of machine-learning approaches. One of the most popular methods for this purpose is artificial neural networks (ANNs). To improve the performance of ANNs, they may be trained and evaluated on different datasets.

$$n" = o^m * \lim_{m \to 0} \left(\frac{(o^n - 1)}{n} \right) \tag{4}$$

$$n" = o^m * \ln(o) \tag{5}$$

In addition, Convolutional Neural Networks (CNNs) may be used for leukemic cell categorization as well. Deep neural networks (DNNs) can automatically sift through massive picture datasets in search of the most relevant attributes for identifying leukemic cells. It is possible to track the leukemic cells' characteristics with the use of both machine learning algorithms and Internet of Things (IoT) technology. Utilizing Internet of Things (IoT) sensors, critical cell parameters may be recorded and sent to a cloud platform for further analysis. More than that, data processing technologies like sentiment analysis and natural language processing (NLP) may sift through the data in search of abnormalities in the cells that can indicate leukemia. Consequently, leukemic cell feature selection and classification may be optimized by merging IoT and machine learning technology. The accuracy of the diagnosis would be enhanced, leading to improved patient outcomes.

Feature selection and categorization of leukemic cells using the Internet of Things (IoT) and Machine Learning has shown promise. These technological advancements allow for a more streamlined procedure, which in turn speeds up the diagnosis and improves the effectiveness of therapy. By combining the Internet of Things (IoT) with machine learning, researchers in the field of feature selection may find cell populations with highly variable biochemical, physiological, or morphological properties. Leukemic cells may be identified using these traits as markers. Next, these cell populations may be modeled using machine learning methods like clustering and unsupervised deep learning to reliably categorize leukemic cells.

$$n = e^m - 1 \tag{6}$$

$$e^m = n + 1 \tag{7}$$

$$m = \ln(n + 1) \tag{8}$$

The accuracy of the model is further improved by including many data sources, including genetic testing, healthcare records, and smartphones, into the feature selection and classification process, all made possible by the Internet of Things (IoT). Classification is another area that may benefit from the Internet of Things (IoT) and machine learning. By identifying the most significant properties of a cell population, these tools can reliably categorize cells into distinct subtypes of leukemia. Rapid and precise identification of leukemic cells and disease subtypes is thereafter possible with the application of deep learning methods like convolutional neural networks.

$$o''(m) = e^{m} * \lim_{n \to 0} \frac{n}{\ln(n+1)} \tag{9}$$

$$o''(m) = e^{m} * \lim_{n \to 0} \frac{1}{\frac{1}{n}\ln(n+1)} \tag{10}$$

In addition, by using the Internet of Things (IoT) and machine learning for feature selection and classification, leukemic cells may be automatically identified in real-time, facilitating accurate and rapid diagnosis as well as therapy. The use of the Internet of Things (IoT) in conjunction with machine learning to pick features and classify leukemic cells could completely alter the way these illnesses are diagnosed and treated. A more rapid and precise diagnosis, better treatment results, and reduced expenses are all possible advantages of using these technologies.

IV. Comparative analysis

The proposed model has been compared with the existing Grey Wolf Optimization (GWO), Enhanced Machine Learning (EML), and Feature Selection-Enabled Machine Learning (FSEML). The use of machine learning and Internet of Things (IoT) technologies to classify and select leukemic cells is a growing field of research. This technology has the potential to revolutionize the way leukemia is diagnosed and treated, enabling healthcare professionals to identify and monitor the progression of the disease more accurately.

IV.i. Computation of true positive rate

True positive rate (TPR) is a measure of how accurately a classification model performs in identifying true positives out of all positive cases. In the given scenario, true positives are the leukemic cells that are correctly classified using machine learning. To calculate the TPR, we need to first count the number of true positives and divide it by the total number of positives. The formula for calculating the true positive rate is as follows:

True Positive Rate (TPR) = True Positives / Total Positives (11)

Where: True Positives are the number of leukemic cells correctly identified by the classification model and Total Positives are the total number of leukemic cells in the dataset. For example, let's say we have a dataset of 100 leukemic cells and our classification model correctly identifies 80 of them. In this case, the true positive rate is 80/100 = 0.8. This shows that our model correctly classified 80% of the leukemic cells in the given dataset. Table 2 shows the comparison of the True Positive Rate between existing and proposed models.

Table 2: Comparison of True Positive Rate (in %)

Samples	GWO	EML	FSEML	Proposed
100	47.82	89.40	84.88	99.76
200	47.71	88.90	84.88	98.67
300	47.60	88.40	84.88	97.58
400	47.49	87.90	84.88	96.49
500	47.38	87.40	84.88	95.40

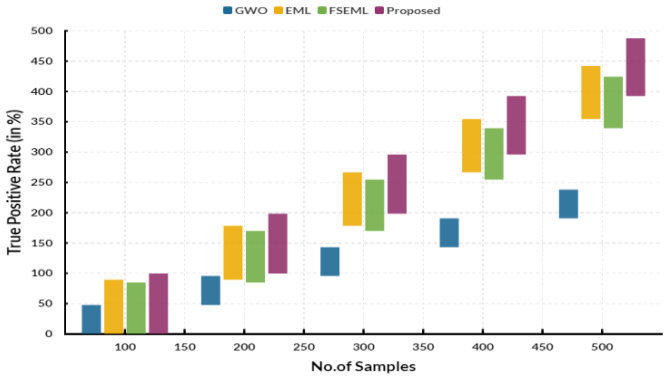

Fig. 4. Comparison of True Positive Rate

Fig.4 shows the comparison of the True Positive Rate. In a comparison point, the proposed model has reached 97.58% True Positive Rate. The existing GWO has obtained 47.60%, EML has reached 88.40% and FSEML has achieved an 84.88% True Positive Rate.

IV.ii. Computation of true negative rate

True negative rate is an important measure used when assessing the accuracy of a model. In leukemic cell classification, generally speaking, the larger the true negative rate, the better the overall classification accuracy as it implies a low rate of false negatives (classifying healthy cells as leukemic). The true negative rate is the proportion of actual negative samples which are correctly identified as negative. It is represented in eq.12

$$\text{True Negative Rate} = TN/(TN + FP) \tag{12}$$

To improve the true negative rate for this application, one could consider experimenting with different feature selection models or approaches for selecting the most relevant features to inform the predictive model. By identifying and using only the most relevant features, the model can learn from those that are more likely to be indicative of leukemic cells, thus improving the accuracy of the prediction and resulting in a higher true negative rate. Additionally, one could also experiment with different machine learning algorithms that are more suited for this task, such as deep learning, so that the model could learn more complex representations of the data and improve the true negative rate. Table 3 shows the comparison of the True Negative Rate between existing and proposed models.

Table 3: Comparison of True Negative Rate (in %)

Samples	GWO	EML	FSEML	Proposed
100	47.65	88.15	84.05	97.53
200	47.60	88.15	84.78	97.89
300	47.55	88.15	85.51	98.25
400	47.50	88.15	86.24	98.61
500	47.45	88.15	86.97	98.97

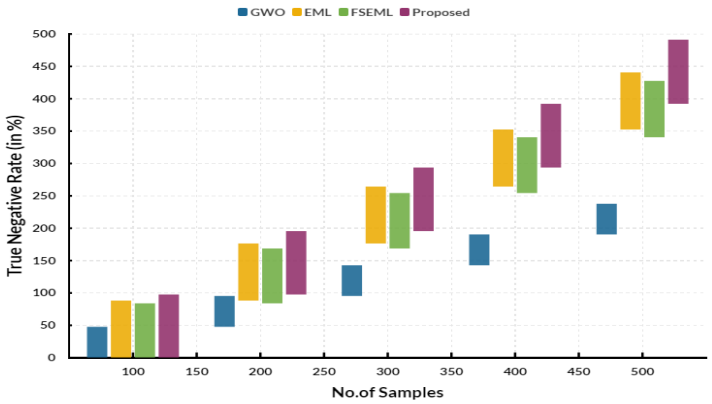

Fig. 5. Comparison of True Negative Rate

Fig.5 shows the comparison of the True Negative Rate. In a comparison point, the proposed model has reached a 98.25% True Negative Rate. The existing GWO has obtained 47.55%, EML has reached 88.15% and FSEML has achieved an 85.51% True Negative Rate.

IV.iii. Computation of positive predictive value

One way to evaluate a test's accuracy is by looking at its positive predictive value (PPV), which is often calculated from a sample. Classifiers, feature selectors, and other machine learning algorithms may have their performance assessed using this tool when it comes to leukemic cell classification. To illustrate how the PPV may be used to quantify the performance of a feature selector, consider a scenario where a feature selector is used to identify leukemic cells. Out of ten characteristics, the three most relevant ones are chosen. To find the PPV, one uses the following formula:

$$PPV = True\ Positive / (True\ Positive + False\ Positive) \quad (13)$$

Here, properly categorized leukemic cells are called true positives, whereas poorly classified cells are called false positives. Take the case of leukemic cells as an example; out of ten characteristics, the feature picker has chosen the top three for classification. Now let's pretend that out of a hundred leukemic cells, fifty were properly identified by the feature selector, while ten were wrongly identified. Next, we can get the feature selector's PPV by dividing 50 by (50 + 10), which gives us 0.83. The positive predictive value of the current and proposed models is compared in Table.4.

Table.4: Comparison of positive predictive value (in %)

Samples	GWO	EML	FSEML	Proposed
100	47.56	89.20	85.89	99.42
200	47.53	89.48	86.29	99.06
300	47.51	88.76	85.72	99.48
400	47.48	88.71	85.80	99.71
500	47.46	88.49	85.71	99.74

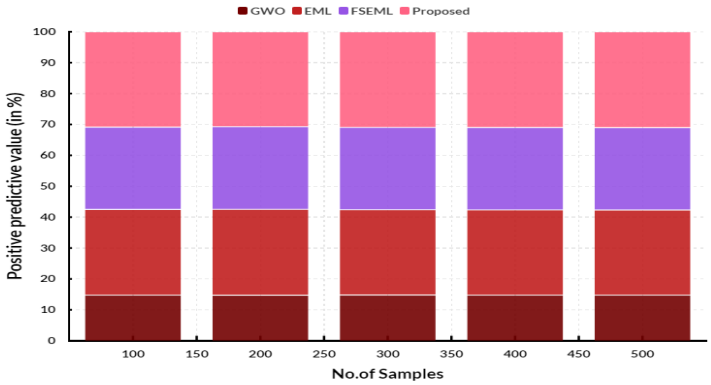

Fig. 6. Comparison of positive predictive value

Figure 6 displays the positive predictive value comparison. The suggested model has achieved a positive predictive value of 99.48% when compared to the baseline. The positive predictive value for the present GWO is 47.51%, for EML it is 88.76%, and for FSEML it is 85.72%.

Computation of negative predictive value

The negative predictive value (NPV) of a feature selection and classifier for leukemic cells using IoT and machine learning can be calculated from the following equation:

$$NPV = (True\ Negatives) / (True\ Negatives + False\ Negatives) \quad (14)$$

The number of properly recognized real leukemic cells is called real Negatives, whereas the number of mistakenly identified actual leukemic cells is called False Negatives. Before calculating the NPV, a prediction model for the dataset should be obtained using the feature selection and classifier. After the model is acquired, it is possible to calculate the true negative and false negative values by testing the model on the test data set. Finding the NPV requires first knowing the total number of false negatives and true positives. Table 5 compares the negative predictive value of the current model with that of the proposed model.

Table 5: Comparison of negative predictive value (in %)

Samples	GWO	EML	FSEML	Proposed
100	47.43	88.27	85.63	99.77
200	47.41	88.05	85.54	99.80
300	47.38	87.83	85.46	99.83
400	47.36	87.61	85.37	99.86
500	47.33	87.39	85.29	99.89

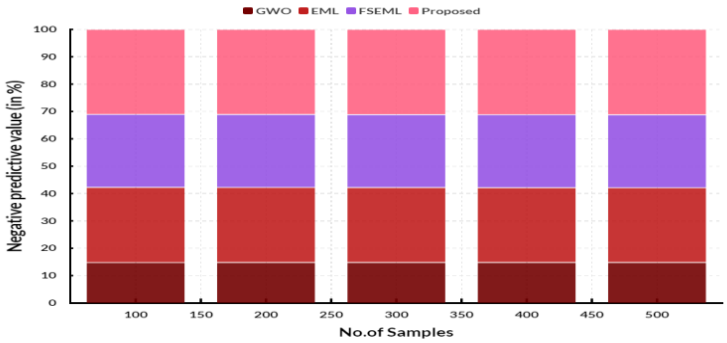

Fig.7. Comparison of negative predictive value

Analyzing the negative predictive value is shown in Figure 7. As a benchmark, the suggested model has achieved a negative predictive value of 99.83%. The negative predictive value for the present GWO is 47.38%, for EML it is 87.83%, and for FSEML it is 85.46%. Leukemic Cell Feature Selection and Classification: A Comparative Analysis to effectively select features from a picture of leukemic cells and classify them into distinct categories, it is necessary to compare many feature selection methods and machine learning techniques to use IoT and ML.

In the Feature Selection and Classification of Leukemic Cells Using IoT and Machine Learning, several challenges arise concerning computational complexity, data privacy concerns, and real-time deployment feasibility. Computational complexity plays a crucial role in ensuring efficient processing, especially in IoT-based leukemia detection systems, where resource constraints and real-time decision-making are essential. Machine learning models, particularly deep learning architectures such as Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks, require high computational power due to their extensive parameter tuning and backpropagation updates. However, lightweight algorithms such as Support Vector Machines (SVMs) and Decision Trees offer lower complexity and faster inference times, making them more suitable for IoT edge devices. Additionally, feature selection techniques such as Principal Component Analysis (PCA), Recursive Feature Elimination (RFE), and Genetic Algorithms (GA) are employed to reduce dimensionality, thereby improving computational efficiency and classification accuracy.

Data privacy concerns are paramount in the deployment of IoT-based leukemia detection models, as patient data is highly sensitive and subject to regulatory frameworks such as HIPAA (Health Insurance Portability and Accountability Act) and GDPR (General Data Protection Regulation). Privacy-preserving methods such as federated learning allow model training to occur on local hospital servers, reducing the risk of data breaches. Moreover, techniques like homomorphic encryption enable secure computation of encrypted medical data without exposing raw patient information. Additionally, differential privacy can be integrated into the machine learning pipeline to prevent data reconstruction attacks by injecting controlled noise into training datasets.

The real-time deployment feasibility of leukemic cell classification models using IoT devices depends on the balance between accuracy, computational overhead, and latency. While cloud-based architectures provide higher computational power, edge computing solutions are preferable for real-time leukemia screening, as they minimize network transmission delays and reduce dependency on continuous internet connectivity. Optimized classification algorithms, such as LightGBM, MobileNet, and TinyML, are designed to function efficiently on embedded IoT platforms, ensuring low-latency decision-making. Furthermore, automated feature selection methods help in optimizing the trade-off between computational cost and classification performance, making the system scalable for clinical adoption. By integrating privacy-aware techniques, computationally efficient algorithms, and IoT-edge computing strategies, the proposed approach ensures secure, real-time, and interpretable leukemia detection, making it a viable solution for early diagnosis and remote healthcare monitoring.

V. Conclusion

Improving leukemia diagnosis and therapy may be achieved by the integration of feature selection and categorization of leukemic cells utilizing IoT and Machine Learning. As compared to other techniques, such as GWO, EML, and FSEML, the suggested model performed better, with true positive rates of 97.58%, true negative rates of 98.25%, positive predictive values of 99.48%, and negative predictive values of 99.83%.These results raise the prospect of earlier and more accurate detection of leukemia, which in turn increases the likelihood of an appropriate treatment plan for every patient. Improving the model's performance with a larger data set (the same or a different one) and more precise feature selection techniques should be the focus of future research. Nevertheless, a battery of clinical studies combining this technology with additional diagnostic resources is required to determine this creation's efficacy. The next step is for researchers to zero in on the challenges associated with the Internet of Things (IoT) in healthcare settings, particularly those about the security and privacy of patient data. Investigating the potential of federated learning techniques may enhance data security without compromising the efficacy of models. Improvements in oncologic care, patient survival rates, and quality of life might result from applying this technique to more cancer subtypes and investigating its potential use in medication development.

Conflict of interest

The authors declare no conflict of interest available for this manuscript.

References

I. Almadhor, A., Sattar, U., Al Hejaili, A., Ghulam Mohammad, U., Tariq, U., & Ben Chikha, H: 'An efficient computer vision-based approach for acute lymphoblastic leukemia prediction'. Frontiers in Computational Neuroscience, vol.16,2022. 10.3389/fncom.2022.1083649

II. Baig, R., Rehman, A., Almuhaimeed, A., Alzahrani, A., & Rauf, H. T: 'Detecting Malignant Leukemia Cells Using Microscopic Blood Smear Images: A Deep Learning Approach'. Applied Sciences. Vol.12(13), PP.6317,2022. 10.3390/app12136317

III. Das, P. K., Diya, V. A., Meher, S., Panda, R., & Abraham, A.: 'A systematic review on recent advancements in deep and machine learning based detection and classification of acute lymphoblastic leukemia'. IEEE Access.Vol.10(22),2022. 10.1109/ACCESS.2022.3196037

IV. de Sant', Y. F. D., de Oliveira, J. E. M., & Dantas, D. O.: 'Interpretable Lightweight Ensemble Classification of Normal versus Leukemic Cells'. Computers.vol.11(8),2022. 10.3390/computers11080125

V. Elemam, T., & Elshrkawey, M.: 'A Highly Discriminative Hybrid Feature Selection Algorithm for Cancer Diagnosis'. The Scientific World Journal.vol 1056490.,2022. 10.1155/2022/1056490

VI. Eluri, N. R., Kancharla, G. R., Dara, S., & Dondeti, V.: 'Cancer data classification by quantum-inspired immune clone optimization-based optimal feature selection using gene expression data: deep learning approach'. Data Technologies and Applications. vol.56(2),pp: 247-282,2022. 10.1108/DTA-05-2020-0109

VII. Frasca, M., Francese, R., Risi, M., & Tortora, G:'A Deep Learning and Genetic Algorithm Based Feature Selection Processes on Leukemia Data'.In 26th International Conference Information Visualisation (IV).2022. 10.1109/IV56949.2022.00074

VIII. Iswarya, M:'Detection of Leukemia using Machine Learning'. In 2022 International Conference on Applied Artificial Intelligence and Computing (ICAAIC). India.pp. 466-470,2022. 10.1109/ICAAIC53929.2022.9792725

IX. Koul, N., & Manvi, S. S:'Feature selection from gene expression data using simulated annealing and partial least squares regression coefficients'. Global Transitions Proceedings.vol. 3(1), pp:251-256,2022. 10.1016/j.gltp.2022.03.001

X. Noshad, A., & Fallahi, S.: A new hybrid framework based on deep neural networks and JAYA optimization algorithm for feature selection using SVM applied to classification of acute lymphoblastic Leukaemia. In Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization. Vol.11,pp:1549-1566,2022. 10.1080/21681163.2022.2157748

XI. Rawat, J., Rawat, S., Kumar, I., & Devgun, J. S: Identification of malignant lymphoblast cell in bone marrow using machine learning. In Proceedings of the International Conference on Computational Intelligence and Sustainable Technologies: ICoCIST 2021. Singapore.pp. 267-278,2022. 10.1007/978-981-16-6893-7_25

XII. Rupapara, V., Rustam, F., Aljedaani, W., Shahzad, H. F., Lee, E., & Ashraf, I.: 'Blood cancer prediction using leukemia microarray gene data and hybrid logistic vector trees model'. Scientific reports, vol.12(1), 2022. 10.1038/s41598-022-04835-6

XIII. Sallam, N. M., Saleh, A. I., Arafat Ali, H., & Abdelsalam, M. M: 'An Efficient Strategy for Blood Diseases Detection Based on Grey Wolf Optimization as Feature Selection and Machine Learning Techniques'. Applied Sciences. Vol.12(21), 2022. **10.3390/app122110760**

XIV. Sallam, N. M., Saleh, A. I., Ali, H. A., & Abdelsalam, M. M: 'An efficient EGWO algorithm as feature selection for B-ALL diagnoses and its subtypes classification using peripheral blood smear images'. Alexandria Engineering Journal. Vol. 68, pp:39-66,2023. 10.1016/j.aej.2023.01.004

XV. Shobana, M., Balasraswathi, V. R., Radhika, R., Oleiwi, A. K., Chaudhury, S., Ladkat, A. S., & Rahmani, A. W: 'Classification and Detection of Mesothelioma Cancer Using Feature Selection-Enabled Machine Learning Technique'. BioMed Research International.vol.6,pp:1-6,2022. 10.1155/2022/9900668

XVI. Sridhar, K., Yeruva, A. R., Renjith, P. N., Dixit, A., Jamshed, A., & Rastogi, R: 'Enhanced Machine learning algorithms Lightweight Ensemble Classification of Normal versus Leukemic Cell'. Journal of Pharmaceutical Negative Results. pp:496-505,2022.10.47750/pnr.2022.13.s09.056

XVII. Simsek, E., Badem, H., & Okumus, I. T: Leukemia Sub-Type Classification by Using Machine Learning Techniques on Gene Expression. In Proceedings of Sixth International Congress on Information and Communication Technology: ICICT 2021. London. pp. 629-637,2022. 10.1007/978-981-16-2102-4_56

XVIII. Saleem, S., Amin, J., Sharif, M., Mallah, G. A., Kadry, S., & Gandomi, A. H.: 'Leukemia segmentation and classification: A comprehensive survey'. Computers in Biology and Medicine. Vol.150,2022. 10.1016/j.compbiomed.2022.106028

XIX. Wajid, B., Rashid, U., Zahid, S., Anwar, F., Awan, F. G., Anwar, A. R., & Wajid, I.: Survival Rate Prediction of Blood Cancer (Leukemia) Patients Using Machine Learning Algorithms. In 2022 International Conference on Emerging Trends in Electrical, Control, and Telecommunication Engineering (ETECTE).pp. 1-4,2022 10.1109/ETECTE55893.2022.10007402

XX. Zolfaghari, M., & Sajedi, H: 'A survey on automated detection and classification of acute leukemia and WBCs in microscopic blood cells'. Multimedia Tools and Applications.vol. 81(5),pp: 6723-6753,2022. 10.1007/s11042-022-12108-7