

RESEARCH ARTICLE

 OPEN ACCESS

Received: 15-10-2024

Accepted: 04-11-2024

Published: 26-11-2024

Citation: Alagu S, Kavitha G, Bagan KB (2024) Advanced Detection of Acute Lymphoblastic Leukemia Using Integrated Deep Features and Metaheuristic Algorithms. Indian Journal of Science and Technology 17(43): 4487-4493. <https://doi.org/10.17485/IJST/v17i43.3398>

* **Corresponding author.**alagupugal@gmail.com**Funding:** None**Competing Interests:** None

Copyright: © 2024 Alagu et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Published By Indian Society for Education and Environment ([iSee](https://www.isee.in))

ISSN

Print: 0974-6846

Electronic: 0974-5645

Advanced Detection of Acute Lymphoblastic Leukemia Using Integrated Deep Features and Metaheuristic Algorithms

S Alagu^{1*}, G Kavitha², K Bhoopathy Bagan³

¹ Assistant Professor, Department of Electronics and Communication Engineering, Sri Sai Ram Engineering College, Chennai, Tamil Nadu, India

² Professor, Department of Biomedical Engineering, CEG campus, Anna University, Chennai, Tamil Nadu, India

³ Professor, Department of Electronics Engineering, MIT Campus, Anna University, Chennai, Tamil Nadu, India

Abstract

Objectives: Acute lymphoblastic leukemia is one form of blood cancer. This research work suggests the impact of meta-heuristic feature optimization techniques on leukemia diagnosis. **Methods:** ALL-IDB2 image database is utilized for this purpose which is publicly available. Techniques for pre-processing images include scaling, SMOTE, and augmentation. Segmentation of the nucleus plays a vital role in detecting leukemia. As a novelty, two distinct deep networks, SegNet and ResUNet are deployed for this purpose in place of conventional segmentation techniques. ResUNet performs better than SegNet with a 0.98 accuracy rate. From the segmented images, about 6144 deep features are extracted using the Xception, Inception-v3, and ResNet50 models. The robust features are selected using mRmR and LASSO methods. The features from the LASSO method are taken for further processing based on the ANOVA test results. Utilizing the whale and moth flame optimization techniques, these features are optimized. SVM and Twin-SVM are used as classifiers in the proposed work. The distinctiveness of the work lies in its combination of metaheuristic algorithms and deep feature integration, which improves detection accuracy. **Findings:** It is found that the optimized features obtained from whale optimization give an accuracy of 0.9828 whereas the obtained accuracy is only about 0.9478 without using the optimization algorithm. **Novelty:** The integration of metaheuristic optimization and deep features is introduced in the proposed work to enhance leukemia detection accuracy. The suggested work will definitely help clinicians to make more accurate decisions on acute lymphoblastic leukemia diagnosis.

Keywords: Inception; Xception; LASSO; Feature concatenation; Whale Optimization; Twin-SVM

1 Introduction

White blood cells, specifically lymphocytes, are affected by a kind of malignancy known as acute lymphoblastic leukemia (ALL). One of its distinguishing features is an excess of immature white blood cells in the blood and bone marrow. The diagnosis of ALL involves a series of tests, including blood tests, bone marrow aspiration and biopsy, and imaging studies. The primary goal of this research work is to develop a unique approach that combines concatenated deep features with machine learning classifiers to detect acute lymphoblastic leukemia cells.

Recent research works are reviewed continuously which is used to identify the research gap in the process of detection of leukemia. A few of them are discussed in this section. Xiaodongchen et al⁽¹⁾ proposed a model with Deeplab v3 for the diagnosis of cancer Hybrid-based CNN system was proposed by Yesim Eroglu et al⁽²⁾ for the purpose of distinguishing benign, malignant, and normal cancer lesions. The Hybrid structure is built using the Alexnet, MobilenetV2, and Resnet50 devices. The complexity of the hybrid architecture leads to excessive computational time. An efficient approach to feature selection using a fused lasso framework and graph-based feature representations was proposed by Lixin Cui et al⁽³⁾. This approach could improve the trade-off between each feature's significance and the pairwise feature redundancy.

Jan Zidek et al⁽⁴⁾ proposed a technique of leukemia detection based on human visual perception. It leads to a decrease the detection accuracy which was the major challenge identified in this work. Nature-inspired metaheuristic algorithms were also introduced for the diagnosis of leukemia which gives significant results^(5–7).

According to studies in the literature, leukemia cells are detected using CNN models^(8,9), but the detection accuracy is low. In several recent research works, hybrid designs were developed to increase the detection accuracy. The system will become even more complex as a result. The development of a system with a basic architecture and deep features is identified as a research gap. To bridge this gap, the proposed work suggests metaheuristic algorithms to choose robust and optimized features that both increase detection accuracy and decrease system complexity.

2 Methodology

This research presents an effective method for the detection of acute lymphoblastic leukemia cells. For this purpose, the ALL_IDB2 dataset (<https://scotti.di.unimi.it/all/>) with 130 microscopic images each of blast and healthy white blood cells are gathered. Figure 1 depicts the block diagram of the suggested work. It is composed of several blocks, including segmentation, pre-processing images, feature extraction, feature selection, feature optimization, and classification.

2.1 Image pre-processing and segmentation

The suggested method handles image pre-processing through augmentation, SMOTE (Synthetic Minority Oversampling Technique), and scaling. All the input images are rescaled to 257 x 257 pixels. SMOTE oversamples minority classes by taking each sample of the minority class and appending synthetic examples along the line segments connecting all k nearest neighbors of the minority class. The cell nucleus is separated from the picture using SegNet and ResUNet. SegNet is a semantic segmentation model comprised of a pixel-level classification layer, followed by an encoder and decoder network.

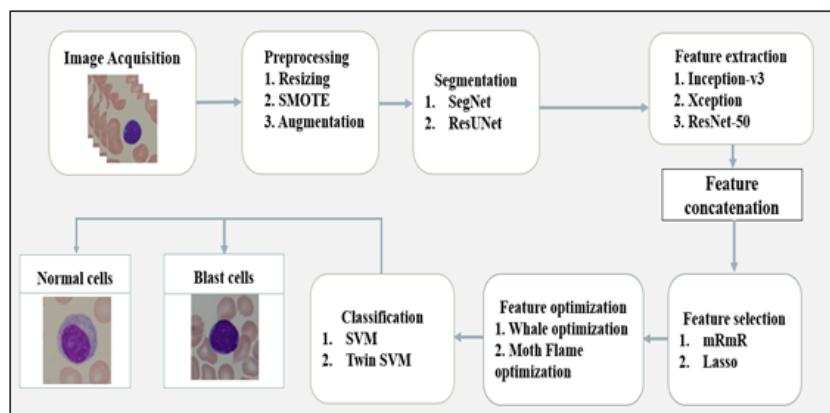


Fig 1. Block schematic of the proposed work

2.2 Feature Extraction

Extraction of features is the process of converting unprocessed data into manageable numerical features while maintaining the information present in the source data collection. Deep learning models like Inception-v3, Xception, and ResNet-50 are used for feature extraction from segmented images. The Inception-v3 model consists of both symmetric and asymmetric building components⁽¹⁰⁾. The inception modules used at each step reduce the number of parameters and channels by replacing larger filters at each level with smaller ones⁽¹¹⁾. Xception is an architecture that developed from the Inception design and is based on the idea of depth-wise separable convolutions⁽¹²⁾. Totally, 6144 features are obtained by concatenating 2048 features from Inception-v3, Xception, and ResNet-50.

2.3 Feature Selection

Feature selection is a method that helps you reduce the input variable of your model by removing noise and only using pertinent data. For feature selection, the Lasso technique and mRmR are employed. With the goal of maximizing relevance to the target variable and minimizing redundancy among the selected characteristics, mRmR seeks to extract a subset of pertinent features from a broader collection of variables. The relationship between each feature and the target variable is quantified by relevance⁽¹³⁾. There is a max-relevance in Equation (1), where S is the size of features, and $I(x_i; c)$ is the mutual knowledge between individual traits (x_i) and classes (c).

$$D = \frac{1}{S} \sum_{x_i \in S} I(x_i; c) \quad (1)$$

Redundancy quantifies the similarity or overlap between features. Equation (2) displays the minimal redundancy condition.

$$R = \frac{1}{|S|^2} \sum_{x_i, x_j \in S} I(x_i, x_j) \quad (2)$$

Where the mutual information between the x_i and x_j features is represented by $I(x_i, x_j)$. The mRmR algorithm aims to select a subset of features that maximizes the objective function shown in Equation (3).

$$\max_{\varphi} (D, R) \varphi = D - R \quad (3)$$

Lasso regression is a type of linear regression that adds a penalty term to the ordinary least squares (OLS) objective function in order to accomplish both feature selection and regularization⁽¹⁴⁾. Mathematically, the lasso regression problem can be formulated as shown in Equation (4).

$$\text{Minimize} : 1/(2n) * ||Y - X\beta||^2 + \lambda * ||\beta|| \quad (4)$$

Where,

- Y Represents a vector of the observed target values,
- X Denotes the matrix of feature values,
- β Represents the vector of coefficients for the linear regression model,
- λ Is the regularization parameter and
- n Notifies the number of data points

2.4 Feature Optimization

The process of selecting and altering the input features to improve the model's functionality and predictive capacity is known as feature optimization. For this, the moth flame optimization (MFO) and whale optimization algorithm (WOA) are employed in this research. Humpback whales use their exploring behavior to look for possible prey. This behavior is represented by randomly exploring the search space to locate potential solutions. Equations of Exploration are shown in Equations (5) and (6) and equations of exploitation are shown in Equations (7) and (8),

$$\vec{D} = |\vec{C} \cdot \vec{X}_{rand} - \vec{X}| \quad (5)$$

$$\vec{X}(t+1) = \vec{X}_{rand} - \vec{A} \cdot \vec{D} \quad (6)$$

$$\vec{D} = |\vec{C} \cdot \vec{X}^*(t) - \vec{X}(t)| \quad (7)$$

$$\vec{X}(t+1) = \vec{X}^*(t) - \vec{A} \cdot \vec{D} \quad (8)$$

In this case, X^* denotes the best solution found, X_{rand} is a random solution drawn from the current population, t denotes the current iteration, X is the current solution, A and C are coefficient vectors, and r is a random vector. The behavior of moths and their attraction to light sources served as the model for the meta-heuristic optimization technique known as Moth Flame Optimization (MFO). The MFO algorithm imitates the movement of moths toward light and utilizes a mathematical model to guide the search for an optimal solution. Equations (9) and (10) show how the moth changes its position in relation to a particular flame.

$$S(M_i, F_j) = D_i \cdot e^{bt} \cdot \cos(2\pi t) + F_j \quad (9)$$

$$D_i = |F_j - M_i| \quad (10)$$

The logarithmic spiral's shape is defined by the constant b , the i^{th} moth is denoted by M_i , the j^{th} flame is represented by F_j , and t is a random number in the interval $[-1, 1]$. The distance of the i^{th} moth from the j^{th} flame is indicated in Equation (10).

2.5 Classification

For classification, Support Vector Machines (SVM) and Twin-Support Vector Machines (Twin-SVM) are employed in the suggested research work. Twin-SVM uses two non-parallel hyperplanes to determine the decision boundary as opposed to the single hyperplane utilized in typical SVMs. The objective is to find these two hyperplanes while maximizing the margin and minimizing the classification errors.

$$\text{Minimize} : \frac{1}{2}(w_1 \cdot w_1 + w_2 \cdot w_2) + C_1 \xi_1 + C_2 \xi_2 \quad (11)$$

Where, w_1 and w_2 represent the weight vectors perpendicular to the respective hyperplanes, b_1 and b_2 are the corresponding bias terms, C_1 and C_2 are the regularization parameters for the two hyperplanes and the variables ξ_1 and ξ_2 are slack variables that allow for misclassification errors.

3 Result and Discussion

Deep learning networks are utilized in the proposed work to segment the nucleus. In order to identify malignant cells, the segmented nucleus images are given to classifiers, after feature extraction and feature selection. The 130 tiny blood cell images in each of two classes such as cancer cells and healthy cells. ALL-IDB2 database is the source of the required input images with a size of 257 x 257 pixels. Totally 1458 images are obtained after augmentation.

Figure 2 presents the segmented nucleus of both healthy and cancerous cells using both SegNet and ResUNet along with ground-truth images. Table 1 shows the performances of deep networks such as SegNet and ResUNet. It is found that ResUNet is made to achieve great performance with a minimum number of parameters.

Xception, ResNet-50, and Inception-v3 are used to extract the deep features from segmented images. 2048 features are extracted from each model and combined into a single feature vector. The validation of all features is carried out by 'ANOVA test'. The metric known as 'p-value' can be used to identify the best features. The features with lower 'p-values' give a better classification.

The extracted deep features are optimized by metaheuristic algorithms such as mRmR (minimum redundancy maximum relevance), LASSO, Whales, and mothflame optimization techniques. ANOVA test is carried out again for the significance of the selected features. The p-value of the ANOVA test shows the statistical difference between the means of the two groups. If the p-value is less than 0.05, it will be simple to identify a significant difference between the two groups. It is observed that the features selected using mRmR are more significant. This is because the feature selection procedure in mRmR explicitly takes

Type of cell	Input images	Actual Ground truth	SegNet output	Segmented images by SegNet	Predicted images by ResUNet	ResUNet output
Healthy cells						
Healthy cells						
Cancer cells						
Cancer cells						

Fig 2. Segmentation results

Table 1. Performance comparison of SegNet and ResUNet

Model	Accuracy	Validation Accuracy	Loss	Validation loss	DICE Coefficient
SegNet	96.3%	94.99%	6.2%	3.6%	97.2%
ResUNet	97.21%	97.96%	3.4%	1.1%	98.1%

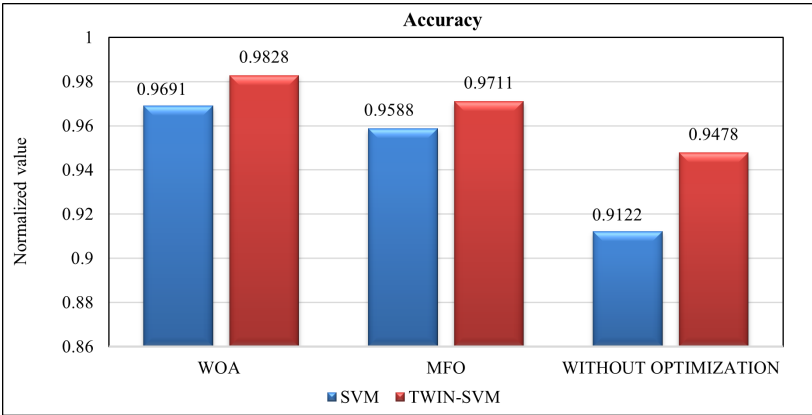


Fig 3. Performance comparison of classifier metrics with and without Optimization

redundancy considerations into account. While LASSO can effectively shrink the coefficients, it may not fully account for the redundancy among features. This means that LASSO may select multiple correlated features.

The classifiers utilized are SVM and Twin-SVM which can be used to classify the features of Whale Optimization, Moth Flame Optimization, and Without Optimization.

Figure 3 compares the accuracy of the SVM and Twin-SVM classifiers with and without optimization. Whale optimization has performed better than Moth flame optimization as Whale optimization has a good balance between exploration and exploitation. Whale optimization has been reported to converge quickly to a satisfactory solution due to its diverse search mechanisms. As per the performance of the classifier, Twin-SVM has a higher performance than SVM. The use of two parallel hyperplanes in Twin SVM provides a more flexible representation of the data. Identifying the blast cells more precisely will be aided by the proposed effort.

A comparison is made between the proposed methodology with some of the recent works performed to detect acute lymphoblastic leukemia based on the detection accuracy, algorithm complexity, and running time. It is observed that in recent works, a single deep network is adopted where the running time and system complexity are much less when compared to the proposed work. VAR-PCANet has moderate complexity as well as running time. Even though the suggested method in this paper has high complexity and more running time, the accuracy is higher when compared to other models. Always there is a trade-off between running time and accuracy. Regarding medical diagnosis problems, accuracy is given more preference than system complexity and running time. In that way, the proposed work is suggested as the best approach to detect leukemia.

Table 2. Performance comparison of proposed work with existing methods

Authors	Methodology	Accuracy	Algorithm complexity	Running time
Deepika Kumar et al ⁽¹⁵⁾	DCNN	97.2%	Less	1min 5sec
Mondal et al ⁽¹⁶⁾	XceptionV3	93.90%	Less	2min 20sec
Genovese et al ⁽¹⁷⁾	VAR-PCANet	97.62%	Medium	5min 30sec
Pradeep Kumar et al ⁽¹⁸⁾	MobilenetV2	97.18%	Less	2min 10sec
Proposed method	Inception-v3+Xception+ ResNet-50+LASSO+Whale optimization+Twin-SVM	98.28%	High	8min 10sec

4 Conclusion

The most prevalent type of malignancy in white blood cells is acute lymphocytic leukemia. The immature cancer cells will spread quickly. The primary objective of this research project is to create a high-accuracy diagnostic technique for acute lymphoblastic leukemia. Meta-heuristic feature optimization algorithms can be utilized to do this. All of the input images are taken from the publicly accessible ALL-IDB2 database for this purpose. Every image that is obtained is pre-processed using various techniques, including augmentation, SMOTE, and scaling. SegNet and ResUNet are utilized for segmentation when the pre-processed results are provided. ResUNet performs better than SegNet, with a 98.26% accuracy rate. Three deep learning methods—Inception-v3, Xception, and ResNet-50 are applied to the segmented findings in order to extract deep features. By concatenating all of the deep features, a total of 6144 features are acquired. By applying the LASSO approach, the robust features are chosen. Subsequently, the selected features are optimized by the Whale optimization technique. Blast cell detection is accomplished via a twin-SVM classifier. Accuracy, precision, recall, and F1-score are a few of the performance indicators that are used to compare classifier performance. The detection accuracy is found to be 0.9828 when the whale optimization technique is used, compared to just roughly 0.9478 when no optimization algorithm is used. This demonstrates the impact of applying a meta-heuristic feature optimization approach on the acute lymphoblastic leukemia diagnosis process. The suggested research will undoubtedly fill the research gap and also help clinicians make better-informed decisions regarding acute lymphoblastic leukemia diagnosis. The suggested approach will thereafter be evaluated using real-time images from hospitals and health institutions.

References

- 1) Xiaodongchen Q, Wu R, Yang Z. Segmentation of lung computed tomography images based on SegNet in the diagnosis of lung cancer. *Journal of Radiation Research and Applied Sciences*. 2021;14(1):396–403. Available from: <https://doi.org/10.1080/16878507.2021.1981753>.
- 2) Eroglu Y, Yildirim M, Cinar A. Convolutional Neural Networks based classification of breast ultrasonography images by hybrid method with respect to benign, malignant, and normal using mRMR. *Computers in Biology and Medicine*. 2021;133:104407–104407. Available from: <https://doi.org/10.1016/j.combiomed.2021.104407>.

- 3) Cui L, Bai L, Wang Y, Philip S, Edwin Y, Hancock R. Fused lasso for feature selection using structural information. *Pattern Recognition*. 2021;119. Available from: <https://doi.org/10.1016/j.patcog.2021.108058>.
- 4) Zidek J, Bodzas A, Kodytek P. Automated Detection of Acute Lymphoblastic Leukemia from Microscopic Images Based on Human Visual Perception. *Frontiers in Bioengineering and Biotechnology*. 2020;8:1–13. Available from: <https://doi.org/10.3389/fbioe.2020.01005>.
- 5) Cui EH, Zhang Z, Chen CJ. Applications of nature-inspired metaheuristic algorithms for tackling optimization problems across disciplines. *Sci Rep*. 2024;9403(14):1–10. Available from: <https://doi.org/10.1038/s41598-024-56670-6>.
- 6) Kaur S, Kumar Y, Koul A. A Systematic Review on Metaheuristic Optimization Techniques for Feature Selections in Disease Diagnosis: Open Issues and Challenges. *Arch Computat Methods Eng*. 2023;30:1863–1895. Available from: <https://doi.org/10.1007/s11831-022-09853-1>.
- 7) Ekinici E, Garip Z, Serbest K. Meta-heuristic optimization algorithms-based feature selection for joint moment prediction of sit-to-stand movement using machine learning algorithms. *Computers in Biology and Medicine*. 2024;178:5–15. Available from: <https://doi.org/10.1016/j.combiomed.2024.108812>.
- 8) Ram M, Afrash MR, Moulaei K. Application of artificial intelligence in chronic myeloid leukemia (CML) disease prediction and management: a scoping review. *BMC Cancer*. 2024;24:1026–1026. Available from: <https://doi.org/10.1186/s12885-024-12764-y>.
- 9) Alabdulqader EA, Alarfaj AA, Umer M. Improving prediction of blood cancer using leukemia microarray gene data and Chi2 features with weighted convolutional neural network. *Sci Rep*. 2024;2:15625. Available from: <https://doi.org/10.1038/s41598-024-65315-7>.
- 10) Diakogiannis F, Waldner F, Caccetta P, Chen WuResUNet-a: A deep learning framework for semantic segmentation of remotely sensed data. *ISPRS Journal of Photogrammetry and Remote Sensing*. 2020;162:94–114. Available from: <https://doi.org/10.1016/j.isprsjprs.2020.01.013>.
- 11) Wang C, Deleichen, Linhao, Xuebolu, Yuzeng J, Zhang G. Pulmonary Image Classification Based on Inception-v3 Transfer Learning Model. *IEEE Access*. 2019;7:146533–146541. Available from: <https://doi.org/10.1109/ACCESS.2019.2946000>.
- 12) Sharma S, Kumar S. The Xception model: A potential feature extractor in breast cancer histology images classification. *ICT Express*. 2022;8(1):101–108. Available from: <https://doi.org/10.1016/j.icte.2021.11.010>.
- 13) Eroglu Y, Yildirim M, Cinar A. Convolutional Neural Networks based classification of breast ultrasonography images by hybrid method with respect to benign, malignant, and normal using mRmR. *Computers in Biology and Medicine*. 2021;133:1–9. Available from: <https://doi.org/10.1016/j.combiomed.2021.104407>.
- 14) Cui L, Bai L, Wang Y, Philip S, Yu ER. Fused lasso for feature selection using structural information. *Pattern Recognition*. 2021;119:1–14. Available from: <https://doi.org/10.48550/arXiv.1902.09947>.
- 15) Kumar D, Jain N, Khurana A, Mittal S, Satapathy SC, Senkerik R, et al. Automatic Detection of White Blood Cancer from Bone Marrow Microscopic Images Using Convolutional Neural Networks. *IEEE Access*. 2020;8:142521–142531. Available from: <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9149873>.
- 16) Mondal MC, Hasana MK, Jawada T, Duttal MA, Islama MR, Awalc A, et al. Ensemble of Convolutional Neural Networks to diagnose Acute Lymphoblastic Leukemia from microscopic images. *Informatics in Medicine unlocked*. 2021;27:1–14. Available from: <https://doi.org/10.1016/j.imu.2021.100794>.
- 17) Genovese MS, Hosseini V, Piuri KN, Plataniotis F, Scotti. Acute Lymphoblastic Leukemia Detection Based on Adaptive Unsharpening and Deep Learning. *IEEE International Conference on Acoustics, Speech and Signal Processing*. 2021. Available from: <https://doi.org/10.1109/ICASSP39728.2021.9414362>.
- 18) Das PK, Diya VA, Meher S, Panda R, Ajith Abraham A. Systematic Review on Recent Advancements in Deep and Machine Learning Based Detection and Classification of Acute Lymphoblastic Leukemia. *IEEE Access*. 2022;10:81741–81763. Available from: <https://doi.org/10.1109/access.2022.3196037>.