

Classification of Acute Lymphocytic Leukemic Blood Cell Images using Hybrid CNN-Enhanced Ensemble SVM Models and Machine Learning Classifiers.

Saranya Vijayan¹, Radha Venkatachalam²

¹Avinashilingam Institute of Home Science and Higher Education for Women, Dept. of Computer Science, Coimbatore, India

²Avinashilingam Institute of Home Science and Higher Education for Women, Dept. of Computer Science, Coimbatore, India

saranyav101@gmail.com¹, 19phcsf003@avinuty.ac.in²

Abstract: Acute Lymphocytic Leukemia is a dangerous kind of malignant cancer caused due to the overproduction of white blood cells. The white blood cells in our body are responsible for fighting against infections, if the WBC increases the immunity will decrease and it would lead to serious health conditions. Malignant cancers such as ALL is life threatening if the disease is not diagnosed at an early stage. If a person is suffering from ALL the disease needs to be diagnosed at an early stage before it starts spreading, if it starts spreading the person's chances of survival would also reduce. Here comes the need of an accurate automated system which would assist the oncologists to diagnose the disease as early as possible. In this paper some of the algorithms that are enhanced to detect and classify ALL are incorporated. In order to classify the Acute Lymphocytic Leukemia a hybrid model has been deployed to improve the accuracy of the diagnosis and it is termed as Hybrid CNN Enhanced Ensemble SVM for the classification of malignancy. Machine Learning classifiers are also used to design the system and it is then compared with enhanced CNN based on the performance metrics.

Keywords : Acute Lymphocytic Leukemia , Ensemble System, Hybrid CNN-SVM.

I. INTRODUCTION

Digital Image Processing has been proved for solving many challenging problems in the field of medical images that could greatly contribute in the discovery of diseases and it would provide the physicians with valuable inputs in the process of diagnosing the diseases without any flaws. This paper proposes a hybrid model that uses Enhanced SVM and CNN for the classification of Acute Lymphocytic Leukaemia (ALL) . To identify malformations in white blood cells, generally, a manual inspection is carried out by an expert pathologist. Manual inspection and identification has several drawbacks like being time-consuming, high cost as expert pathologists are expensive and diagnosis accuracy depends on the experience and workload of the expert. To avoid these drawbacks, an automatic system is preferred by pathologists, who use these systems to help them aid in the correct identification of the disease. Moreover, automated systems help to avoid or reduce human intervention during diagnosis and are cost-effective.

Convolutional neural networks is one of the efficient deep neural networks known as CNN. A CNN will extract features automatically without any human intervention. CNNs will eradicate the need for extracting features manually and it removes the need to identify the features that used to classify

images. The CNN will directly extract features from images given as input. The features are not pretrained whereas they are learned when the network trains the images thereby making the automatic features extracted by deep learning models are accurate for challenging tasks in image processing.

Leukemia is also known as blood cancer, in which it would result in the rapid overproduction of abnormal white blood cells. Blood cancer occurs when the white blood cells in the bone marrow quickly increase and will eventually end up in destroying normal blood cells. It is considered to be the 11th top cancer type worldwide [Lin *et al.*, 2021]. In order to reduce death related to leukemia, it is important to treat them at the early stage.

An automatic ALL classification (ALL-C) system consists of four steps, namely, preprocessing, identification of white blood cells, feature extraction and classification. [Bukhari *et al.*, 2022; Mustafa *et al.*, 2022]. The proposed system is designed by enhancing the working of each step of the automated system and then combining them to further improve the performance of the ALL-C system.

“To design an Automatic ALL Classification System of the form

$$\mathcal{R} = DC + EMC$$

where \mathcal{R} is the output indicating the classification of ALL using hybrid deep learning and ensemble machine learning classifiers. The classification operation is of the form

$$\mathcal{R} = \oplus(m_i) \rightarrow \rho(m_i) \rightarrow \zeta(m_i)$$

where \rightarrow denotes the sequential application of operations. The classification output is any one of the pre-defined target label set {Normal, L1, L2, L3} and is to be performed using machine learning and deep learning classifiers.”

Extraction of white blood cells from microscopic image is the most important and challenging task in ALL (Acute Lymphoblastic Leukemia) detection and classification. The challenges arise mainly because of the high variations of cells in shape, size, edge, and position.

Each microscopic blood cells image has three main colors,

-Blue, which Indicates White Blood Cells (WBC)

-Red, it Indicates red blood cells

-Gray-white, Indicates background

Separation of these three cells is vital and are performed using image segmentation algorithms. Segmentation is defined as a task that partitions an image into disjoint and homogeneous regions based on some characteristic of the image. The main objective here is to create stable segments that are less sensitive to parameter changes.

To accomplish this problem statement, the primary research objective was set to strengthen the clinical decision support system by designing an automatic system that enhances the operation of each step involved during ALL-C in order to increase the overall accuracy and speed of leukemia classification.

II. LITERATURE SURVEY

Xiang li et al. (2018) have investigated a method for the classification of blood cell, to segregate white blood cells and red blood cells. They have used deep convolutional neural networks.

A method was analysed by Khot s et al. (2013) in which they have extracted the features from the images and had applied it to the classifier.

Himali et al. (2015) have investigated that when compared with watershed transform , histogram equalizing methods and k means clustering, the shape based features were more accurate for counting leukemic cells.

Khot s et al. (2013) have) used Support Vector Machine. They have extracted the features from the images and had applied it to the classifier.

Emad A. Mohammed, Mostafa M.A. Mohammed et al. (2017) have adopted a method for the cell segmentation of leukemia cells. In their research work they have used otsu method by using an optimal threshold value. They have also performed canny edge detection. The dilation and erosion were also carried out and the pixels that are isolated were eliminated and they have derived a segmented nucleus.

Subrajeet Mohapatra, Dipti Patra and et al. (2017) have examined a method known as color based clustering for the segmentation of the images of blood. They have done a comparison on the performances of some of the standard clustering techniques. The clustering techniques were k Means, FCM and FPCM. They have also used contour signature and hausdroff dimension for finding out the irregularities of the boundary of the nucleus. SVM classifier has been used to derive the results.

Sonal G. Deore Prof. Neeta Nemade et al. (2015) have proposed a method in which they have extracted the lymphocyte cells followed by extracting morphological indexes and then classification was done. They have identified the single cells by enhancing the input image. The filter used was adaptive pre filtering. The second step of their research work was to identify the white cells by separating it from other components of blood. The third step was identifying the lymphocytes associated with the other white cells. The accuracy of their research work was 93.63%

Acute leukemia grows rapidly and becomes severe within a short period, while chronic spreads slowly and takes longer to reach the advanced stage. Leukemia Classification is based on the type of white blood cells involving myeloid or lymphoid.

Himali et al. (2015) have identified that when compared with watershed transform , histogram equalizing methods and k means clustering, the shape based features are more accurate for counting leukemic cells. The accuracy of their method was 97.8%. They made use of shape based features to detect different cells like basophils , monocytes, eosinophils and lymphocytes. Finally they have diagnosed the disease based on the immature cell count

Xiang li et al. (2018) have investigated a method for the classification of blood cell, to segregate white blood cells and red blood cells. They have used deep convolutional neural networks.

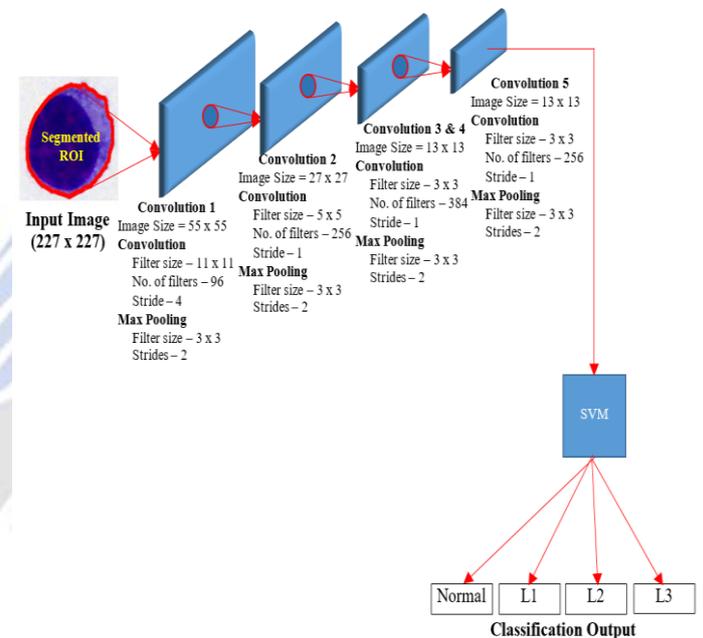
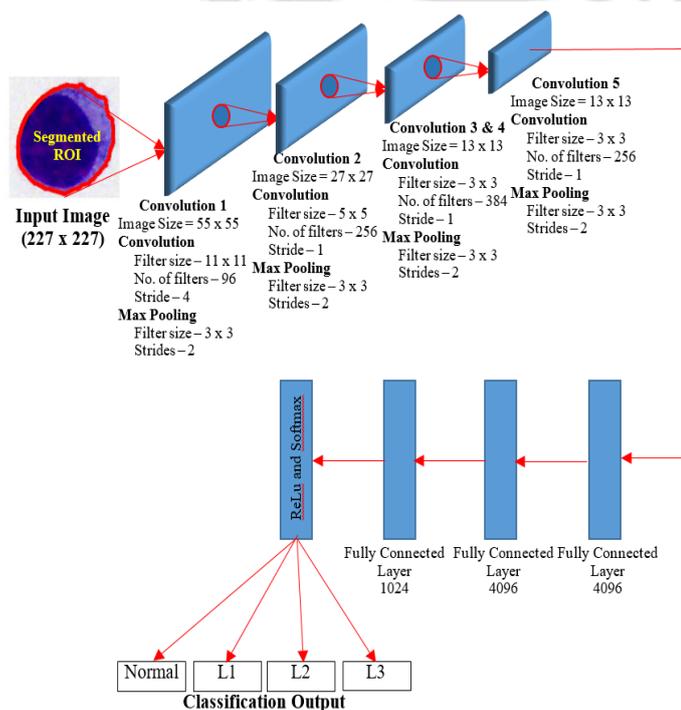
A method was analysed by Khot s et al. (2013) in which they have extracted the features from the images and had applied it to the classifier..

III. PROPOSED METHODOLOGY

The research work has implemented an AlexNet that is pretrained for ALL detection and classification. The concept of transfer learning to retrain the pretrained model with respect to the existing problem was used to design the deep neural network. The RGB images with 227*227 are given as input. The AlexNet Model has got 5 convolutional layers with 3 max polling layers. Each of the convolutional layers have Rectied Linear Unit (ReLU). The last 3 layers such as fully connected layer, softmax layer and classification layer of the pretrained AlexNet were used for Transfer Learning. In the proposed AlexNet architecture the data will be first classified into normal or abnormal. The feature vector size has been reduced by using another fully connected layer with 1024 neurons was added. The layer has been fully connected binary classification (normal, abnormal) probability using softmax function followed by further classifying it into L1, L2 and L3 . In order to further classify it the last fully connected layer is modified from binary classes to four classes output probability by keeping the remaining layers the same.

Transfer learning in machine learning is retraining the pre trained model with respect to the problem. Transfer learning has got advantages such as it is fast as well as effective. The parameters such as the filter size, the number of filters, and stride for each layer are shown in Figure 1.

Through experiments, it was evident that the CNN base model outperforms the conventional SVM-based model in terms of accuracy and its performance is at par with enhanced ensemble SVM from Phase II. However, CNN has more steps hence the time needed to run it is longer than SVM. The CNN models are very good at automatically learning the optimal features. However, its performance is not always the best during classification, as the fully connected layers use parameters that have to be fine-tuned manually. To solve this issue, hybrid models that combine CNN and SVM are proposed by several researchers. These systems use CNN for feature extraction and SVM for classification. This research work, moving in the same path, also combines CNN and SVM and includes procedures that can improve its performance



The proposed hybrid system is designed to combine CNN with ensemble SVM. The base SVM classifiers are created by differing the kernel functions used. The kernel functions used in the proposed system are Linear Kernel, Polynomial Kernel, (RBF), Gaussian Kernel, exponential kernel, Laplacian kernel, Bessel function kernel, Gaussian Radial Basis Function, ANOVA RBF kernel, Laplace RBF and hyperbolic or Sigmoid Kernel. Another point from Phase II is that not all base classifiers in the classification system help with classification. For this purpose, the hybrid CNN-ensemble SVM classifier is enhanced to include the selection step described in Phase II. Thus, Phase III of the research methodology proposes three systems as given below:

- Hybrid CNN-ensemble SVM classifier
- Hybrid CNN-ensemble SVM using DCS
- Hybrid CNN-ensemble SVM using DES

The proposed hybrid classifiers works in two stages.

In the first stage, the AlexNet CNN extracts the deep features and feeds them to the second stage. The second stage then uses the ensemble SVM classifier to classify the deep feature maps extracted from AlexNet CNNs. The hybrid model proposed replaces the last three fully connected layers of base model by the ensemble SVM classifier. In this stage, dynamic selection methods, DCS and DES, are used, so that only optimal base classifiers are used during classification. Thus, the hybrid model is enhanced by combining CNN with ensemble SVM with DCS/DES method. The DES method selects five optimal classifiers during the construction of the ensemble system.

Aggregation of results, while using ensemble with DES, is done using a weighted ensemble of networks method. In this method, the proposed model considers the performance of each SVM using a weight that denotes its contribution to the final classification. The main aim of this method is to enhance classification system performance by using individual SVM's performance to the result of the respective base classifier. The weights are estimated by first calculating the probability of each class and then using it to calculate an evaluation score. The main advantages of the proposed hybrid system are to reduce the chance of overfitting, number of parameters and the time & process complexities.

3.1 Best Base Classifier Selection Algorithm

- Input

Set of Base Classifiers (S), Classification Accuracy (A)

Output

Set of best performing classifiers (BHC) that are used to construct SVM ensemble

- The Algorithm

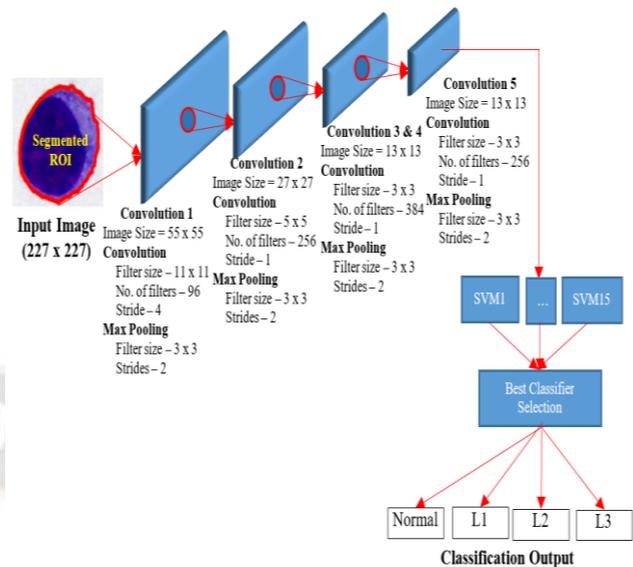
Step 1 : $BHC = \{ \}$

Step 2 : Arrange base classifiers in S in descending order of its associated accuracy

Step 3 : $BHC = BHC + \{C_1\} + \{C_2\} + \{C_3\} + \{C_4\} + \{C_5\}$

– Step 4 : Return BHC

- The returned BHC has the best performing classifier set



3.2 Classification using Deep Learning Algorithm

The nodes will process the input and the results will be communicated to the next layer of nodes. This process will be repeated until it reaches an output layer. One major problem with ML models is with the feature extraction process, which is hand-crafted. DL models are capable of focusing on the right features by themselves with very little or no human intervention.

To perform early detection of ALL, Phase III proposes a DL classification model based on Convolutional Neural Network (CNN or ConvNet) classifier. This model, called as the base model in this research work, is designed using the AlexNet CNN model based on the transfer learning method, in which deep feature maps were extracted and classified.

CNN will be convolved across the image in the input data and uses convolutional layers for feature extraction. The CNN eliminates the need for human intervention by identifying the features by itself and classifying the images. CNN will extract the features directly from the images. The features are not pre-trained and they will be learned while the network trains on a collection of images. The automatic feature extraction makes deep learning models highly accurate for challenging tasks in Image Processing.

One major issue with CNN is when the size of the training dataset is small, then the issue of overfitting arises (Xiao *et al.*, 2021). To solve this issue, the training set size has been increased through the use of augmentation methods. Data Augmentation methods are implemented through the use of image manipulation methods to increase the size of the training set. Seven manipulation methods were used. They are height shift, width shift, zoom, horizontal flip, vertical flip,

rotation and shearing. By applying image manipulation methods, the training set size was increased and normalized.

The input for the proposed base model AlexNet are the RGB colored images with 227x227 pixel resolution. The base model is designed with five convolutional layers along with three max pooling layers followed by Rectified Linear Unit. Transfer learning was performed using the FCL, softmax layer and classification layer of the pre-trained AlexNet are used. The proposed AlexNet architecture was fine-tuned to perform detection and classification of ALL, where the data is first classified into normal or abnormal.

In the next step, the abnormal category is further classified into L1, L2 and L3 subtype categories. The last fully connected layer is modified from binary classes (normal, abnormal) to four classes classes, while keeping the rest of the layers the same.

Through experiments, it was evident that the CNN base model outperform the conventional SVM-based model in terms of accuracy and its performance is in par with enhanced ensemble SVM from Phase II. However, CNN has more steps, so, the time needed to run it is longer than SVM. The CNN models are very good at automatically learning the optimal features. However, its performance is not always the best during classification, as the fully connected layers use parameters that have to be fine-tuned manually. To solve this issue, hybrid models that combine CNN and SVM are proposed by several researchers (Kang *et al.*, 2018; Liu *et al.*, 2018). These systems use CNN for feature extraction and SVM for classification. This research work, moving in the same path, also combines CNN and SVM and includes procedures that can improve its performance.

The kernel functions used in the proposed system are Linear Kernel, Polynomial Kernel, (RBF), Gaussian Kernel, exponential kernel, Laplacian kernel, Bessel function kernel, Gaussian Radial Basis Function, ANOVA RBF kernel, Laplace RBF and hyperbolic or Sigmoid Kernel. The hybrid CNN-ensemble SVM classifier is enhanced to include the selection step described in Phase II. Thus, Phase III of the research methodology proposes three systems as given below:

1. Hybrid CNN-ensemble SVM classifier
2. Hybrid CNN-ensemble SVM using DCS
3. Hybrid CNN-ensemble SVM using DES

All three classifiers listed above works in two stages. In the first stage, the AlexNet CNN extracts the features and feeds those features extracted to the second stage. The second stage then uses the ensemble SVM classifier to classify the

features extracted from AlexNet CNNs. The hybrid model proposed replaces the last three fully connected layers of base model by the ensemble SVM classifier. In this stage, dynamic selection methods, DCS and DES, are used, so that only optimal base classifiers are used during classification. Thus, the hybrid model is enhanced by combining CNN with ensemble SVM with DCS/DES method.

The DES method selects five optimal classifiers during the construction of the ensemble system. During aggregation of results, while using ensemble with DES, a weighted ensemble of networks method is used. In this method, the proposed model considers the performance of each SVM using a weight that denotes its contribution to the final classification. This method has been used to enhance classification system performance by using individual SVM's performance to the result of the respective base classifier. The weights are estimated by first calculating the probability of each class and then using it to calculate an evaluation score. The proposed hybrid system will reduce the chance of overfitting, number of parameters and the time & process complexities.

In order to meet the above-listed objectives, the research methodology was designed in three phases. Each phase was designed separately to satisfy two points. The first was to improve the performance of the task connected to it, while the second was to integrates these tasks together in a manner that could increase the system's performance of ALL classification. The phases were integrated using a simple I/O (Input/Output) interface, where the output from Phase I was used as input to Phases II and III. The working of each phase along with the optimization methods used is described in the following sections. The first phase, pre-processing, performs two tasks, namely, enhancement and white blood cell identification.

3.1. Noise Removal

Noise in images are visual distortions caused due variations in brightness or color information. The noise in microscopic images is handled by an algorithm that combines the advantages of two frequently used transformation-based algorithms, namely, Discrete Wavelet Transformation (DWT) and K-Singular Value Decomposition (K-SVD).

These issues are solved, in this research work, by proposing a unified algorithm that combines contrast adjustment algorithm with noise removal and edge enhancement algorithms. The contrast variations are corrected using an adaptive histogram equalization algorithm. The distortions in the image is removed using a hybrid DWT and K-SVD algorithm. This algorithm beings with DWT coefficients to obtain LL, LH, HL and HH subbands, The LL subband is then

divided into edge and non-edge regions using its contrast information. The edge region is enhanced using the sigmoid function, while the noise in the non-edge regions are reduced using K-SVD algorithm.

3.2. ROI-Extraction

The second task of preprocessing phase is the extraction of White Blood Cells (WBC) from the enhanced microscopic image.

The methodology behind the proposed segmentation method involves two steps. The first step enhances the working of two conventional algorithms, whose results are then combined to form a final set of segments in the second step. The two algorithms considered during the design of the proposed algorithm are the watershed algorithm and K-means clustering-based algorithm.

The watershed algorithm is enhanced through the use of a set of techniques, that when applied sequentially can produce accurate segments in a fast manner. The proposed enhanced watershed algorithm is designed using color intensity, Otsu's threshold algorithm, enhanced watershed segmentation algorithm, region merging algorithm and pruning algorithm. The K-means clustering-based segmentation algorithm is enhanced through the use of an automatic technique to determine the initial seeds using a subtractive clustering algorithm. The algorithm sets K as 3 since there are three types of blood cells in microscopic images. The algorithm is further enhanced through the use of a computation reduction algorithm, which can speed up the process of clustering and thus, segmentation. The results of the two enhanced segmentation algorithms are then combined by first generating a mean segmentation image, using which a distance map is constructed. Using this distance map, a weight for each algorithm is estimated. Finally, a majority voting algorithm is used to determine the best segment.

3.3 Steps in CSA-WBC

-The CSA-WBC is designed using two segmentation algorithms, which are combined synergistically to produce accurate grouping of blood cells. The first algorithm enhances watershed algorithm, while the second is a clustering based algorithm.

Steps Involved in CSA-WBC are,

Input : Microscopic Image, I

Step 1 : Segment input image using enhanced watershed algorithm and perform region merging

Step 2 : Segment input image using enhanced clustering algorithm and perform region merging

Step 3 : Combine segment results to produce a single set of segments

Step 4 : Identify Lymphocytes

Step 5 : Use a post processing procedure to further refine the segmented result

Output : Three Segments

Input : Microscopic Image M

3.4 Steps in PFKM Algorithm are,

Step 1: Assign $K=3$

Step 2: Estimate K initial seeds (c_j) using Subtractive Clustering Algorithm

Step 3: Repeat

a. For each pixel of an image, Calculate the Euclidean Distance d , between the centre and each pixel of an image using equation given below

$$D = |p(x,y) - c_k|$$

b. Find the closest centre c_j and assign pixels to cluster j

c. Store label of cluster centre j along with the distance and store them in an array Cluster [] and Distance [] respectively

d. Set $\text{cluster}[i] = j$ (j is the nearest cluster)

e. Set $\text{Dist}[i] = D_{ij}$ (Distance between x_i to the closest centre c_i)

f. Recalculate Cluster Centres

g. Compute New Distance to new cluster centres

h. Calculate D with all the cluster centre assign cluster $i =$ cluster j ,

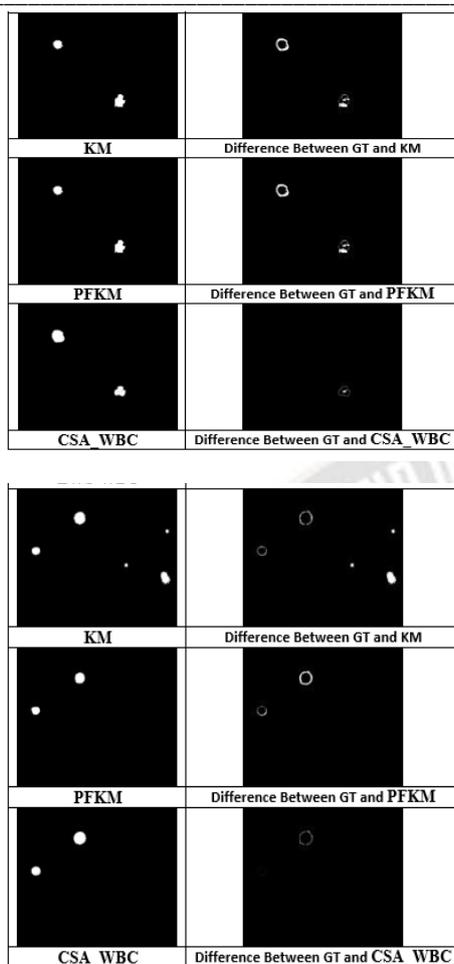
$$\text{Distance} = D_{\text{new}}$$

End if

Until Convergence

Step 4: Output clustered results

The segmented images are shown below,



3.5. Classification

Ensemble systems is used to obtain a more accurate and robust classification by considering multiple views of the same problem. Several researchers have used this idea and proved that multiple classifiers work better than single classifier [Sesmero *et al.*, 2021].

In this research work, the above-described ensemble system is enhanced in two manners, as listed below.

- (i) Usage of optimization procedure that pre-treat the training feature set
- (ii) Usage of base classifier selection methods.

Dynamic Classifier Selection - It will Select one single best classifier from the set of base classifiers generated

- Dynamic Ensemble Selection –Among the set of base classifiers generated a subset of best classifier will be selected.

This research work, to further enhance the process of classification, proposes two enhanced methods that combine static selection and dynamic selection in order to maximise the performance of the proposed EC system. The proposed hybrid systems are

- (i) EC system using static and dynamic classifier selection
- (ii) EC system using static and dynamic ensemble selection

The static selection is done using a pruning algorithm that selects optimal classifiers among the base classifiers constructed before the training step. The resultant set of classifiers are then supplied to a dynamic ensemble or dynamic classifier selection method, whose results are reported as the final classification output. The methodology used by the proposed ensemble systems is presented in Figure 1.

In this research work, a static pruning technique is used, as a preprocessing function to reduce optimal candidate classifiers. Static techniques work to construct a subset of base classifiers of fixed size to improve its performance with respect to the full ensemble, removing the rest of the classifiers that do not meet this objective [Margineantu and Dietterich, 1997; Zhang *et al.*, 2006; Munoz *et al.*, 2009]. The reason for using a static pruning technique with the proposed enhanced EC system is to produce a smaller-sized base classifier set, which can produce the same advantages of the full ensemble system with added advantages like low time complexity.

3.3 Classification using Machine Learning Algorithms

Phase II of the research methodology uses the segmented results to classify the identified white blood cells. The steps involved are, feature engineering and classification. Feature engineering consists of two tasks, namely, feature extraction followed by selection of relevant features. In the classification step, the feature vector obtained from feature engineering is used to classify a cell as normal or cancerous. If cancerous, then to classify them into their types L1, L2 and L3.

3.4 Feature Engineering

In this research work, multiple features are extracted from the segmented image. They are, texture features (Energy, Entropy, Contrast, Correlation, Homogeneity), shape features (Area, Perimeter, Eccentricity, Elongation, Compactness, Minor Axis, Major Axis, Solidity, Form Factor, Nucleus-Cytoplasm Ratio), color features (Mean, Standard Deviation) and irregularity of the nucleus boundary (Horizontal Direction, Vertical Direction). Thus, a total of 19 features are extracted.

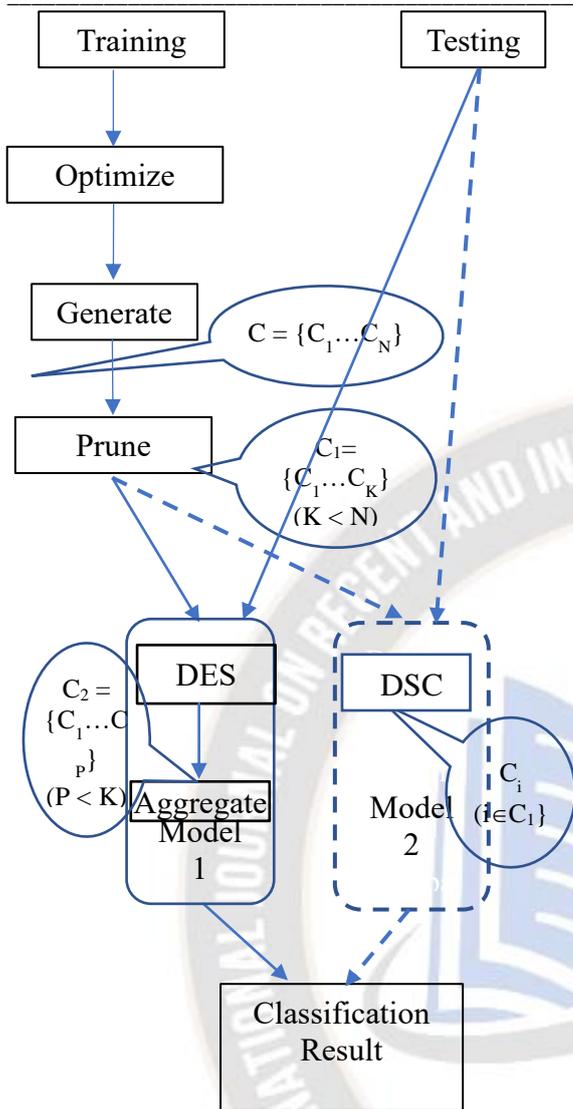


Figure 1 : Proposed Enhanced Ensemble Classification System

The first step selects classifiers that reduce the classification error rate. The rate of error is the percentage of number of classifications that are incorrect by the total number of classifications. The classifiers are then arranged in ascending order of their error rate and this step returns the top 30 classifiers. Let this set be denoted as $\{C_s\}$. In the second step, the kappa statistic pruning technique is used to select the final set of optimal classifiers, $\{C_{ss}\}$.

The kappa statistics returns a value between zero and one, with zero indicating poor agreement and 1 indicating perfect agreement (Landis and Koch, 1977). The resulting set of classifiers is denoted as C_p .

The main aim of the DES is to find a subset of base classifiers, C_o , that can classify a test sample, such that $C_o \square C_p$ and size (C_o) < size (C_p). The classification of test data can be done in three steps.

- Step 1 : Region of Competence Identification. A region surrounding the test data , is used to estimate the base classifier.
- Step 2 : In order to determine the level of the base classifiers Selection Criteria is used in this research work. In this research work, the criteria used is the classification accuracy.
- Step 3 : Determine the selection mechanism, that is, DES or DCS.

That is, all classifiers having the highest accuracy are selected as the most efficient ones, suitable to maximise the performance of the Ensemble system. The number of classifiers selected by DCS is one, while for DES, it is set to 15. The best-performing classifier is selected using an automated procedure that uses the accuracy as the prime metric.

IV. RESULTS AND DISCUSSIONS

Stage 3 experiments focus on the algorithms proposed in Phase III of the research methodology. This stage of experiments also used sensitivity, specificity, accuracy and speed to evaluate the classifiers. The two proposed algorithm CNN-EDCS-SVM and CNN-EDES-SVM were compared with the conventional SVM, conventional CNN and conventional hybrid CNN-SVM classifiers. Figure 4 depicts the accuracy obtained by these classifiers.

The major benefits of the hybrid system proposed is to reduce the chance of overfitting, number of parameters and the time & process complexities.

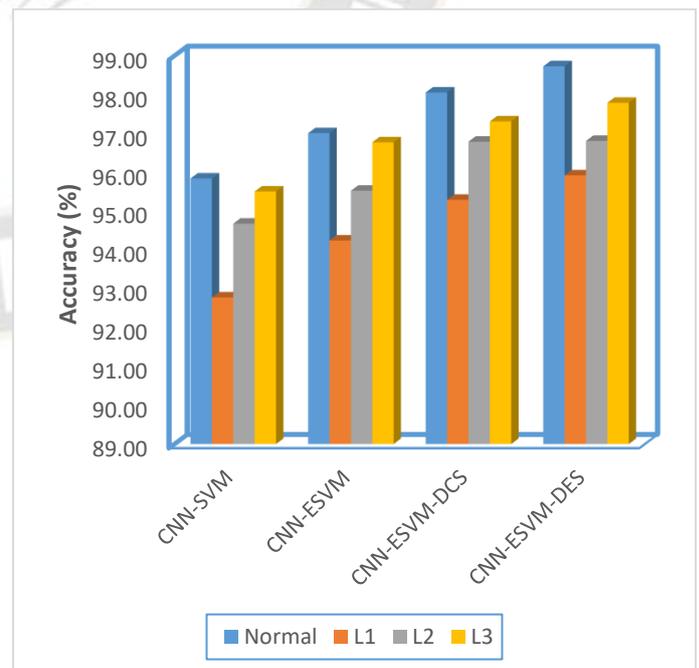
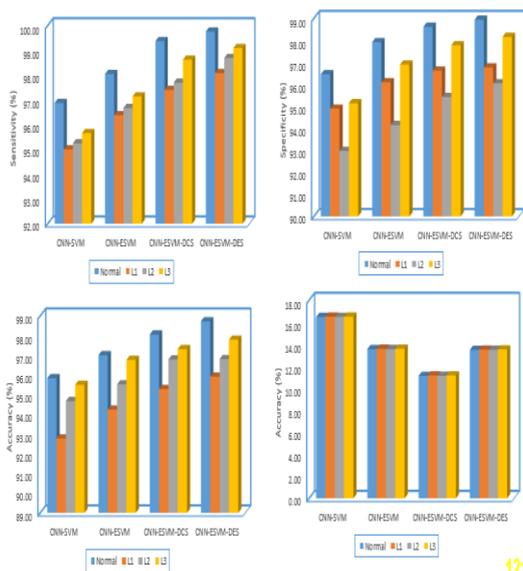
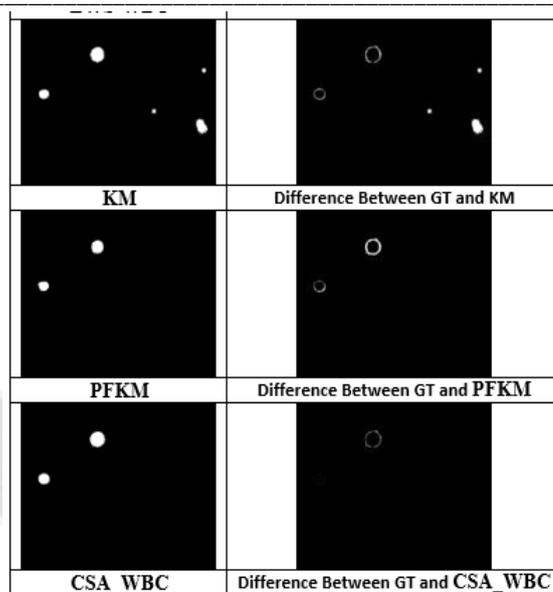
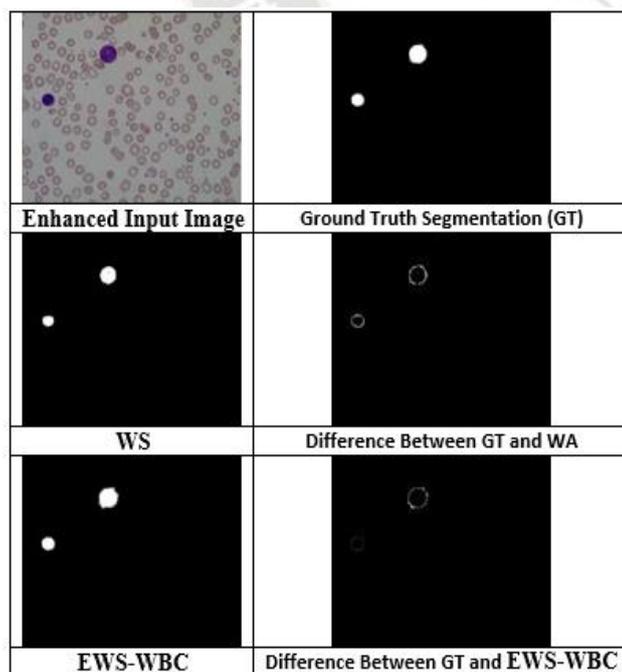


Figure 4 : Accuracy (%) of Hybrid Classifiers

From Phase III experimental results, the hybridization of classifiers have a positive impact on the performance of ALL classification. Comparison of deep and machine learning classifiers showed that CNN outperformed SVM classifier. The proposed hybrid model that used CNN as feature extractor, ensemble SVM with DES for best classifier selection produced maximum advantage during ALL classification. This model produced, on average, a high accuracy of 97.31% accuracy. Thus, from the various results, it could be concluded that the hybrid model that combined CNN with ensemble SVM with dynamic ensemble selection is best suited for ALL classification.



121



The performance of the proposed algorithms in each phase and its cumulative effect on ALL classification was performed using Acute Lymphoblastic Leukemia Image Database obtained from <https://homes.di.unimi.it/scotti/all>

The image enhancement algorithms were analysed using four performance metrics, such as PSNR, MSSl, FoM and speed of enhancing a single image in seconds. The proposed unified algorithm was compared with three conventional methods, discrete wavelet transformation, K-singular value decomposition and one existing DWT-KSVD algorithm. From the results, it was proved that the proposed algorithm unified enhancement algorithm outperforms in terms of all the selected performance metrics. Figure 2 shows the PSNR values obtained by five randomly selected test images, where UCED indicates the proposed unified algorithm. This trend envisaged was the same with all the images in the database.

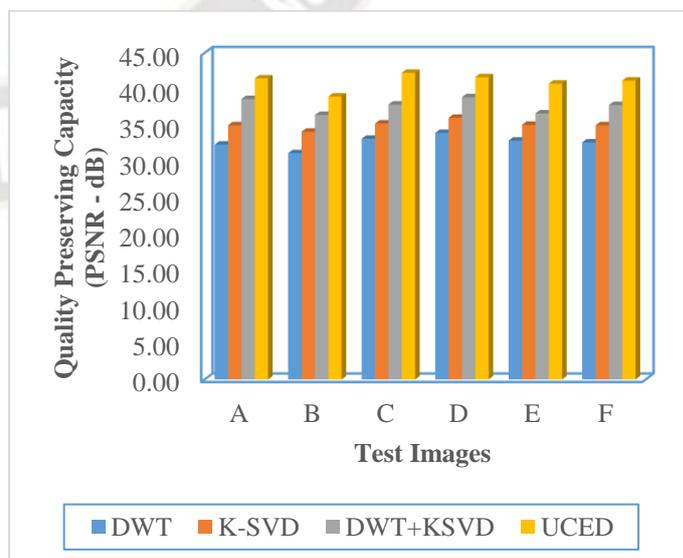


Figure 2 : PSNR (dB) of Enhancement Algorithms

From the results obtained while evaluating the segmentation algorithms, it could be seen both the enhanced watershed and enhanced clustering-based segmentation algorithms work better when compared with their respective conventional counterparts. However, the proposed algorithm that combines the results of the two enhanced algorithms finds white blood cells in the most efficient manner.

Stage 2 experiments were used to evaluate the machine learning classification algorithm. Four performance metrics, namely, sensitivity, specificity, accuracy and classification speed were used during evaluation. From Phase II experimental results, it could be understood that the proposed ensemble classifiers, enhanced ensemble SVM using dynamic classifier selection (EDCS-SVM) and enhanced ensemble SVM using dynamic ensemble selection (EDES-SVM), were more powerful, when compared to single SVM (SVM) and conventional ensemble classification systems (ESVM). Moreover, comparison between the proposed classifiers showed that enhanced ensemble SVM using dynamic ensemble selection is slightly more efficient than the enhanced ensemble SVM using dynamic classifier selection. The accuracy of the classifiers is shown in Figure 3.

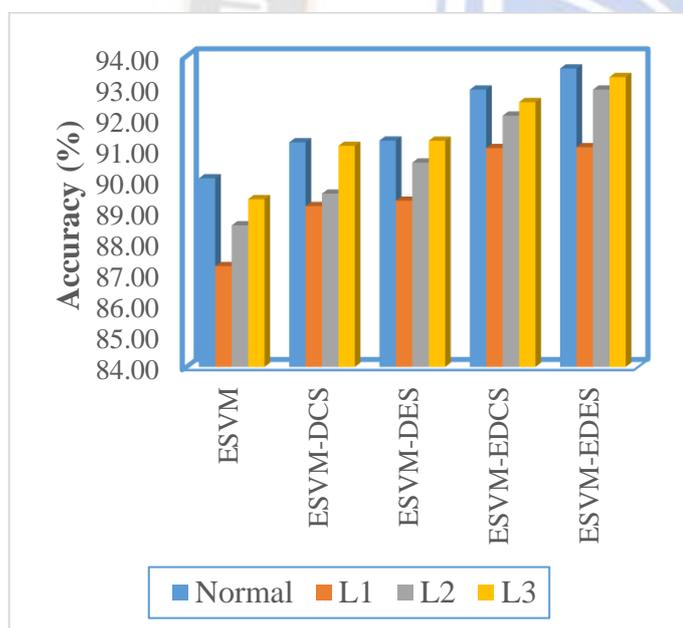


Figure 3 : Accuracy (%) of Machine Learning Classifiers

V. CONCLUSION

From Phase III experimental results, both ensembling of deep learning classifier and hybridization have positive impact on the performance of ALL detection and classification. Comparison of deep and machine learning classifiers showed that CNN outperformed SVM classifier. Among the ensemble models proposed, the CNN-based ensemble model improved

ALL classification when compared with SVM-based enhanced ensemble model (EDES-SVM). Stage 3 experiments showed that proposed hybrid model that used CNN as feature extractor, ensemble SVM with best classifier selection as classifier produced maximum advantage during ALL classification. This model produced a high accuracy of 97.31% accuracy. Thus, from the various results, it could be concluded that CNN-EESVM model is best suited for ALL classification.

Machine Learning algorithms have been used in healthcare industries as powerful analytical and diagnostic tool that can assist physicians with maximum efficiency.

The research work was divided into three phases. The first phase focused on preprocessing, which works on improving the visual quality of the input microscopic images and identification of the white blood cells. The enhancement of microscopic images was done using a unified approach that combined contrast adjustment, edge enhancement and denoising into a single algorithm. The second task of Phase I is the extraction of white blood cell regions from microscopic images using segmentation algorithm. The proposed segmentation algorithm merged segments produced by watershed and enhanced K-Means-based clustering algorithm.

Phase II of the research methodology proposed an enhanced SVM based ensemble classifier. The enhancement was achieved by using procedures that reduce the time complexity of ensembling and also to improve its accuracy. The proposed classifier performs classification using two tasks, namely, feature engineering and classification. Feature engineering first extracted four groups of features (texture, shape, color and Irregularity of the nucleus boundary) to form a 19 attribute feature vector. From this feature vector, optimal features were selected using an algorithm namely MRMR algorithm. The constructed ensemble model was enhanced through the use of an algorithm that improved the quality of the training data along with dynamic ensemble selection and dynamic classifier selection algorithms. Apart from this, a pruning algorithm that removed irrelevant base classifiers was also proposed.

REFERENCES

- [1] Sesmero, M.P., Iglesias, J.A., Magán, E., Ledezma, A. and Sanchis, A. (2021) Impact of the learners diversity and combination method on the generation of heterogeneous classifier ensembles, Applied Soft Computing, Vol. 111, Article ID 107689, Pp. 1-17.
- [2] Sharma, N., Mukopadhyay, A., Garg, A., Shrivastava, A., Pathan, S. (2022) Detection of Acute Lymphoblastic Leukemia and its Subtypes using Deep Learning, International

- Journal of Engineering Research & Technology, Vol. 11, Issue 05, Pp. 339-341.
- [3] Starza, D.I., Chiaretti, S., De Propriis, M.S., Elia, L., Cavalli, M., De Novi, L.A., Soscia, R., Messina, M., Vitale, A., Guarini, A. and Foà, R. (2019) Minimal Residual Disease in Acute Lymphoblastic Leukemia: Technical and Clinical Advances, *Frontiers in Oncology*, Vol. 9, Article ID 726, Pp. 1-17.
- [4] Stork, J., Ramos, R., Koch, P. and Konen, W. (2015) SVM Ensembles Are Better When Different Kernel Types Are Combined, Lausen, B., Krolak-Schwerdt, S., Böhmer, M. (eds) *Data Science, Learning by Latent Structures, and Knowledge Discovery, Part of Studies in Classification, Data Analysis, and Knowledge Organization Book Series*, Springer, Pp. 191–201.
- [5] Terwilliger, T. and Abdul-Hay, M. (2017) Acute lymphoblastic leukemia : A comprehensive review and 2017 update, *Blood Cancer Journal*, Vol. 7, Article ID e577, Pp. 1-12.
- [6] Vijay, M. and Bhupendra, D. (2014) Segmentation of Microscopic Images: A Survey, *Proceedings of International Conference on Electronic Systems, Signal Processing, and Computing Technologies*, Pp. 362-364.
- [7] Vohra, S.K. and Prodanov, D. (2021) The Active Segmentation Platform for Microscopic Image Classification and Segmentation, *Brain Sci.*, Vol. 11, Article ID 1645, Pp. 1-23.
- [8] Xiao, M., Wu, Y., Zuo, G., Fan, S., Yu, H., Shaikh, Z.A. and Wen, Z. (2021) Addressing Overfitting Problem in Deep Learning-Based Solutions for Next Generation Data-Driven Networks, *Wireless Communications and Mobile Computing*, Vol. 2021, Article ID 8493795, Pp. 1-10.
- [9] Yang, L. (2011) Classifiers selection for ensemble learning based on accuracy and diversity, *Procedia Engineering*, Vol. 15, Pp. 4266-4270.
- [10] Zhang, Y., Burer, S. and Street, W.N. (2006) Ensemble pruning via semi-definite Programming, *Journal of Machine Learning Research*, Vol. 7, Pp. 1315–1338.
- [11] Arivuselvam, B. and Sudha, S. (2022) Leukemia classification using the deep learning method of CNN, *J Xray Sci Technol.*, Vol. 30, No. 3, Pp. 567-585.
- [12] Britto, A.S.J., Sabourin, R. and Oliveira, L.E.S. (2014) Dynamic selection of classifiers — A comprehensive review, *Pattern Recognition*, Vol. 47, No. 11, Pp. 3665–3680.
- [13] Bukhari, M., Yasmin, S., Sammad, S. and Abd El-Latif, A.A. (2022) A Deep Learning Framework for Leukemia Cancer Detection in Microscopic Blood Samples Using Squeeze and Excitation Learning, *Mathematical Problems in Engineering*, vol. 2022, Article ID 2801227, Pp. 1-18.
- [14] Chen, R., Pu, D., Tong, Y. and Wu, M. (2021) Image-denoising algorithm based on improved K-singular value decomposition and atom optimization, *CAAI Transactions on Intelligence Technology*, Vol. 7, Issue 1, Pp. 117-127.
- [15] Chennamadhavuni, A., Lyengar, V. and Shimanovsky, A. (2022) Leukemia, *StatPearls [Internet]*, StatPearls Publishing.
- [16] Fang, H., Tang, P. and Hao, S. (2020) Feature Selections Using Minimal Redundancy Maximal Relevance Algorithm for Human Activity Recognition in Smart Home Environments, *Hindawi Sensor-Based Systems for Independent Living of Ageing People*, Vol. 2020, Article ID 8876782, Pp. 1-13.
- [17] Kang, J., Park, Y.J., Lee, J., Wang, S.H. and Eom, D.S. (2018) Novel Leakage Detection by Ensemble CNN-SVM and Graph-Based Localization in Water Distribution Systems, *IEEE Transactions on Industrial Electronics*, 65(5), 4279–4289.
- [18] Krishnaveni, N. and Radha, V. (2021) Ensemble of Multiple Kernel SVM Classifiers for Detection of Online Spam Reviews, *International Conference on IoT based Control Networks and Intelligent Systems*, Pp. 479-488.