CHAPTER 8

SUMMARY AND CONCLUSION

Artificial intelligence algorithms have been used in healthcare industries as powerful analytical and diagnostic tool that can assist physicians with maximum efficiency. ML and DL models can be used to address the various challenges faced during manual diagnosis, differing opinions of experts and time consuming tracking of blood samples. Both models play a key role in the early detection of ALL. This research work proposed classification models that were enhanced to improve its performance in terms of accuracy during the detection and classification of leukemia.

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 enhancement, edge enhancement and denoising into a single algorithm. Contrast enhancement was performed using CLAHE. The noise in microscopic images was removed using hybrid DWT and K-SVD algorithm. The edges were improved using sigmoid function.

During image enhancement, the unified approach proposed offers advantages as listed below.

i. The algorithm is efficient in maintaining structure and edge details to the maximum extent,

ii. Reduces time complexity considerably, when compared to the individual application of each algorithms, as the unified approach focus on specific regions, and

iii. Do not introduce artifacts or blur as noise-free region is not processed at all.

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 enhanced watershed and enhanced K-Means-based clustering algorithm. The watershed algorithm was enhanced to overcome its issues on over-segmentation and sensitivity to noise. The K-Meansbased clustering algorithm was enhanced by making it parameterless and reducing the number of computations.

Phase II of the research methodology focused on developing an enhanced machine learning classifier to perform ALL-C. This phase 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, starts with the extraction of 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 minimum redundant maximum relevant (MRMR) feature selection algorithm. This optimal feature vector was then used to construct a homogeneous SVM-based ensemble classifier.

The constructed ensemble model was enhanced through the use of an algorithm that improved the quality of the training data along with two selection algorithms, namely, static and dynamic selection algorithms. Initially, a static selection using a pruning algorithm is used to select only the relevant classifiers. The dynamic selection can be performed using either dynamic classifier selection or dynamic ensemble selection algorithms. Thus, in Phase II two classifiers that combine static selection with dynamic classifier selection and static selection with dynamic ensemble selection were proposed.

The final phase of the research methodology, proposed a enhanced deep learning model for ALL classification. Initially, a CNN AlexNet base model was designed. This was modified so that the CNN works as a feature extractor, whose results are then fed to SVM to perform classification. This conventional manner of designing hybrid CNN-SVM classifier was then enhanced by using a SVM-based ensemble classifier in the place of single SVM classifier. The heterogeneous classifier used different variant of classifiers that used the different kernels of SVM to construct the ensemble. The kernels used in this research work are linear kernel, polynomial kernel, gaussian RBF kernel, exponential kernel, laplacian kernel, bessel function kernel, anova RBF kernel and hyperbolic or sigmoid kernel. The SVM ensemble was further enhanced through the use of dynamic classifier selection and dynamic ensemble selection methods. Thus, the two hybrid models proposed in Phase II are CNN-Ensemble SVM with dynamic classifier selection algorithm and CNN-Ensemble SVM with dynamic ensemble selection algorithm.

The proposed algorithms were tested vigorously using Acute Lymphoblastic Leukemia Image Database along with images from Google image search. The size of the dataset was increased by using image manipulation methods. The inclusion of noise removal algorithm improved the quality of microscopic image in terms of contrast, edge and detailed coefficients, and achieved a high PSNR of 41.05 dB. This algorithm improved accuracy of the ALL-C system using SVM classifier by 1.49%.

The WBC identification algorithm using the proposed segmentation algorithm resulted with stable (78.07 - average stability criterion) and being highly resistant to the noise (79.7 - average anti-noise criterion) in the microscopic images. The segmentation algorithm resulted with an accuracy gain of 1.87% when incorporated in the ALL-C system.

Among the machine learning classifiers, the ALL-C system that used preprocessing algorithm along with optimal features and SVM-based ensemble classifier enhanced using dynamic ensemble selection algorithm produced maximum efficiency during ALL classification. This classifier produced, on average, 93.99% (sensitivity), 93% (specificity) and 92.73% (accuracy). Maximum speed efficiency was obtained when the ensemble SVM was combined with dynamic classifier selection algorithm.

Among the deep learning classifiers, the hybrid system that combined CNN with ensemble SVM classifiers with dynamic ensemble classifier selection algorithm showed maximum performance during leukemia detection. This system produced a maximum of 98.91% sensitivity, 97.48% specificity and 97.31% accuracy, on average, obtained while using preprocessing methods along with the hybrid model that combined CNN with enhanced SVM-based ensemble model. The fastest ALL-C system was the one that used CNN with ensemble SVM classifier with dynamic classifier selection algorithm.

Experimental results analyzing algorithms proposed in each phase revealed that the enhancement operations are successful in both improving their respective tasks and improve the performance of ALL classification, thus meeting the research objectives. Thus, it can be concluded that the proposed system meets the research objectives framed and can be used to improve the detection of leukemia disease and can assist pathologist and oncologist to diagnose the leukemia in better way efficiently.