ABSTRACT

Disease misdiagnosis is a major challenged faced during blood cancer detection in early stages. The use of image processing and machine learning algorithms have been of tremendous use to detect leukemia in its early stage. This research work uses microscopic images to design enhanced classification systems that can aid in minimizing the rate of misdiagnosis and act as a tool to assist the physicians during Acute Lymphocytic Leukemia (ALL) detection and classification. This goal was achieved by enhancing the various steps, namely, image enhancement, white blood cell detection and classification, involved during Acute Lymphocytic Leukemia detection.

The algorithms proposed were grouped into three research phases. The first phase focused on preprocessing, which performed image enhancement and segmentation to detect white blood cells. A unified approach that combined three enhancement tasks, namely, contrast enhancement, noise removal and edge enhancement was proposed. Contrast enhancement was performed using Contrast Limited Adaptive Histogram Equalization (CLAHE). The noise in microscopic images was removed using hybrid Discrete Wavelet Transformation (DWT) and K-Singular Value Decomposition (K-SVD) algorithm. The edges were improved using sigmoid function. The identification of white blood cells was done using a segmentation algorithm. The proposed segmentation algorithm combined an enhanced watershed algorithm with an enhanced K-Means clustering-based segmentation algorithm. The results of these two algorithms were merged and used during classification.

The second and third phases of the research methodology focused on improving the classification of ALL classification system. The second phase proposed an enhanced SVM-based ensemble classification system. This system used multiple features (texture, shape, color and irregularity of the nucleus boundary) to form a feature vector having 19 feature sets. Optimal features from this set is obtained using a feature fusion and maximum relevancy minimum redundancy feature selection algorithm. This optimal feature set is then used to train the ensemble classifier. The proposed classifier was enhanced in three steps. The first step used a training set optimization algorithm to refine the training feature set. The second step used a static selection algorithm that used pruning technique to remove irrelevant classifiers from the ensemble system. The third

step used either a dynamic classifier selection or dynamic ensemble selection algorithms, to identify a single best performing classifier or a set of best performing classifiers.

The third phase proposed an enhanced hybrid CNN-SVM classifer, where the CNN model was used as a feature extractor, while SVM was used to classify the microscopic images. The SVM component of the hybrid deep learning classification system was enhanced to form an ensemble system that used the dynamic classifier selection or dynamic ensemble selection algorithms, to improve the ALL classification. The base classifiers were created using eight different Kernel functions of SVM, namely, linear kernel, polynomial kernel, gaussian RBF kernel, exponential kernel, laplacian kernel, bessel function kernel, anova RBF kernel and hyperbolic or sigmoid kernel.

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. 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. The proposed SVM-based classifier produced an average accuracy of 92.73%, while using preprocessing algorithms with SVM ensemble classifier that used static pruning and dynamic ensemble classifier. The enhanced deep learning classifier produced an average accuracy of 97.31%, while using CNN-SVM hybrid model enhanced through the incorporation of ensemble method along with preprocessing and dynamic ensemble selection algorithm.