ABSTRACT

Among the women around the world, breast cancer is the most common cancer and the leading cause of cancer-related deaths. In India, breast cancer is the second most common cancer among females. Analyzing and classifying breast tumor cell with minimum number of feature set is the objective of this work. The mammogram screening is found an effective technique to detect the breast cancer in the early stage. Efforts have been made to develop a breast cancer classification system to help radiologist in the examination of breast tumor by using mammogram, which enhances the accuracy of diagnosis, as well as to improve the uniformity of interpretation of images by the use of computer results as a reference.

The following steps have been used to develop breast cancer classification system. They are preprocessing, segmentation, feature extraction, feature selection and classification. Preprocessing of mammogram includes noise elimination, artifact removal and background elimination, contrast enhancement and pectoral muscle elimination. Low pass filter has applied to eliminate the noise, thresholding technique has been used to eliminate artifacts and background information. Region growing algorithm has been used to remove the pectoral muscle region of the breast and histogram equalization has utilized for contrast adjustment.

The Region of Interest contains masses which have been extracted from the preprocessed images. Region growing algorithm has been used to locate the mass portion from the ROI. Once mass boundary is identified, then feature extraction is done to generate feature vector which is used in classification stage. From mass portion various features namely texture, shape, intensity histogram features and radial distance features have been
extracted. Totally 123 features were extracted from mass portion and had been used as input for feature selection algorithm.

In this thesis, heuristic optimization techniques for feature selection have been proposed for breast cancer classification system. Proposed feature selection algorithms are Hybrid Harmony Search (HHS), Enhanced Cuckoo Search (ECS) algorithm and Improved Lion Optimization Algorithm (ILOA). For classifying the breast tumor cells, Minimum Distance Classifier, k-Nearest Neighbor (k-NN) and Support Vector Machine (SVM) classifier are used as a classifier. The overall accuracy of a classifier on validation samples are used as a fitness value for HHS, ECS and ILOA.

The hybrid scheme for feature selection algorithm is obtained by combining cuckoo search and harmony search. The Minimum Distance Classifier, k-NN classifier and SVM classifier produces 98.19%, 98.34% and 97.18% average classification accuracy respectively with minimum number of features. The performance of the new hybrid algorithm is compared with the Genetic Algorithm, Particle Swarm Optimization algorithm, Cuckoo Search and Harmony Search.

Enhanced Cuckoo Search (ECS) is a modified version of Cuckoo search with different types of host nest and multiple eggs. ECS with Minimum Distance Classifier, k-NN classifier, and SVM classifier produces 98.75%, 99.13% and 99.22% average classifier accuracy respectively with minimum set of features. The performance of the new ECS is compared with the Particle Swarm Optimization, Cuckoo Search and Harmony Search.

To improve the performance of Lion optimization algorithm, ILOA had been introduced. The proposed ILOA with Minimum Distance Classifier, k-NN classifier and SVM produces 98.92%, 99.31% and 98.62% average
classifier accuracy respectively with minimum number of feature sets. The performance of the new ILOA is compared with Genetic Algorithm, Cuckoo Search, Harmony Search and Lion Optimization Algorithm.

The performance of the developed breast cancer classification system has been estimated and found that the classification accuracy is high with optimal features, which confirm to be a reliable system for medical pathology.