ABSTRACT

A mammogram is a radiograph of the breast tissue. It is an effective non-invasive means of examining the breast, commonly searching for masses and/or microcalcifications. Cancer is not preventable, but early detection leads to a much higher chance of recovery and lowers the mortality rate from this disease. Mammography plays a central part in early detection of breast cancers because it can show changes in the breast up to two years before a patient or physician can feel them.

Initial mammographic images themselves are not usually enough to determine the existence of a benign or malignant disease with certainty. Most of the image processing techniques may be applied to process the mammograms in order to extract the hidden information. Mammography systems allow manipulation of fine differences in image contrast by means of image processing algorithms. Challenges in mammography become the real motivation for this research.

The preprocessing of mammogram image is essential before detection and segmentation of microcalcification. However, the presence of artifacts and pectoral muscle can disturb the detection of microcalcification and reduce the rate of accuracy in the Computer Aided Diagnosis (CAD). Its inclusion can affect the results of intensity-based image processing methods and needs to be identified and removed before further analysis.

Image segmentation is one of the most critical tasks in automatic image analysis. Segmentation consists of subdividing an image into its constituent part and extracting those of interest. Many techniques for global thresholding have been developed over the years to segment images and
recognize patterns but the error on the segmentation leads to misclassification. In this thesis mammogram is segmented using Rough Set Theory.

Since the classification algorithm requires the classified data to be composed of feature vectors, data mining cannot be directly performed on the original image. The Gray Level Co-occurrence Matrix (GLCM) is a well-established robust statistical tool for extracting second order texture information from images. The GLCM characterizes the spatial distribution of gray levels in an image.

Feature selection refers to the process of selecting those input attributes that are most predictive of a given outcome. Unlike other dimensionality reduction methods, feature selectors preserve the original meaning of the features after reduction. The benefits of feature selection are twofold: it considerably decreases the running time of the induction algorithm, and increases the accuracy of the resulting model.

Fuzzy-Rough feature selection builds on the notion of the fuzzy lower approximation to enable reduction of datasets containing real-valued features. The process becomes identical to the crisp approach when dealing with nominal well-defined features. Fuzzy-Rough- Quick-Reduct algorithm has been applied for feature selection.

Classification is assigning the objects in the dataset into a predefined set of classes. It is a type of supervised learning, because the set of classes are introduced to the system before executing classification algorithm. Classification of objects in a dataset is very useful both to understand the characteristics of existing objects and to predict the behaviours of new objects.
In the literature mainly Neural Network is used for classifying the segmented portion. It requires a number of parameters that are typically best determined empirically, such as the network topology or structure. In this thesis, the metaheuristic Ant-Miner algorithm and its variants have been proposed to improve the classification accuracy by considering all the types of mammograms in the MIAS data base since Neural Networks have been criticized for their poor interpretability. The main contributions of the thesis are:

- Fuzzy-Rough Quick-Reduct has been proposed and implemented in the domain of mammogram for feature selection in order to improve the classification accuracy.
- The metaheuristic Ant-Miner algorithm has been applied for mammogram image classification on the extracted feature set over the segmented image and it is compared for classification accuracy.
- The Ant-Miner algorithm has been modified for optimizing the parameters such as the number of Ants, minimum number of cases per rule and the number of rule convergence.
- Another improvement is made in the Ant-Miner algorithm by implementing Tsallis Entropy instead of Shannon Entropy.

It is observed that the best classification accuracy is achieved at the angle $90^0$ using Ant-Miner after applying feature selection.