Chapter 6: Conclusion and Future Work

The conclusions of work of this thesis and suggestions for future research are presented in this chapter.

6.1 Conclusions

CBIR is browsing, searching and navigation of images from large image databases based on their visual contents. CBIR has been an active area of research for more than a decade and many CBIR systems have been developed in the different application areas. The focus of this thesis is to develop some efficient algorithms of CBIR systems for general images as well as mammogram application. For the general image, this thesis contributed an improved image retrieval system using varying weighted similarity measures, fusion of fast features and machine learning approaches. This contribution reduced the semantic gap and search space, and retrieved the most relevant images in less searching time.

For the mammogram application, this thesis contributes towards the improvement of breast cancer diagnosis through evidence- based medicine by supporting the physicians or radiologist in their clinical decision making process. The thesis presented mammogram classification and retrieval system using informative WCS-LBP texture features and random forests. Also resolved the issues of artifacts suppression, and segmentation by using connected component labelling, morphological image processing, adaptive k-means and modified seeded region growing algorithm. Further, reduced the semantic gap and search space, by using SOM with WCS-LBP features.

The contribution towards pre-processing, feature extraction, and supervised and unsupervised stages of the content-based approach is a highlight of this research work. The chapter-wise conclusion of this thesis is also being summarized as follows:

Chapter 1 discussed the motivation, background and problem description for the presented work including thesis scope/objectives, and contributions.

Chapter 2 discussed the theoretical background for CBIR system. This chapter introduced an overview of different variants of feature extraction techniques, and similarities measures. Also, a literature survey of prominent CBIR approaches for general images, mammogram classification and retrieval are given.

Chapter 3 demonstrated a simple and effective approach for image retrieval based on the fusion of fast feature sets, weighted similarity measure and machine learning approaches. This chapter contribution was divided into two sections. In the first section, varying weighted similarity measure using conventional approach gives encouraging retrieval performance. Further, this contribution was extended on pre-clustered images and trained the random forests using all ground truth of database. In the second section, fast and effective image retrieval was proposed which uses supervised learning with the fusion of OCLBP and statistical moments. In both contributions, used feature sets are compact with rich information, fast in indexing and constant in size. In addition, as opposed relationship between low-level features and higher level semantics viz. object and their inter-relationships, this chapter resolved the semantic level problem by applying the supervised machine learning techniques with different variants of colour and texture features. Both contributions narrowed down the search space and got promising retrieval results as compared to the conventional and other state-of-art methods. Therefore, using supervised learning based techniques; we

not only improved the efficiency of the CBIR systems but also improved the accuracy of the overall process.

Chapter 4 presented a framework for the mammogram classification and retrieval using WCS-LBP feature and random forests. For pre-processing of mammograms, labels, and background having no details were cropped and ROI's were extracted, after that adaptive median filter was used to suppress the noises. Further, we introduced random forests with selected local wavelet based CS-LBP features which improved the classification performances. To describe the most dominating wavelet-based texture characteristics, we first extracted CS-LBP features from the non-overlapping local region in wavelet domain and applied SVM-RFE which selects a small subset of most relevant features. Then we applied these selected features to decision trees, construct random forests which are an ensemble of random decision trees. Further, using same extracted features from small set of known ground truth, we trained the random forest classifier, and this trained model managed the remaining mammogram in different libraries. At the retrieval time, trained model predicted the category of the query and searching is automatically narrowed down in the predicted cluster or library. This work was evaluated on MIAS database and found advantageous in terms of feature dimension, fast modelling and prediction, significantly encouraging classification as compared to state-of-art methods, and significantly encouraging retrieval accuracy and response time as compared to the conventional approach.

Chapter 5 resolved the issues of artifacts suppression, and segmentation by using the application of connected component labelling, morphological image processing, adaptive k-means and seeded region growing algorithm. There are four key contributions of the work presented in this chapter. Firstly, fully automated removal of artifacts and pectoral muscles. Secondly, modified region growing for effective

segmentation which provided perfect breast contour representation for breast profile region. Further, designed wavelet-based CS-LBP features captured the strong texture characteristics and finally introduced SOM clustering based retrieval framework. This framework retrieved the most relevant mammograms in less searching time as compared to the traditional exhaustive search method.

6.2 Future Work

The research work presented in this thesis can help further into different directions. The scope for future works are as follows:

- Despite the advance in CBIR system, there is still significant gap towards semantic-aware retrieval from visual contents. To proceed towards semantic-aware retrieval, scalable supervised or semi-supervised learning are promising to learn semantic aware representation so as to boost the content-based retrieval quality.
- Deep learning techniques can be applied for reducing the semantic gap. Since computational efficiency and memory overhead are emphasized in particular in CBIR, it would be beneficial to consider those constraints in the structure design of deep learning networks.
- New similarity measures can be defined for further improvement of the retrieval.
- For the application of mammogram retrieval, the proposed methods can be tested in an extensive collection of mammograms that are pooled from diverse databases. In the case of feature extraction, a bigger neighbourhood and new variant of texture features can be thought of, and more localized features from digital mammograms can be taken. For the segmentation of mammogram, dynamically thresholds as per the input mammogram can be designed and

according to these thresholds termination criteria for region growing algorithm are estimated.