# PREFACE

With the advancement in internet and multimedia technologies, a huge amount of multimedia data in the form of audio, video, and images has been used in many fields like medical treatment, satellite data, video and still images repositories, digital forensics, and surveillance system. This has created an on-going demand for systems that can store and retrieve multimedia data in an effective way. The development of efficient content-based image retrieval (CBIR) system is a growing research area where relevant images are retrieved as per the content of the query. This system is applied in various area i.e. medical, remote sensing, geographical information system, journalism, crime detection, advertising, education, and entertainment, etc. Many multimedia information storage and retrieval systems have been developed till now for catering these demands. An effective CBIR needs efficient extraction of low-level features, and for this many different methods have been recently proposed using colour, texture, and shape features.

In this thesis, the problems of semantic gap, searching time, fast indexing, and accurate retrieval are addressed, and finally, the application of CBIR for the diagnosis of breast cancer is taken.

In the last several years, developing computer-aided diagnosis (CAD) schemes that use CBIR to search for the clinically relevant and visually similar medical images depicting suspicious lesions has also been attracting researcher's interest. In mammographic context, CBIR can contribute more reliable diagnosis by classifying the query mammograms and retrieving similar mammograms already annotated by diagnostic descriptions and treatment results, which help radiologist to interpret and analyse the current case with historical cases. Though the retrieval of a mammogram for

the diagnosis of breast cancer is a well-explored problem in the field of computer vision but still many challenges remain to be investigated. These challenges include reliable feature extraction, pre-processing efforts for the removal of artefacts, scratches, labels, and pectoral muscles. Further, for segmentation, it is necessary to find the accurate region of interest from the mammogram. To address the issues of CBIR system for general images as well as for mammogram application, comprehensive literature review and comparative study of various conventional as well as state-of-the art methods are given. Further, for resolving these issues, the major contribution of this thesis is divided into three parts.

Firstly, this thesis work focuses on the problem of the semantic gap and searching time for general images. As we know that, the colour feature is quite remarkable for the implementation of CBIR but due to achromatic surfaces or unevenly colored objects, the role of texture is also important. This thesis introduces an efficient and fast CBIR system, which is based on the combination of computationally light weighted colour and texture features viz. chromaticity moments, colour percentile, and local binary patterns. For searching, this work proposes inverse variance based varying weighted similarity measure (low for high variance feature and high for low variance feature), which reduces the effect of redundancy by assigning the priority to each feature and effectively retrieves relevant images. A query image classification and retrieval model by filtering out irrelevant class images using random forests classifier is also proposed, which recovers the class of a query image based on distinct learning (supervised) of various decision trees. This successful ensemble classification of query images reduces the semantic gaps, searching space, and enhances the retrieval performance. Extensive experimental analysis on two benchmark databases confirms the usefulness and effectiveness of this work.

Taking the same concern, an additional fast and effective CBIR system is proposed which uses supervised learning based image management and retrieval techniques. It utilizes machine learning approaches as a prior step for speeding up image retrieval in the large database with enhanced accuracy. For the implementation of this, method, at first statistical moments and the orthogonal combination of local binary pattern based computationally light weighted colour and texture features are proposed. Further, using some ground truth annotation of images, the multi-class support vector machine (SVM) classifier is trained. This classifier works as a manager and categorizes the remaining images into different libraries. However, at the query time, same features are extracted and fed to the SVM classifier. SVM detects the class of query and searching space is narrowed down to the corresponding library. This supervised query image classification and retrieval model filters out maximum irrelevant images and reduces the searching time significantly. This work is evaluated and compared with the conventional model of the CBIR on two benchmark databases, and it is found that the proposed work is significantly encouraging in terms of retrieval accuracy and response time for the same set of features.

Secondly, this thesis work focuses on a mammogram classification cum retrieval framework for the diagnosis of breast cancer. This work firstly proposes an effective method for classifying mammograms using random forests with wavelet-based center-symmetric local binary pattern (WCS-LBP) then for the classification of mammograms, multi-resolution CS-LBP texture characteristics i.e. features from non-overlapping regions of the mammograms are captured. Further, we examine most relevant features using support vector machine- recursive feature elimination (SVM-RFE) based method. Finally, we feed the selected features to decision trees and construct random forests which is an ensemble of random decision trees. Using wavelet-based local CS-LBP

features with random forest, the test images are classified into different categories having the maximum posterior probability. Further, we have used this classification accuracy and retrieved most relevant mammograms in less searching time.

Finally, this thesis includes automated and effective content-based mammogram retrieval using wavelet-based CS-LBP feature and self-organizing map. This fully automated approach supports radiologists in their decision to find similar mammograms out of a database to compare the current case with past cases. Labels, scratches, and pectoral muscles present in mammograms can bias the retrieval procedures. For the removal of these, generally manual cropping is performed which is labor intensive and time-consuming process. In this work, an automated fast and effective content based-mammogram image retrieval system is proposed. The proposed pre-processing steps include automatic labelling-scratches suppression, automatic pectoral muscle removal, and image enhancement. For segmentation, selective thresholds based seeded region growing algorithm is introduced. Further, 2-level discrete wavelet transform (DWT) is applied on the segmented region, and WCS-LBP features are extracted. Then, extracted features are fed to self-organizing map (SOM) which generates clusters of images, having similar visual content. SOM produces different clusters with their centers and query image features are matched with all cluster representatives to find the closest cluster. Finally, images are retrieved from this closest cluster using Euclidean distance similarity measure. So, at the searching time, the query image is searched only in small subset depending upon cluster size and is not compared with all the images in the database and it reflects a superior response time with good retrieval performances.