Intelligent Decision Support Systems for cancer diagnosis through state of the art machine learning algorithms

Budur Ibrahim Alodhaybi

September 2016

Dissertation submitted in partial fulfilment for the degree of
Master of Science in Information Technology

Computing Science and Mathematics
University of Stirling

Abstract

Breast cancer is a dangerous disease that exists in breast tissue. The fatality rate of this cancer can be decreased by diagnosing it at an early stage and then applying the optimal treatment. The computer-aided decision support system (CADSSs) consists of the following four steps: pre-processing for images, feature extraction, feature selection, and classification, together with mammography, will help doctors diagnose breast cancer at the early stage. For the recognition of abnormalities in mammograms, the using of a novel local energy-based shape histogram (LESH) as the feature set has been studied in some detail. The study also explores the use of this technique on the mammogram datasets of the Mammographic Image Analysis Society (MIAS). In the estimation, regions of interest (ROI) were extracted from the mammograms. Then, their LESH features were calculated, and used to support the vector machine (SVM) classifiers. Furthermore, the effect on classification performance of choosing a subset of LESH features was recognized and benchmarked alongside the state-of-the art feature extraction method. The suggested method achieved a greater classification accuracy of 99% and the Az value achieved 0.989 with multiple SVM kernels. The linear kernel achieved 99.09 accuracy for discrimination between the abnormalities (masses vs. micro-calcifications). Another experiment was performed using the Echo State Network (ESN) as another classifier to identify the abnormalities in mammogram images. Abnormalities in mammograms can be recognized as different types. An effective system can thus consider all these abnormalities and produce a correct diagnosis. The experiment was conducted with Local Energy based Shape Histogram (LESH) features together with the Echo State Network classifier. The recommended system achieved a high classification accuracy of 98% as well as a good rate of sensitivity and specificity. Then the comparative of the achievement of ESN with the Support Vector Machine (SVM) and the other classifiers and results created showed that ESN had performed well with the benchmark classifier. The good rate of Sensitivity and Specificity also shows the capability of the ESN classifier to detect positive and negative cases correctly. As a final experiment Logistic Regression was applied to the same dataset MIAS and it achieved a good result of approximately 72.72% positive accuracy for discriminating the abnormalities in the mammogram dataset.

Hence, the general capabilities of the suggested methods were recognized, proving that they only differentiate between malignant and benign cases for any kind of abnormality but also have an ability to differentiate between all the types of abnormalities. Using LESH features with different classifiers is consequently a good choice for extracting important clinical information from mammogram pictures. What is more, there is the exciting possibility of applying these techniques 3-D MRI images for further use and investigation.