

Abstract

Breast cancer is the most common life-threatening type of cancer affecting women in The Netherlands. About 10% of the Dutch women have to face breast cancer in their lifetime. The success of the treatment of breast cancer largely depends on the stage of a tumor at the time of detection. If the size of the invasive cancer is smaller than 20 mm and no metastases are found, chances of successful treatment are high. Therefore, early detection of breast cancer is essential. Although mammography screening is currently the most effective tool for early detection of breast cancer, up to one-fifth of women with invasive breast cancer have a mammogram that is interpreted as normal, i.e., a false-negative mammogram result. An important cause are interpretation errors, i.e., when a radiologist sees the cancer, but classify it as benign. In addition, the number of false-positive mammogram results is quite high, more than half of women who undergo a biopsy actually have breast cancer.

To overcome such limitations, Computer-Aided Diagnosis (CAD) systems for automatic classification of breast lesions as either benign or malignant are being developed. CAD systems help radiologists with the interpretation of lesions, such that they refer less women for further examination when they actually have benign lesions.

The dataset we used consists of mammographic features extracted by automated image processing algorithms from digitized mammograms of the Dutch screening programme. In this thesis we constructed several types of classifiers, i.e., Bayesian networks and support vector machines, for the task of computer-aided diagnosis of breast lesions. We evaluated the results with receiver operating characteristic (ROC) analysis to compare their classification performance. The overall conclusion is that support vector machines are still the method of choice if the aim is to maximize classification performance. Although Bayesian networks are not primarily designed for classification problems, they did not perform drastically lower. If new datasets are being constructed and more background knowledge becomes available, the advantages of Bayesian networks, i.e., incorporating domain knowledge and modeling dependencies, could play an important role in the future.