Development and Comparative Analysis of Machine Learning Algorithms for Breast Cancer Detection

Huang, Yuchen

Breast cancer is the second only to lung cancer as a cause of cancer deaths in American women and causes around 400,000 deaths per year worldwide. The primary screening tool for breast cancer is mammography (85-90% accurate). The objective of this project is to create a new Machine Learning (ML) method for breast cancer detection that gives a higher accuracy than the existing methods. Binary Sum of Products (SOP) method was developed and compared with 3 other methods - Decision Trees, Naïve Bayes, Support Vector Machines (SVM). In contrast to other methods, SOP hasn't been used for breast cancer detection in the past. I developed a program that implements and compares the four methods. In addition, I used the Wisconsin Breast Cancer Database, which is used in all top world researches in this area. All methods go through the training, classification and tuning phases. Results showed that using a 90%/10% training/testing split, SOP gave 99.5% accuracy, highest ever reported in the literature, in contrast to the three other methods (97.5%-99.3%). This shows that if given a large dataset, SOP is the best method. If given a small dataset, SVM outperforms the other methods by a wide margin. This is the first time SOP was used for breast cancer detection and the highly reliable Wisconsin data base was used. The results in this project give higher accuracy than any existing ML accuracy documented in literature. My research seems to show the advantage of using logic-based methods in medical diagnosis.

Awards Won: