Chapter 79

Classification Approach for Breast Cancer Detection Using Back Propagation Neural Network:
A Study

Aindrila Bhattacherjee

Bengal College of Engineering and Technology, India

Sourav Roy

Bengal College of Engineering and Technology, India

Sneha Paul

Bengal College of Engineering and Technology, India

Payel Roy

JIS College of Engineering, India

Noreen Kausar

Malaysia University of Science and Technology, Malaysia

Nilanjan Dey

Techno India College of Technology, India

ABSTRACT

According to the recent surveys, breast cancer has become one of the major causes of mortality rate among women. Breast cancer can be defined as a group of rapidly growing cells that lead to the formation of a lump or an extra mass in the breast tissue which consequently leads to the formation of tumor. Tumors can be classified as malignant (cancerous) or benign (non-cancerous). Feature selection is an important parameter in determining the classification systems. Machine learning methods are the most commonly used methods among researchers for breast cancer diagnosis. This paper proposes to investigate the WBCD (Wisconsin Breast Cancer Dataset) which comprises of 683 patients and implements the chosen features to train the back propagation neural network. The performance is then analyzed on the basis of classification accuracy, sensitivity, specificity, positive and negative predictor values, receiver operating characteristic curves and confusion matrix. A total of 9 features has been used to

classify breast cancer with an accuracy of 99.27%. According to the recent surveys, breast cancer has become one of the major causes of mortality rate among women. Breast cancer can be defined as a group of rapidly growing cells that lead to the formation of a lump or an extra mass in the breast tissue which consequently leads to the formation of tumor. Tumors can be classified as malignant (cancerous) or benign (non-cancerous). Feature selection is an important parameter in determining the classification systems. Machine learning methods are the most commonly used methods among researchers for breast cancer diagnosis. This paper proposes to investigate the WBCD (Wisconsin Breast Cancer Dataset) which comprises of 683 patients and implements the chosen features to train the back propagation neural network. The performance is then analyzed on the basis of classification accuracy, sensitivity, specificity, positive and negative predictor values, receiver operating characteristic curves and confusion matrix. A total of 9 features has been used to classify breast cancer with an accuracy of 99.27%.

INTRODUCTION

Breast cancer is the malignant tumor, which starts in the cells of the breast and can attack surrounding tissues or extend to remote areas of the body. Typically, breast cancer either starts in the cells of lobules that are the milk-producing glands, or the ducts which over time can occupy the nearby healthy breast tissue and create their way into the underarm lymph nodules or the small organs which filter out the foreign substances in the body. If the cancer cells get into the lymph nodules, then they have an alley way into the other parts of the body. There were more than 2.8 million women with the history of breast cancer in the U.S that comprises women presently being treated and the women who have completed the treatment in 2014. A woman's possibility of breast cancer more or less doubles if she has a first-degree relative like mother, sister, and daughter who has been analyzed with breast cancer. About 15% of women who get the breast cancer have a family member identified with it. About 5-10% of breast cancers can be connected to gene mutations hereditary from one's mother or father. Mutations of the BRCA1 and BRCA2 genes are the most ordinary. Women with the BRCA1 mutation have a 55-65% risk of increasing the breast cancer before age 70, and frequently at a younger age that it classically develops. For women with the BRCA2 mutation, this risk is about 45% (Acharya *et al*., 2008; Sengupta *et al*., 2014).

Although breast cancer is considered to be one of the most fatal diseases, it is also one of the most curable diseases if it can be diagnosed at an early stage. Early diagnosis needs a precise and reliable diagnosis procedure that allows physicians to distinguish between benign breast tumors and malignant ones (Abbadi *et al*., 2014). Since, the radiologists find difficulty in mammographic examination the scope of human error in precise diagnosis of the disease increases, thus resulting in the need for the development Computer Assisted Detection (CAD) (Wani *et al*., 2014). The concept of CAD involves taking into account the roles of both the computer and the physicians on equal grounds. The results of the computer need not be better than the result developed by the physicians; rather they need to be complementary. When lateral chest images are diagnosed with the help of CAD, the general presentation in the recognition of lung swellings is enhanced when integrated with another CAD scheme for testing PA chest images. CAD also finds its way in reliable and precise detection of vertebral fractures, thus making the early detection of osteoporosis possible (Doi, 2007).

The evolution of automated diagnosis was spurred by the need to assist physician in decision making. Machine learning methods have become indispensable in the field of medical science and hence are currently being employed in various classification methods due to their meticulous prediction performance.

Related Content

A General Knowledge Representation Model for the Acquisition of Skills and Concepts
Carlos Ramirezand Benjamin Valdes (2012). *Breakthroughs in Software Science and Computational Intelligence (pp. 28-45).*
www.irma-international.org/chapter/general-knowledge-representation-model-acquisition/64601

Multi-Objective Adaptive Manta-Ray Foraging Optimization for Workflow Scheduling with Selected Virtual Machines Using Time-Series-Based Prediction
Sweta Singh, Rakesh Kumarand Udai Pratap Rao (2022). *International Journal of Software Science and Computational Intelligence (pp. 1-25).*
www.irma-international.org/article/multi-objective-adaptive-manta-ray-foraging-optimization-for-workflow-scheduling-with-selected-virtual-machines-using-time-series-based-prediction/312559

Variants of VRP to Optimize Logistics Management Problems
Claudia Gómez Santillán, Laura Cruz Reyes, María Lucila Morales Rodríguez, Juan Javier González Barbosa, Oscar Castillo López, Gilberto Rivera Zarateand Paula Hernández (2012). *Logistics Management and Optimization through Hybrid Artificial Intelligence Systems (pp. 207-237).*
www.irma-international.org/chapter/variants-vrp-optimize-logistics-management/64923

An Islamic Distributed Information Retrieval Approach
 Falah Hassan Ali Al-akashi (2020). *International Journal of Software Science and Computational Intelligence (pp. 38-54).*
www.irma-international.org/article/an-islamic-distributed-information-retrieval-approach/258865

Computational Intelligence Techniques for Pattern Recognition in Biomedical Image Processing Applications
D. Jude Hemanthand J. Anitha (2012). *Machine Learning Algorithms for Problem Solving in Computational Applications: Intelligent Techniques (pp. 195-209).*
www.irma-international.org/chapter/computational-intelligence-techniques-pattern-recognition/67704