Chapter 71
Survey of Breast Cancer Detection Using Machine Learning Techniques in Big Data

Madhuri Gupta
Jaypeee Institute of Information Technology, Noida, India

Bharat Gupta
Jaypee Institute of Information Technology, Noida, India

ABSTRACT

Cancer is a disease in which cells in body grow and divide beyond the control. Breast cancer is the second most common disease after lung cancer in women. Incredible advances in health sciences and biotechnology have prompted a huge amount of gene expression and clinical data. Machine learning techniques are improving the prior detection of breast cancer from this data. The research work carried out focuses on the application of machine learning methods, data analytic techniques, tools, and frameworks in the field of breast cancer research with respect to cancer survivability, cancer recurrence, cancer prediction and detection. Some of the widely used machine learning techniques used for detection of breast cancer are support vector machine and artificial neural network. Apache Spark data processing engine is found to be compatible with most of the machine learning frameworks.

1. INTRODUCTION

As reported by World Health Organization (WHO, 2016), breast cancer is the most prominent problem in the area of medical diagnosis, which is increasing every year. A consistent advancement in technology has been accomplished for breast cancer research (Hanahan & Weinberg, 2011). Researchers has applied different methods, for example screening and biopsy, to discover different stages of breast cancer before symptoms occur. An unprecedented amount of healthcare data (Marx, 2013) is produced by the plethora of technology such as magnetic resonance imagery (MRI), super-resolution digital microscopy, mass spectrometry, etc. These technologies mainly provide healthcare data, but their focus is not on analysis,

DOI: 10.4018/978-1-6684-6291-1.ch071

knowledge extraction or interpretation (Rider & Chawla, 2013). Therefore, there is need of data storing, data pre-processing and data management in medical research. Medical science is also ushering into the field of big data and there is a need to analyze the huge amount data by applying machine learning techniques (Mattmann, 2013). The effectiveness of big data and machine learning approaches applies appropriate methods to create efficient models for analysis. The fusion of data has significantly supported data-oriented research in breast cancer (BC) field (Dinov, 2016). It involves diagnosis and prediction of human-threatening disease.

Breast cancer detection is one of the main priorities in medical research, because mortality rate in India is growing fast (breast cancer statistics presented by Globocan Project, 2016). Here, mortality rate is the number of women who died because of breast cancer in a particular year. Breast cancer certainly generates large amount of gene expression data. So, this disease has encouraged interest in the improvement of machine learning, data analysis techniques and tools which can accurately extract information from massive data. Therefore, in breast cancer research, data analytics and machine learning approach are the main concern when it arises to management, diagnosis and other clinical aspects. Hence, it is necessary to review the current literature on data analytics and machine learning approaches in breast cancer research.

This paper has mainly 6 sections: section 2 represents an aspect of the breast cancer disease, section 3 provides the necessary background knowledge on machine learning (ML), section 4 introduces big data analytics, section 5 provides challenges, and section 6 represents conclusion and future work.

2. Breast Cancer

Cancer is a disease that causes cells in the body to grow and change uncontrollably. Breast cancer is one of them (Breast State, 2017). Majority of breast cancer cases begin in breast tissues which are made up of lobules (glands), and ducts. The remaining portion of breast is made up of lymphatic and fatty tissues. Most of the time, breast cancer is detected during breast screening or after a patient notices a lump. These breast lumps can be benign (non-cancerous) and malignant (cancerous). Breast cancer can arise in any area of the breast, the lobules, the ducts and sometimes, the tissue in between. This section deals with different types and stages of breast cancer (Breast cancer Symptoms, 2017).

Types of Breast Cancer

1. **Non-invasive:** Cancer in duct and lobules in breast, they do not spread in normal tissue within breast. It is also called pre-cancers or carcinoma in situ (CIS).
2. **Invasive**: Cancer spread in normal tissue of breast. Most breast lumps are invasive.
3. **Multifocal**: Original tumor divides itself, but remains in the same section of the breast.
4. **Multicentric**: All the tumors arise independently in different region of the breast.
5. **Recurrent:** Cancer that arise again in the same or opposite breast after a period of time when the cancer couldn't be detected.
6. **Metastatic:** Cancer spread to other portions of body. It is advanced stage cancer.
7. **Paget's disease:** In this disease, cancer cells collect around the nipple. This type of cancer, first affects the ducts of the nipple, and then spreads to the nipple.