

Chapter 61

Prediction of Breast Cancer Recurrence With Machine Learning

Mohammad Mehdi Owrang O.

American University, USA

Ginger Schwarz

American University, USA

Fariba Jafari Horestani

American University, USA

ABSTRACT

Medical prognostication is the science of estimating the complication and recurrence of a disease. A Breast cancer recurrence (BCR) event is characterized by the cancer “coming back” after at least a year of remission after the treatment. Many factors, including tumor grade, tumor size, and lymph node status may influence or correlate with prognosis for breast cancer patients. Early detection of recurrence events (i.e., while still asymptomatic) is more likely to be curable than after the cancer symptoms are seen again. Machine learning techniques can help to provide some necessary information and knowledge required by physicians for accurate predictions of BCR and better decision-making. The aim of this chapter is to use machine learning classifiers to examine the factors that are most predictive of the BCR. Several attributes/features selection schemes have been used to find the most significant features contributing to BCR. Five different machine learning algorithms were tested and compared for the prediction of BCR. The decision tree was found to be the best model for the dataset.

INTRODUCTION

Breast cancer is the most common female cancer in the U.S., the second most common cause of cancer death in women (“American Cancer Society”, 2021), and the main cause of death in women ages 40 to 59 (Siegel et al., 2012). In 2023, an estimated 297,790 new cases of invasive breast cancer are expected to be diagnosed in women in the U.S., along with 55,720 new cases of non-invasive (in situ) breast cancer (“Cancer.net”, 2023). In 2023, about 43,170 women in the U.S. were expected to die from breast cancer.

DOI: 10.4018/978-1-6684-7366-5.ch061

There is a one in six probability that a woman will develop breast cancer in their lifetime (one in eight for invasive disease) (“American Cancer Society”, 2021; Siegel et al., 2012). Breast cancer incidence is increasing while mortality is declining in many high-income countries (Torre et al., 2016). In recent years, great progress has been made in the understanding of breast cancer, with new classification techniques that have significant prognostic value and provide guides to treatment options (Ahmed et al., 2013; Lafourcade et al., 2018).

Medical prognostication is an essential part of medicine that encompasses the science of estimating the complication and recurrence of disease and predictive survival of patients (Ohno, 2001). Medical prognosis plays an increasing role in health care outcomes. Many factors, including tumor grade, tumor size, and lymph node status may influence or correlate with prognosis for breast cancer patients (Bradley, 2007; Lafourcade et al., 2018).

Early detection of breast cancer allows doctors to treat the disease aggressively giving women a better chance of survival. After the diagnosis and treatment of the primary incidence of breast cancer, the cancer may come back (recurrence event) after a year in the same organ (local recurrence), or in a close-by-organ (regional recurrence) or in another part of the body (distant recurrence) (Fan et al., 2010; Ahmad et al., 2013; Guo et al., 2017; Alva, 2018; Lafourcade et al., 2018). However, local, or distant recurrence can occur at any time. In cancer, relapse-free survival is the length of time after primary treatment for a cancer that the patient survives without any signs or symptoms of that cancer. It is one way to measure how well a new treatment works.

Because of the high prevalence and mortality of breast cancer among women, it is important and necessary to explore and develop new techniques to help predict a patient’s chance for recurrence and develop a treatment process to prevent it. A significant topic in breast cancer research is understanding and improving the effectiveness of breast cancer treatments thus increasing relapse-free survival time. In addition, the practical application of data mining can help to provide some necessary information and knowledge required by physicians for accurate prediction of Breast Cancer Recurrence (BCR) and better decision-making.

Current studies reported in the literature have utilized machine learning/data mining techniques to predict breast cancer recurrence, and to identify asymptomatic women at risk of cancer recurrence (Mannell, 2017; Mulatu & Gangardle, 2017; Mosayebi et al., 2020; Lou et al., 2020; Massafra et al., 2021). However, it is still an open task which requires the gathering of suitable and quality datasets, using proper features selection schemes, and defining more accurate prediction models.

To predict the probability of recurrence, it is important to know which features/factors of a patient have the highest risk factors (Chang & Lin, 2008; Mannell, 2017; Darst et al., 2018; Lafourcade et al., 2018). This chapter presents the preliminary results of a feature important analysis collected by patient records aimed to develop an automated system to predict BCR. The goal of this study is to use machine learning classifiers to examine the factors that are most predictive of the BCR. The main objective is to compare different data mining/machine learning algorithms to select the most accurate model for predicting BCR.

BREAST CANCER RECURRENCE OVERVIEW

Medical prognosis is a field in medicine that includes estimating the complication and recurrence of disease and predicting survival of a patient (Bradley, 2007; Ohno-Machado, 2001). Survival analysis is

31 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/prediction-of-breast-cancer-recurrence-with-machine-learning/332794

Related Content

Evaluation of Information Infrastructures and Social Development Among the Visegrad-Four Countries of Central Europe

László A. Pookand Norman E. Pence (2004). *Journal of Global Information Management* (pp. 63-83).

www.irma-international.org/article/evaluation-information-infrastructures-social-development/3607

Mapping the Diffusion of the Internet Technology Cluster: An Examination of Irish SMEs

James Griffin (2008). *Global Information Technologies: Concepts, Methodologies, Tools, and Applications* (pp. 1795-1805).

www.irma-international.org/chapter/mapping-diffusion-internet-technology-cluster/19074

Warehouse Management Systems: Comparison of Two Pittsburgh-Based Manufacturing Firms

Alan D. Smith (2025). *Encyclopedia of Information Science and Technology, Sixth Edition* (pp. 1-17).

www.irma-international.org/chapter/warehouse-management-systems/320496

Modeling the Free/Open Source Software Community: A Quantitative Investigation

Gregory Madey, Vincent Freehand Renee Tynan (2008). *Global Information Technologies: Concepts, Methodologies, Tools, and Applications* (pp. 3296-3298).

www.irma-international.org/chapter/modeling-free-open-source-software/19179

How Does Time Synchronization Affect Innovation Quality?: Based on Information and Knowledge Acquirement

Xianpeng Si, Ying Liu, Jie Wu, Chia-Huei Wuand Wei Liu (2024). *Journal of Global Information Management* (pp. 1-20).

www.irma-international.org/article/how-does-time-synchronization-affect-innovation-quality/344045