RADemics

Machine Learning Techniques for Early Breast Cancer Detection and Risk Prediction

Pravin Gopalrao Sarpate, Shaik Balkhis Banu
Gopikabai Sitaram Gawande Mahavidyalaya
Umarkhed, Fatima college of health sciences, Al Ain

Machine Learning Techniques for Early Breast Cancer Detection and Risk Prediction

[1]Pravin Gopalrao Sarpate, Assistant Professor, Department of Computer Science, Gopikabai Sitaram Gawande Mahavidyalaya Umarkhed, District Yavatmal, Maharashtra, India. pravingsarpate@gmail.com

[2]Shaik Balkhis Banu, Assistant Professor, Department of Physiotherapy, Fatima college of health sciences, Al Ain, United Arab Emirates. drshaikbalkhis@gmail.com

Abstract

Breast cancer remains one of the leading causes of cancer-related morbidity and mortality worldwide, underscoring the urgent need for effective early detection and accurate risk prediction. Recent advancements in machine learning (ML) have shown considerable promise in transforming breast cancer diagnosis, offering new avenues for improving detection accuracy and predicting individual patient risk. This chapter explores the integration of machine learning techniques in breast cancer detection, focusing on the application of ML models for tumor classification, risk prediction, and prognosis. Emphasis is placed on the combination of medical imaging data, such as mammograms and MRIs, with clinical and genomic information to develop robust hybrid models that enhance diagnostic precision. Key challenges, including data imbalance, model interpretability, and the need for standardization, are discussed, along with emerging solutions that address these barriers. The role of radiomics and advanced data augmentation techniques in improving model performance is also highlighted. Through the development of more accurate, explainable, and personalized machine learning models, this chapter aims to shed light on the future of breast cancer care, where early detection, personalized treatment, and improved patient outcomes can be achieved.

Keywords: Breast Cancer, Machine Learning, Risk Prediction, Tumor Classification, Radiomics, Hybrid Models.

Introduction

Breast cancer continues to be one of the most common cancers globally, accounting for a substantial portion of cancer-related morbidity and mortality among women [1]. Early detection is critical to improving survival rates, as early-stage cancers are more treatable and less likely to result in metastasis [2]. Traditional diagnostic methods, such as mammography, ultrasound, and biopsy, have played a central role in identifying breast cancer over the years [3]. While effective in many cases, these methods have limitations in sensitivity, accuracy, and the ability to detect early-stage tumors, particularly in dense breast tissue [4]. As the demand for more precise, non-invasive, and timely diagnostic tools increases, machine learning (ML) has emerged as a promising solution to address these challenges and transform the landscape of breast cancer detection and risk prediction [5].

Machine learning offers a significant advantage by automating the analysis of complex data and identifying subtle patterns that might be missed by traditional diagnostic techniques [6]. In medical imaging, ML models can be trained to analyze large volumes of data from mammograms, MRIs, and ultrasound images, extracting key features related to tumor size, shape, texture, and density [7]. These models can not only detect visible lesions but also classify the type of tumor and predict its potential malignancy [8]. With the integration of clinical data such as age, family history, and genetic predisposition, ML models can provide personalized risk assessments, offering a more comprehensive approach to cancer diagnosis [9]. This personalized approach ensures that individuals receive the most appropriate screening and treatment options based on their specific risk factors [10].

The integration of imaging data with clinical and genomic information has given rise to hybrid machine learning models that offer superior accuracy in breast cancer risk prediction and detection [11]. These models combine multiple data sources to generate more reliable and precise predictions [12]. For instance, radiomic analysis extracts quantitative features from medical images, which, when integrated with clinical data such as genetic markers or family history, can significantly improve risk prediction accuracy [13]. Hybrid models offer the advantage of multi-dimensional data analysis, which allows healthcare professionals to make more informed decisions about patient care [14]. As these models evolve, they are expected to play an increasingly critical role in the early detection of breast cancer and the development of personalized treatment plans [15].