Dynamic Modeling for Breast Cancer Risk Prediction and Demographic Stratification

Shamiha Binta Manir, PhD Student, Marquette University
Priya Deshpande, PhD

Introduction

Breast cancer represents a substantial health threat to women worldwide, necessitating innovative predictive tools to enable early detection and personalized care. This research leverages a time-adjusted risk prediction model combined with demographic clustering to support targeted public health interventions. Utilizing the Breast Cancer Surveillance Consortium (BCSC) dataset, comprising over 6 million mammograms and 1 million detailed demographic records, we explore dynamic risk modeling to enhance prediction accuracy and reveal critical demographic trends.

Hypothesis

We hypothesize that incorporating temporal demographic variables into machine learning frameworks will improve predictive accuracy and identify high-risk population clusters. By tracking changes in age, BMI, and menopausal status over time, we aim to construct dynamic risk trajectories and design customized interventions for diverse demographic groups.

Methods

A time-adjusted risk scoring system will be implemented using recurrent neural networks (RNNs) and transformer-based architectures to capture temporal variations in risk factors. Annual demographic changes, including age and hormonal transitions, dynamically update risk profiles. Clustering methods, such as K-means and hierarchical clustering, will be applied to identify unique population subgroups with distinct risk profiles. To address data imbalance, resampling techniques like Tomek Links are employed to optimize classifier training. Robustness and generalizability are ensured through cross-validation across demographic segments.

Results

Preliminary findings indicate that the Random Forest classifier achieves 87.47% accuracy using Tomek Link resampling. Clustering analysis identifies distinct high-risk cohorts, including post-menopausal women with elevated BMI, enabling precision-focused public health strategies. Additionally, the ROC curve demonstrates the model's predictive capability, achieving an AUC of 0.81, signifying strong classification performance.

Conclusion

This study highlights the potential of combining time-adjusted modeling with demographic clustering for breast cancer risk prediction. By dynamically refining risk profiles and identifying high-risk groups, the proposed approach advances precision medicine and offers scalable solutions for early detection and strategic public health planning.

Figure(s)

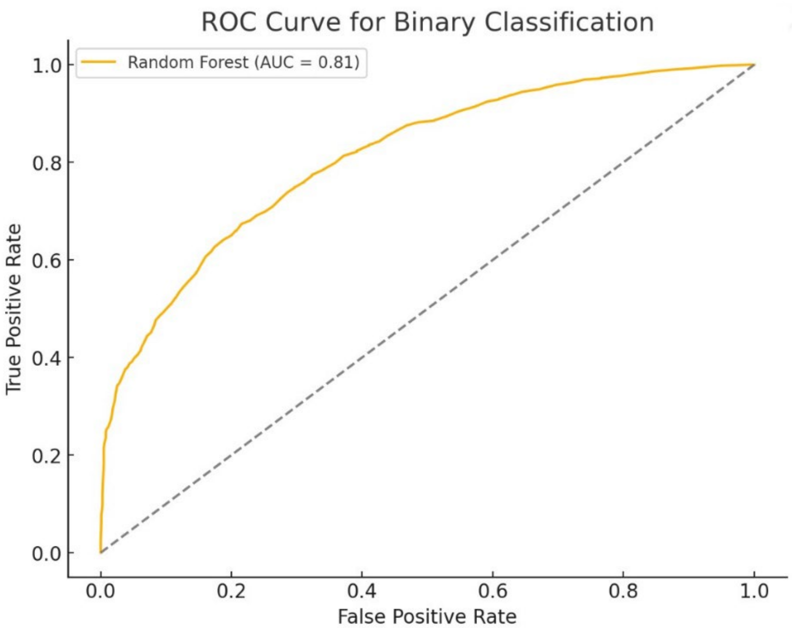

Figure 1. ROC Curve Demonstrating the Predictive Performance of the Random Forest Classifier for Breast Cancer Risk Prediction (AUC = 0.81)

Keywords

Applications; Artificial Intelligence/Machine Learning; Emerging Technologies; Imaging Research; Quality Improvement & Quality Assurance