

BREAST CANCER DIAGNOSIS PREDICTOR APPLICATION

CHAPTER 1- LITERATURE REVIEW

STUDENT NAME: JOHNSON KANYI WAMWEYA

REGISTRATION NUMBER: BCSC01/0019/2021

2.1 Introduction

Breast cancer is among the most common types of cancer worldwide right now, necessitating advanced diagnostic tools to aid in early detection and treatment planning. Machine learning and Artificial Intelligence (AI) have shown tremendous potential in medical diagnostics, particularly in cancer detection. This chapter reviews existing literature and related systems in breast cancer diagnosis, highlights their limitations, and discusses how the proposed **Breast Cancer Diagnosis Predictor Application** addresses these challenges.

2.2 Related Systems

Several systems have been developed to assist in breast cancer diagnosis. Below are two prominent examples:

2.2.1 Breast Cancer Risk Assessment Tool (BCRAT)

The BCRAT is a statistical tool developed to estimate a woman's risk of developing breast cancer by using patient-provided data, including age, reproductive history, and family medical history, to calculate a risk score.

While this tool is widely used for risk assessment, it does not provide diagnostic predictions for existing breast masses.

2.2.2 Wisconsin Breast Cancer Diagnosis (WBCD) Dataset-Based Models

The Wisconsin Breast Cancer Diagnosis dataset has been used extensively for developing machine learning models aimed at classifying breast masses as benign or malignant. These models rely on diagnostic measurements, such as cell radius, texture, and perimeter, to train algorithms like support vector machines (SVMs) and neural networks.

They typically serve as research prototypes rather than user-friendly applications for usability and UX for clinical use.

2.3 Limitations of the Existing Systems

The above-mentioned systems possess notable limitations; they include the following:

2.3.1 Breast Cancer Risk Assessment Tool (BCRAT)

- ***Focus on Risk Prediction:*** The BCRAT estimates *future risk* but does *not* *diagnose existing breast masses*, limiting its utility in immediate diagnostic workflows.
- ***Lack of Machine Learning Integration:*** The tool uses statistical modeling rather than AI or machine learning, restricting its ability to adapt to *complex patterns* in diagnostic data.

2.3.2 Wisconsin Breast Cancer Diagnosis (WBCD) Dataset-Based Models

- ***Limited Clinical Usability:*** These models are often developed for *academic research* and are *not deployed as user-friendly clinical applications*.
- ***Narrow Focus:*** Many models *lack visualization features or user interfaces* that support medical professionals in interpreting the results.
- ***Generalizability Issues:*** The models are *trained on limited datasets* and may *not perform well across diverse patient demographics or real-world clinical scenarios*.

2.4 Addressing Weaknesses with the Proposed Solution

The **Breast Cancer Diagnosis Predictor Application** aims to address the above limitations by providing a comprehensive, user-friendly, and clinically relevant tool. The following key features illustrate how it resolves the identified weaknesses:

✓ Diagnostic Predictions Instead of Risk Assessment

Unlike the BCRAT, which focuses on future risk estimation, the proposed application will provide real-time diagnostic predictions for breast masses based on measurable features enabling immediate decision-making in lab settings.

✓ *Integration of Machine Learning Models*

The application will leverage a machine learning algorithm trained on a comprehensive dataset to identify complex patterns in diagnostic measurements to ensure a greater predictive accuracy and adaptability compared to traditional statistical methods.

✓ *User-Friendly Interface for Clinical Use*

To address the usability limitations of WBCD-based models, the application will include a well-designed interface that allows medical professionals to manually input diagnostic data, view predictions, and interpret results through a visual aid that is a radar chart

✓ *Improved Generalizability and Scalability*

The application's machine learning model will be trained on diverse datasets to ensure robust performance across different patient populations.

Furthermore, the system is designed with scalability in mind, allowing for future integration with laboratory equipment for automated data input.

✓ *Enhanced Visualization for Better Interpretation*

By incorporating the visual aid and probability scores, the application will bridge the gap between algorithmic predictions and human interpretation, making it easier for medical professionals to understand the results to help them make a decision.

References

- American Cancer Society. (2022). *Breast cancer statistics*. Retrieved from <https://www.cancer.org>
- Gail, M. H., & Greene, M. (1989). *Breast cancer risk calculation methods: The Gail Model*. Journal of the National Cancer Institute, 81(24), 1879-1886. doi:10.1093/jnci/81.24.1879

- Street, W. N., Wolberg, W. H., & Mangasarian, O. L. (1993). *Nuclear feature extraction for breast tumor diagnosis*. Retrieved from ([PDF](#)) [Nuclear Feature Extraction For Breast Tumor Diagnosis](#)
- World Health Organization. (2023). *Breast cancer fact sheet*. Retrieved from <https://www.who.int>