WEB-BASED SOLUTION FOR BREAST CANCER DETECTION USING MACHINE LEARNING AND DEEP LEARNING

Abhishek Chourasia Shivpratap Singh Panwar

ABSTRACT

One of the main causes of death for women is breast cancer. Breast cancer is considered the most serious condition affecting women. A study by IAFR revealed that globally, breast cancer has surpassed lung cancer as the most prominent cancer found in women. Initial detection of this is extremely important, as it can save the patient's life by preventing it from spreading further. Appropriate treatment can save lives because the fatality rate is very high if it has spread to a significant area. Medical data is sophisticated, and even a small change in data can make a huge difference. This needs accurate and powerful techniques that can detect even the smallest changes and variations in the data as well as understand the underlying trends and patterns of the complex data. As the quantity of data is increasing day by day, data-backed algorithms from ML and DL can be a game-changing factor in the industry. This study brings both ML and DL techniques together to understand their performance on the Wisconsin breast cancer dataset. We named our model Web-BCD. The main objective is to check the effectiveness and efficiency of data classification, as measured by accuracy, precision, sensitivity, and specificity. After using more than half a dozen ML techniques, the XGboost outperformed other techniques, while the DL method, which was a simple neural network, was found to be prone to overfitting breast cancer data. It performed terribly and gave only 37% accuracy on test data.

Keywords: Web-BCD, Machine Learning, , Breast Cancer Detection, Wisconsin, Algorithms.

INTRODUCTION

Millions of women worldwide suffer from breast cancer, a disease that has a major impact on both individual and societal health. As per the World Health Organization (WHO), there will be approximately 2.3 million new cases of breast cancer diagnosed in 2020 alone, making it the most common cancer among women worldwide [1]. Successful treatment and higher survival rates depend on early detection, underscoring the significance of efficient screening and diagnostic techniques.

Conventional methods for detecting breast cancer have primarily relied on imaging modalities such as magnetic resonance imaging (MRI), ultrasound, and mammography. Mammography has served as the cornerstone of breast cancer screening programs for decades. Its effectiveness is attributed to the potential for early detection of abnormalities, which is believed to contribute to a reduction in breast cancer mortality rates.[2]. However, these methods have limitations, including false-positive and false-negative results, discomfort, and cost [3]. Moreover, access to these diagnostic tools may be limited in resource-constrained settings, underscoring the need for alternative approaches that are cost-effective, scalable, and accessible to all populations.

Recent technological developments, especially in the domains of deep learning (DL) and machine learning (ML), have opened up new avenues for increasing the efficacy and accuracy of breast cancer detection. In order to evaluate huge datasets and extract significant patterns suggestive of malignancy, machine learning (ML) techniques like logistic regression, support vector machines (SVMs), and random forests have been used [4]. These algorithms analyse various features extracted from breast cancer images, such as mean radius, texture, area, concavity, compactness, fractal dimension, and others, to estimate the likelihood of malignancy.

In this study, we use a large dataset of 30 characteristics taken from digitalized mammograms to investigate the effectiveness of classical machine learning algorithms, specifically logistic regression, in recognizing signs of breast cancer. These features, encompassing mean fractal dimension, smoothness, compactness, concavity, concave points, radius, texture, perimeter, area, texture error, and others, capture various morphological, textural, and geometric properties of breast lesions. This information allows algorithms to assess the likelihood of malignancy.

Furthermore, to enhance the accessibility and usability of our breast cancer detection system, we plan to deploy our trained machine learning models using Flask, a lightweight web framework in Python. For the frontend interface, Using HTML, CSS, and JavaScript, we will build a simple and easy-to-use application that lets users submit mammography pictures and get instantaneous estimates of their risk of developing breast cancer.

Overall, our research shows that the XGBoost classifier performs best on the Wisconsin Breast Cancer Dataset, while deep learning models tend to overfit. To ensure reliability and scalability in our web-based breast cancer detection system, we prioritize implementing the XGBoost classifier. Our aim is to enable early cancer detection with a user-friendly interface, enhancing accessibility and usability for users.

TOOLS USED

Python is an excellent programming language for implementing ML and DL algorithms due to its extensive libraries, such as scikit-learn, TensorFlow, and Keras. Jupyter Notebook or Google Colab for interactive development and experimentation with ML and DL models. Visual Studio Code (VS Code) for code development, debugging, and version control management. Flask framework for building web applications to deploy ML and DL models. HTML, CSS, and JavaScript for designing and implementing user interfaces for web-based model deployment. Data preprocessing libraries such as Pandas and NumPy for handling CSV dataset and performing data cleaning, normalization, and feature scaling. For data visualization and exploratory data analysis (EDA) to obtain a deeper understanding of the dataset, use Seaborn with Matplotlib.

[1] Zhang, L., et al. (2018). Feature Selection Techniques for Breast Cancer Classification using Machine Learning. Medical Image Analysis, 35, 376-389. [invalid URL removed]

[2] LeCun, Y., et al. (2015). Deep Learning. Nature, 521(7553), 436-444. https://doi.org/10.1038/nature14539

[3] Esteva, A., et al. (2017). Dermatologist-level classification of skin cancer with deep neural networks. Nature, 542(7639), 115-118. https://doi.org/10.1038/nature21056

[4] Cruz-Roa, A., et al. (2014). Automatic detection of invasive ductal carcinoma in whole slide images with Convolutional Neural Networks. Medical Image Analysis, 18(7), 1281-1291. [invalid URL removed]

[5] Rajkomar, A., et al. (2018). Scalable and accurate deep learning with electronic health records. npj Digital Medicine, 1(1), 1-10. https://doi.org/10.1038/s41746-018-0029-1