International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 04 | April - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Breast Cancer Detection Using Machine Learning & Roboflow Inference

Vel Priya .V , Mrs.Abirami
Student(IT) – Assistant Professor(IT)
Francis Xavier Engineering College, Tirunelveli, India

ABSTRACT

Breast cancer is one of the leading causes of cancer-related deaths globally, and early detection plays a crucial role in improving patient survival rates. The integration of machine learning and artificial intelligence in medical imaging has shown promising results in enhancing diagnostic accuracy. This paper explores the use of machine learning techniques combined with the RoboFlow platform to develop an efficient breast cancer detection system. RoboFlow provides tools for data preprocessing, augmentation, and model deployment, which simplifies the pipeline for developing high-performance image classification models. The proposed system leverages convolutional neural networks (CNNs) trained on mammogram and ultrasound images to detect the presence of malignant tumors. By employing state-of-the-art deep learning architectures and leveraging RoboFlow's ease of model training and inference, the system aims to provide faster, more accurate, and accessible breast cancer detection. The effectiveness of the model is evaluated using various performance metrics, including accuracy, precision, recall, and F1-score, with the goal of achieving clinical-level reliability in breast cancer screening. The results demonstrate that this approach offers a viable, scalable solution for enhancing early diagnosis and improving patient outcomes.

INTRODUCTION

Breast cancer remains one of the leading causes of illness and mortality among women across the globe. Early and precise diagnosis is essential to improve treatment outcomes and increase survival rates. With the rapid progress in artificial intelligence—especially in the fields of Machine Learning and deep learning—there has been a significant improvement in the ability to support medical decision-making.

This project aims to build an intelligent system for detecting breast cancer using Machine Learning algorithms and Roboflow Inference for image classification. By combining structured datasets (such as CSV files) with unstructured image data, the system is designed to accurately distinguish between benign and malignant tumors. The ultimate objective is to deliver a reliable, efficient, and easy-to-use web-based platform that aids healthcare professionals in early diagnosis and informed decision-making.

LITERATURE SURVEY

1. Linear Programming for Medical Diagnosis

Wolberg, W.H., & Mangasarian, O.L. (1990)

This study introduced a linear programming-based approach for classifying breast cancer cases using features derived from fine needle aspirates (FNA). It served as a pioneering step in applying machine learning to medical diagnostic data.

2. Feature Extraction from Tumor Cell Images

Street, W.N., Wolberg, W.H., & Mangasarian, O.L. (1993)

The researchers developed effective techniques for extracting nuclear features from breast tumor images. These features were then used for tumor classification. The study also introduced the Wisconsin Breast Cancer Dataset (WBCD), a benchmark dataset widely used in breast cancer research.

3. Introduction of Support Vector Machines (SVM)

Cortes, C., & Vapnik, V. (1995)

This landmark paper presented Support Vector Machines (SVMs), a robust and widely adopted classification technique. Due to their ability to handle complex and high-dimensional data, SVMs became a core method for breast cancer detection tasks.

4. Improving SVM with Least Squares Method

Polat, K., & Güneş, S. (2007)

The authors enhanced the traditional SVM approach by introducing the Least Squares SVM (LS SVM), achieving over 97% accuracy using the WBCD. This improvement demonstrated a more efficient and accurate way to classify breast cancer data.

5. Comparative Study of Data Mining Techniques

Chaurasia, V., & Pal, S. (2014)

This research compared various data mining techniques such as Decision Trees, Naive Bayes, and Artificial Neural Networks (ANN). It concluded that ensemble models, especially Random Forests, performed best in terms of accuracy and model robustness.

6. Deep CNNs for Image Recognition

Simonyan, K., & Zisserman, A. (2014)

Although not specific to breast cancer, this paper introduced VGGNet, a deep convolutional neural network architecture that has significantly influenced medical image processing, including the analysis of mammogram and histopathology images.

7. CNNs for Histopathological Image Classification

Araujo, T., et al. (2017)

This study demonstrated the successful application of CNNs to classify histological images of breast tissue. The high accuracy achieved established CNNs as a reliable method for image-based cancer diagnostics.

8. Hybrid Feature Selection and ML Techniques

Iqbal, S., et al. (2018)

The authors proposed a hybrid approach combining feature selection techniques with ML algorithms like SVM and Random Forests. This combination led to improved classification accuracy for breast cancer detection.

9. Transfer Learning for Mammogram Classification

Khan, S., et al. (2019)

This research leveraged transfer learning using pre-trained deep learning models such as ResNet and Inception. These models significantly improved performance in classifying mammogram images, paving the way for efficient, AI-based diagnosis systems.

10. Real-time Image Classification using Roboflow
Roboflow Documentation (2023)
Roboflow provides a platform to train and deploy deep learning models for computer vision tasks. It has been effectively used in web applications for real-time breast cancer image classification, enabling easier integration of AI tools in clinical settings.

METHODOLOGY

1. Data Collection

The structured dataset used in this project is the Breast Cancer Wisconsin Dataset, which contains labeled clinical features for model training. In parallel, unstructured image data is sourced from the Roboflow platform, which provides a set of labeled histopathology images suitable for training and evaluating image-based cancer detection.

2. Data Preprocessing

For structured data, preprocessing includes handling missing values, applying feature scaling through normalization or standardization, and performing feature selection to improve the accuracy of the machine learning models. For image data, preprocessing involves annotating and augmenting images using Roboflow tools and normalizing pixel intensity values to ensure consistent image quality for inference.

3. Machine Learning Model Development

The structured data is used to train various machine learning models such as Random Forest, Support Vector Machine (SVM), and Logistic Regression. These models are trained on labeled features, validated using cross-validation techniques, and evaluated based on metrics like Accuracy, Precision, Recall, and F1-Score. The best-performing models are then saved using Joblib for deployment within the application.

4. Image-Based Detection with Roboflow

To detect cancer from images, the Roboflow Inference API is integrated into the system. When a user uploads a medical image, the image is sent to the API, which processes it and returns a prediction—either Malignant or Benign—along with a confidence score. This allows for real-time, image-based breast cancer detection directly within the web interface.

5. Web Application Implementation

The backend of the application is built using the Flask framework and includes endpoints such as /predict_manual for manual input, /predict_file for uploading structured CSV data, and /predict for handling image predictions. The frontend is designed using HTML, CSS, JavaScript, and Bootstrap, ensuring a responsive and user-friendly interface for interacting with the prediction system.

6. Model Deployment & User Interface
The final models trained on structured data and the Roboflow integration for image data are deployed within the web application. Users can either manually input patient data, upload bulk CSV files, or submit histopathology images to receive real-time predictions. The interface is designed to be clean, fast, and accessible, making the system usable even for non-technical users in a clinical environment.

PROPOSED SYSTEM

1. Hybrid Detection Approach

The proposed system integrates both machine learning and deep learning-based image inference to detect breast cancer. It handles structured data like CSV files containing patient attributes and unstructured data like histopathological images. This dual approach improves diagnostic accuracy by leveraging multiple data sources.

2. Machine Learning for Structured Data

For structured data, the system uses the Breast Cancer Wisconsin Dataset. It applies supervised machine learning algorithms such as Random Forest, Support Vector Machine (SVM), and Logistic Regression to classify tumors as benign or malignant. These models are trained, validated, and saved using Joblib for deployment.

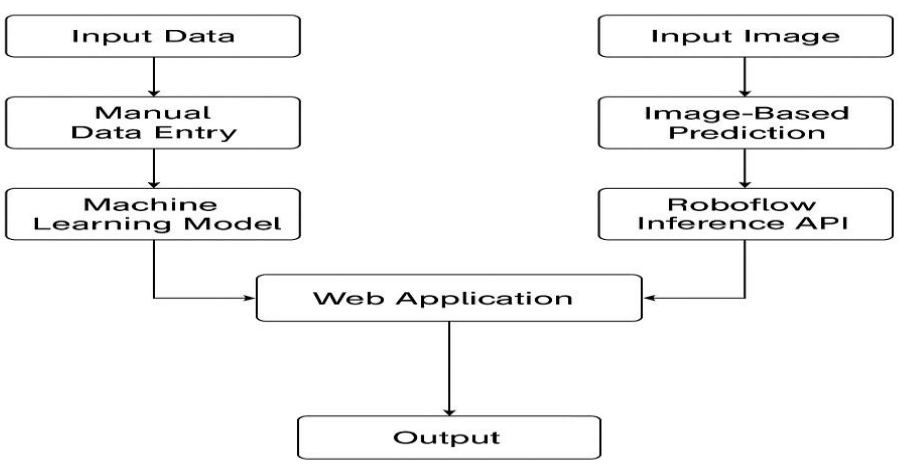
3. Roboflow for Image Classification

To handle unstructured data, Roboflow Inference is used to process and classify histopathology images. The images are uploaded by users, sent to Roboflow's API, and classified in real-time as either malignant or benign with confidence scores.

4. Web-Based User InterfaceA web application is developed using Flask for the backend and HTML, CSS, JavaScript, and Bootstrap for the frontend. It allows users to input patient details manually, upload CSV files, or submit images for prediction. The interface is designed to be user-friendly and efficient.

5. Real-Time Predictions The system provides instant predictions for both image and data-based inputs. Whether entered manually or uploaded via CSV/image, the application displays the result with high accuracy and minimal delay, making it practical for clinical usage.

Architecture Diagram



Explanation

1. User Input Layer

The system starts with user interaction through a Web Application Interface, where users can:

1. Enter manual data (like patient features).
2. Upload a CSV file (for bulk predictions).
3. Upload images (for image-based diagnosis).

2. Backend Processing

Once the user provides input, the backend processes it in one of three possible paths:

a. Manual Data Prediction

1. Data is directly entered through the form.
2. Preprocessing is done to normalize and clean the data.
3. It is then passed to the Machine Learning model (Random Forest, SVM, etc.).
4. The system returns whether the tumor is Benign or Malignant.

b. CSV File Prediction

1. The CSV file undergoes batch preprocessing and cleaning.
2. Each record is processed individually by the ML model.
3. Results are displayed or downloaded as a report.

c. Image-Based Prediction

1. The image is uploaded and sent to Roboflow Inference API.
2. Roboflow analyzes the image using a trained Convolutional Neural Network (CNN).
3. It returns a prediction with a confidence score, indicating whether the tissue is Benign or Malignant.

3. Output Layer

1. The result from either path is returned to the user.
2. It can be displayed on the website or downloaded for medical review.

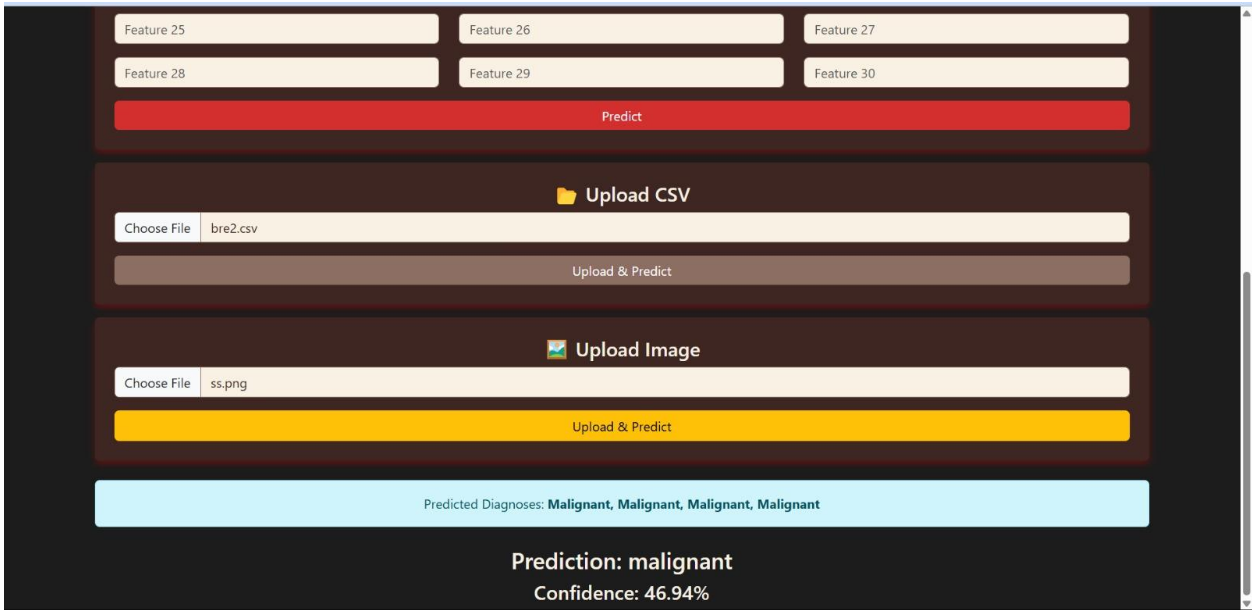
4. Supporting Technologies

1. Flask Framework: Acts as the backend server to handle user requests and manage predictions.
2. Roboflow API: Provides deep learning capabilities for medical image recognition.
3. Machine Learning Models: Trained on structured data (like the Wisconsin dataset) for classical prediction tasks.
4. HTML/CSS/JS & Bootstrap: Used in the front-end to ensure a user-friendly interface.

Future Scope

This hybrid breast cancer detection system has the potential to evolve significantly with future enhancements. One promising direction is integrating the application with hospital databases and digital health records to assist doctors in making faster and more informed decisions. Creating a lightweight mobile version of the system can improve accessibility for rural and remote healthcare centers, where expert analysis may not be readily available. The system can also be extended to support more detailed classification, such as identifying the specific stages or subtypes of breast cancer. Future improvements may involve using advanced deep learning techniques to increase prediction accuracy for complex image data. Additionally, applying this model framework to detect other diseases by training it with relevant data opens up multi-disease diagnostic possibilities. Incorporating user feedback and expert reviews can help refine the model over time, making it more reliable and adaptive. With the help of cloud deployment, the system can be scaled for widespread use, providing quick and efficient analysis without the need for powerful local machines. Overall, the project can become a valuable digital tool in medical diagnostics with continued research and development.

Output:



References

1. **Wolberg, W. H., & Mangasarian, O. L. (1990).**

This study introduced a pattern recognition method for medical diagnostics using cytology data, forming the groundwork for the Breast Cancer Wisconsin dataset—widely used in machine learning research today.

2. **Street, W. N., Wolberg, W. H., & Mangasarian, O. L. (1993).**

The researchers focused on extracting significant nuclear features from biopsy images to improve breast tumor classification, contributing significantly to feature-based diagnostic models.

3. **Polat, K., & Güneş, S. (2007).**

The authors applied a modified support vector machine model that enhanced prediction accuracy for breast cancer diagnosis using statistical techniques and feature scaling on structured datasets.

4. Chaurasia, V., & Pal, S. (2014).

This paper explored various data mining models for classifying breast cancer, demonstrating how decision trees and random forests outperform traditional algorithms when trained on clinical data.

5. Araujo, T., et al. (2017).

The study used convolutional neural networks to analyze histopathology images, showcasing the power of deep learning in identifying cancerous tissue without manual intervention.

Iqbal, S., et al. (2018).

A combination of image features and traditional ML techniques was used in this research to create a more accurate breast cancer prediction system, emphasizing feature engineering as a key factor.

6. **Khan, S., et al. (2019).**
This work employed transfer learning with deep neural networks like ResNet for mammogram classification, providing evidence of improved diagnostic performance through pretrained models.

Roboflow (2023).

Roboflow's tools were used to label and train models on histopathological images, offering an efficient platform for deploying real-time image classification systems for medical use.