MADAN MOHAN MALAVIYA UNIVERSITY OF TECHNOLOGY GORAKHPUR

Final Year Presentation on

Breast Cancer Detection Using Deep Learning

Presented By

Rohan Mishra (2020021116)
Sudhanshu Kumar (2020021158)
KM Poonam (202102009)

Under the Guidance of
Dr. P.K Singh

Department of Computer Science and Engineering
Madan Mohan Malaviya University of Technology,
Gorakhpur

Table of Contents

Introduction

- Breast cancer remains a significant global health challenge, affecting millions of individuals each year.
- Early detection is paramount for effective treatment outcomes in breast cancer cases.
- Histopathology, involving the microscopic examination of tissue samples, is essential for diagnosing breast cancer.
- Manual interpretation of histopathology images is time-consuming and susceptible to human error.
- There is a pressing need for innovative approaches to improve the accuracy and efficiency of breast cancer detection.
- This presentation focuses on exploring the potential of deep learning techniques, particularly Convolutional Neural Networks (CNNs), in revolutionizing breast cancer diagnosis from histopathology images.

Description

- This project aims to develop a deep learning model for automated classification of breast cancer histopathology images.
- The goal is to train a convolutional neural network (CNN) to accurately categorize small image patches from breast biopsies as either containing invasive ductal carcinoma (IDC) or being IDC negative.
- Automated analysis of histopathology slides is an active area of research, as it can potentially improve the efficiency and accuracy of breast cancer diagnosis.

Objective

- Develop a deep learning model, specifically a Convolutional Neural Network (CNN), for automated detection of breast cancer from histopathology images.
- Train the CNN model on a labeled dataset of histopathology images to learn discriminative features associated with cancerous and non-cancerous tissues.
- Evaluate the performance of the CNN model using various metrics such as accuracy, sensitivity, specificity, and area under the ROC curve.
- Compare the performance of the CNN model with existing methods or benchmarks to assess its effectiveness in breast cancer detection.

Methodology

- Data Collection and Preprocessing
- Model Selection and Architecture Design
- Training and Validation
- Evaluation and Testing
- Comparison and Validation
- Documentation and Reporting:

Working of Model

- Data Input: The model takes histopathology images of breast tissue samples as input.
- Feature Extraction: Convolutional layers extract features from the input images through convolutional operations, capturing patterns and structures at different spatial scales.
- Classification: Softmax activation at the output layer computes the probability distribution over the classes (cancerous or non-cancerous), indicating the likelihood of each class.
- Training: The model is trained using labeled histopathology images and corresponding ground truth labels (cancerous or non-cancerous).
- Validation: The trained model's performance is assessed on a separate validation set to monitor its generalization ability and prevent overfitting.

Result and Discussion

- **Accuracy (87.81%):** The accuracy metric indicates the overall correctness of the model's predictions. An accuracy of nearly 88% suggests that the model is proficient at correctly classifying both cancerous and non-cancerous histopathology images.
- **AUC Score (0.94):** The Area Under the ROC Curve (AUC) is a measure of the model's ability to distinguish between positive and negative cases. With an AUC score of 0.94, the model demonstrates strong discriminatory power, indicating high sensitivity and specificity in detecting breast cancer from histopathology images.
- **Number of Epochs (20):** The number of epochs represents the number of times the entire dataset is passed forward and backward through the neural network during training.

Result and Discussion

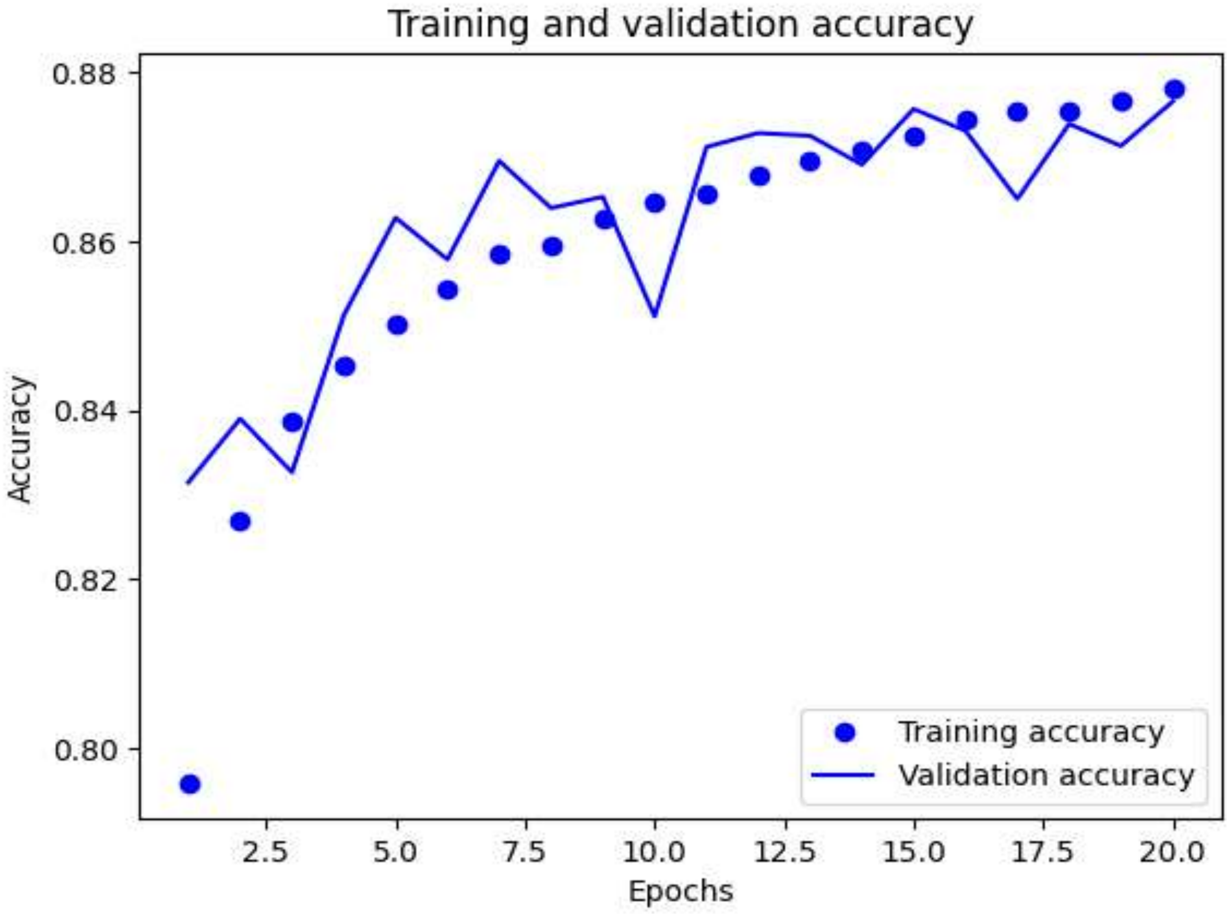
- **Validation Loss (0.30) and Validation Accuracy (0.8765):** The validation loss measures the difference between the predicted and actual labels during model training. A lower validation loss indicates better model performance. Meanwhile, the validation accuracy reflects the accuracy of the model on unseen data. The validation loss of 0.30 and validation accuracy of 0.8765 suggest that the model generalizes well to new data and is not overfitting.
- **IDC Cases (Positive: 8506, Negative: 7252):** The distribution of IDC cases (Invasive Ductal Carcinoma) indicates the prevalence of cancerous and non-cancerous tissue samples in the dataset.
- **F1 Score (0.81):**The F1 score is the harmonic mean of precision and recall, providing a balanced measure of the model's performance across both classes.

Result and Discussion

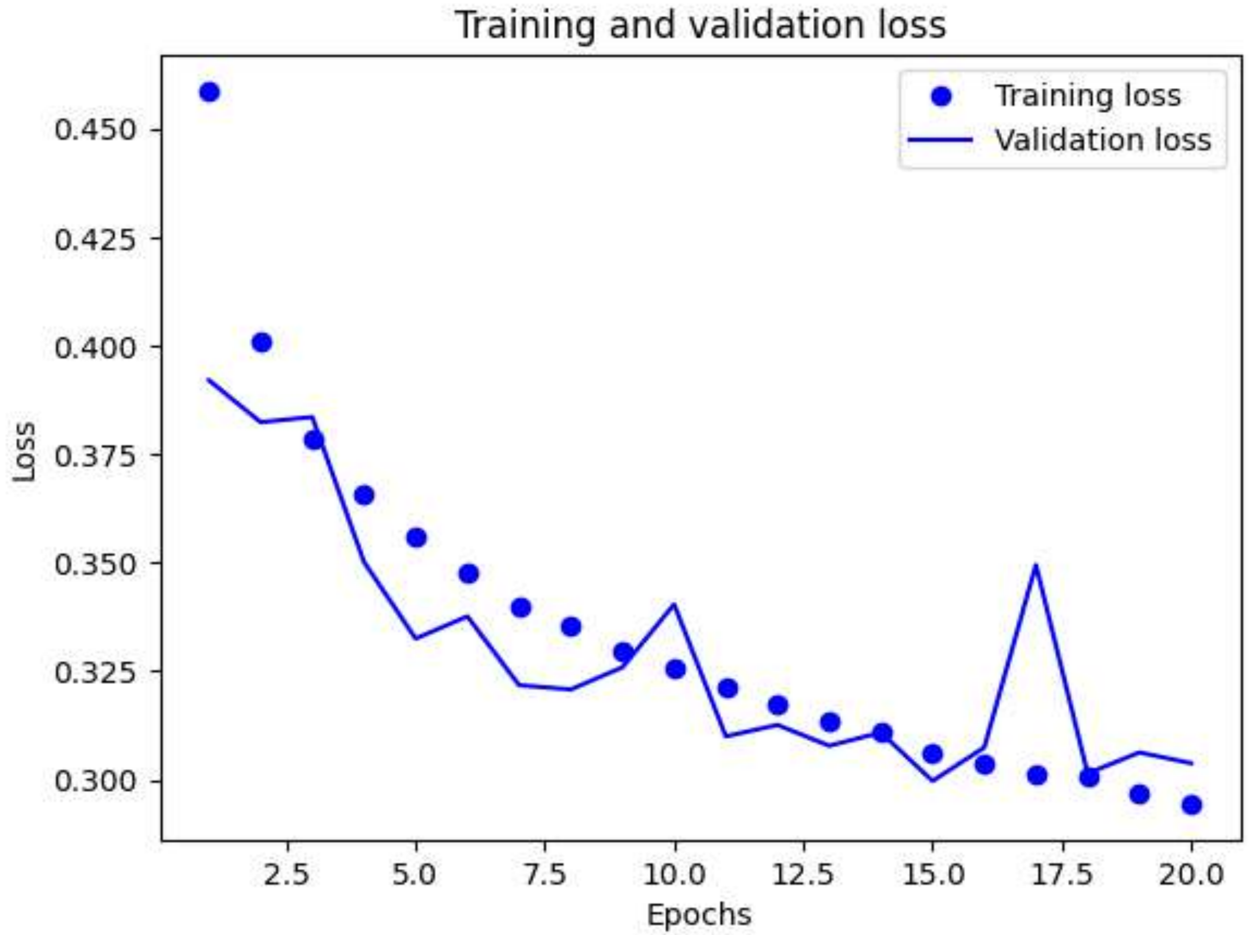
- With an F1 score of 0.81, the model achieves a good balance between precision and recall, indicating robust performance in detecting both cancerous and non-cancerous cases.

- **Precision (Positive: 0.85, Negative: 0.91):** Precision measures the proportion of true positive predictions among all positive predictions made by the model. A high precision indicates that the model makes fewer false positive predictions. In this case, precision values of 0.85 for positive (cancerous) cases and 0.91 for negative (non-cancerous) cases indicate high precision in both classes.

Overall Model Accuracy



Overall Model Loss



Challenges

- **Imbalanced Dataset:** The dataset may have an unequal distribution of cancerous and non-cancerous samples, leading to biased model predictions and reduced performance on minority classes.

- **Overfitting:** The model may become overly complex and learn to memorize the training data rather than generalize to unseen samples, resulting in poor performance on validation or test sets.

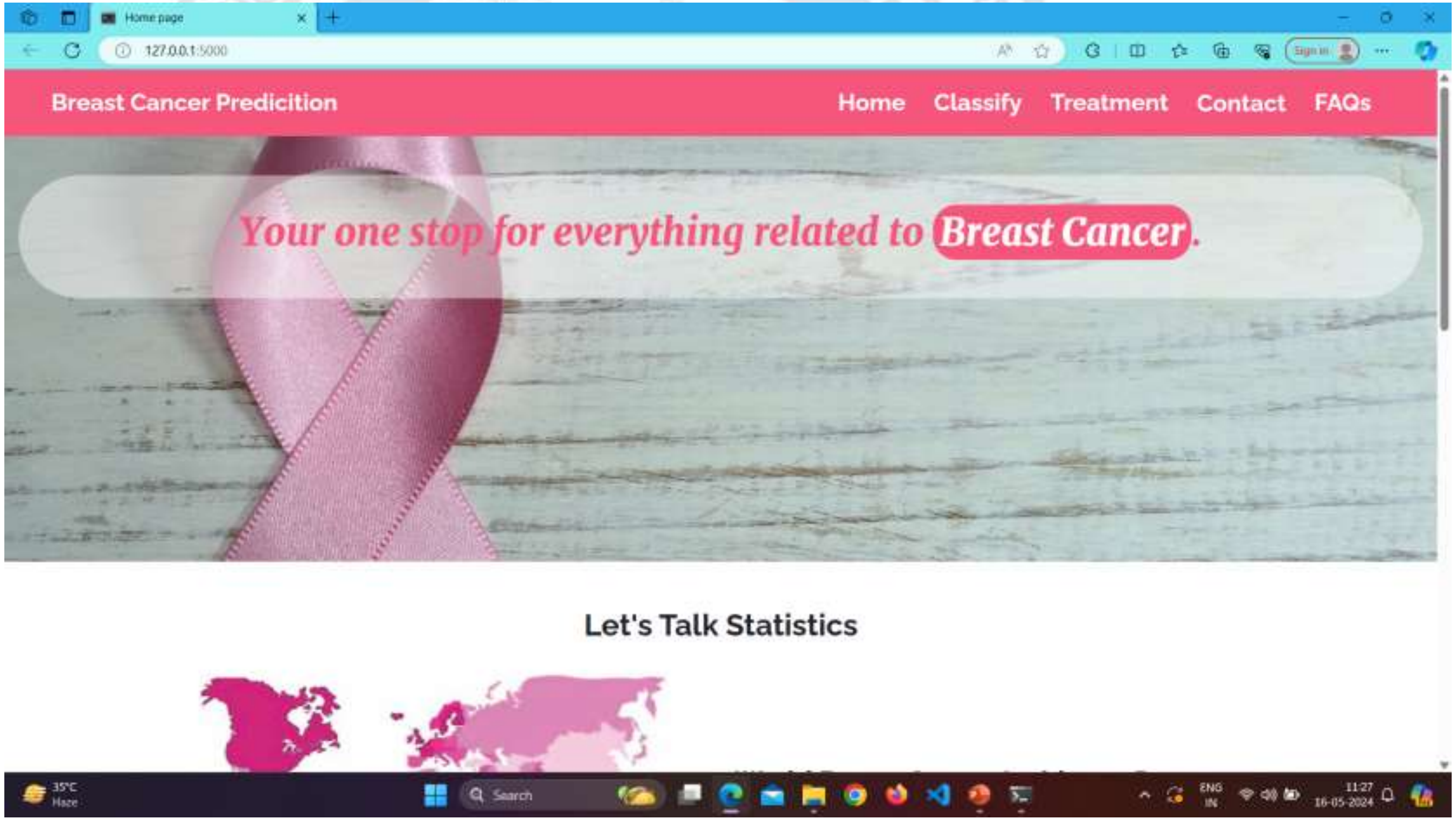
Solution

- **Balanced Sampling:** Employ techniques such as oversampling, undersampling, or class weights to balance the dataset and ensure equal representation of cancerous and non-cancerous samples, mitigating the impact of class imbalance.

- **Transfer Learning:** Leverage pre-trained deep learning models trained on large datasets (e.g., ImageNet) and fine-tune them on the histopathology dataset to transfer knowledge and improve the model's performance with limited labeled data.
- **Data Augmentation:** Augment the dataset with techniques such as rotation, flipping, zooming, or adding noise to increase the diversity of training samples and improve the model's robustness to variations in input data.

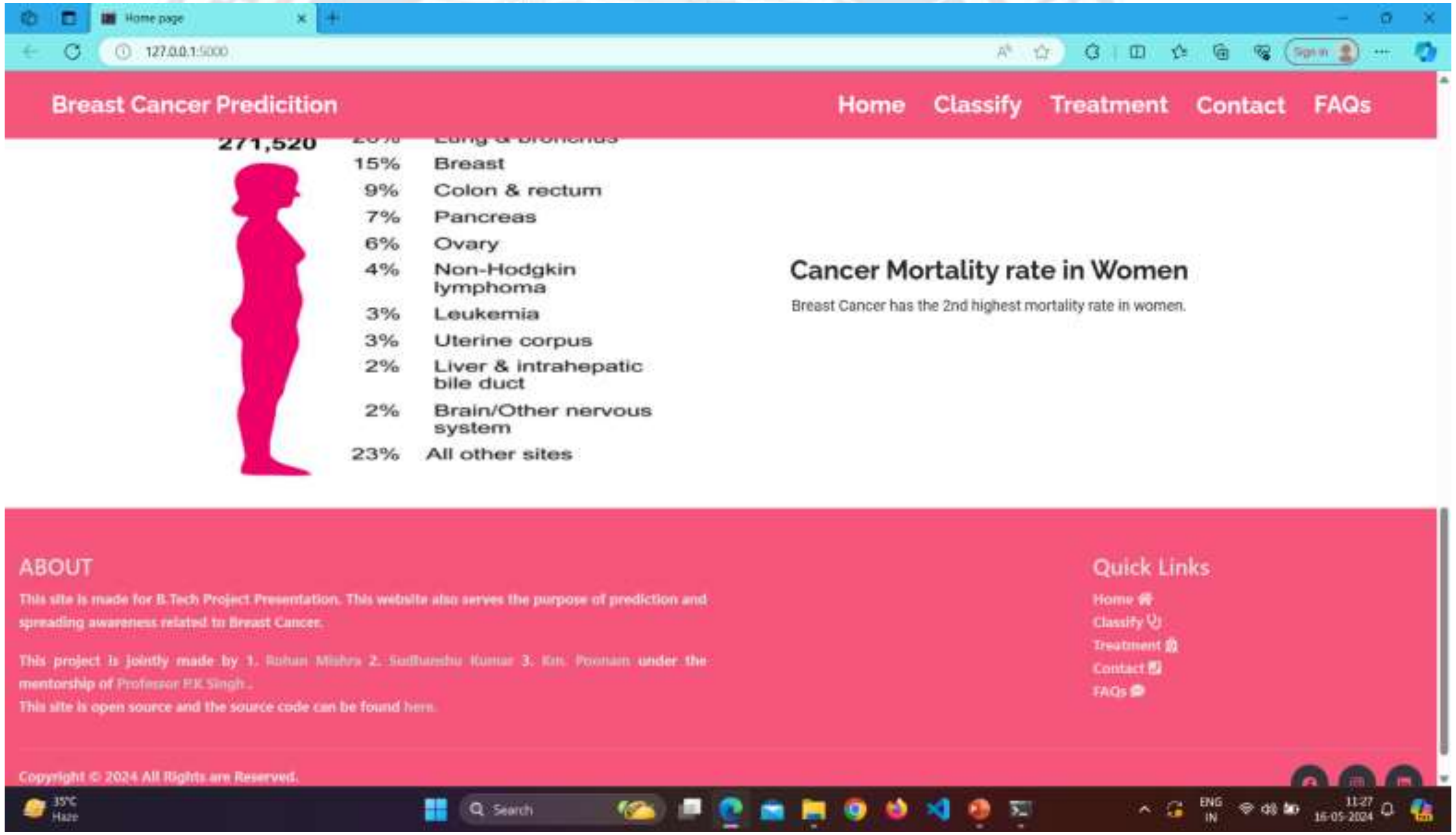
Website Design

- **Home Page :**



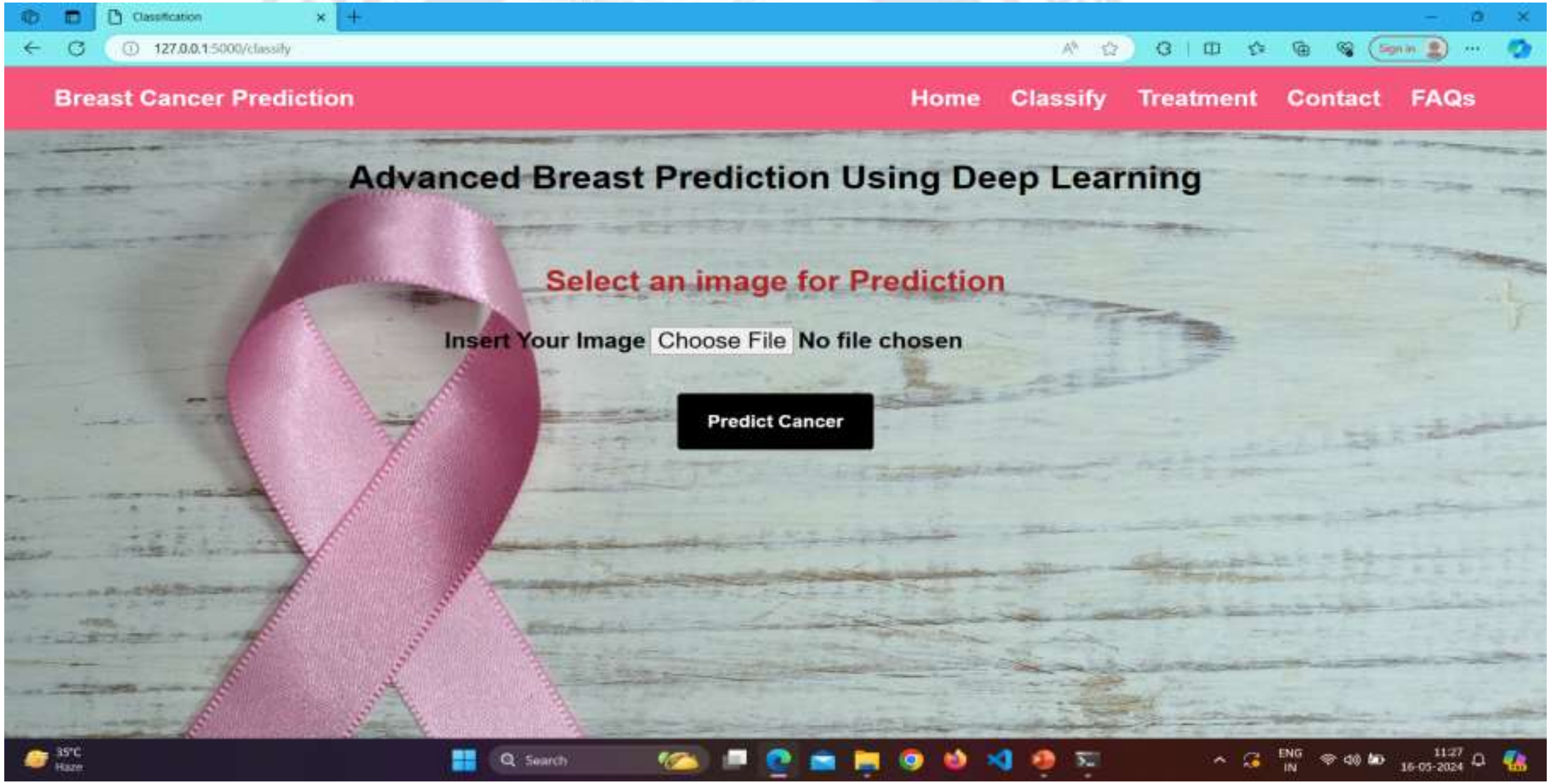
Website Design

- **Home Page :**



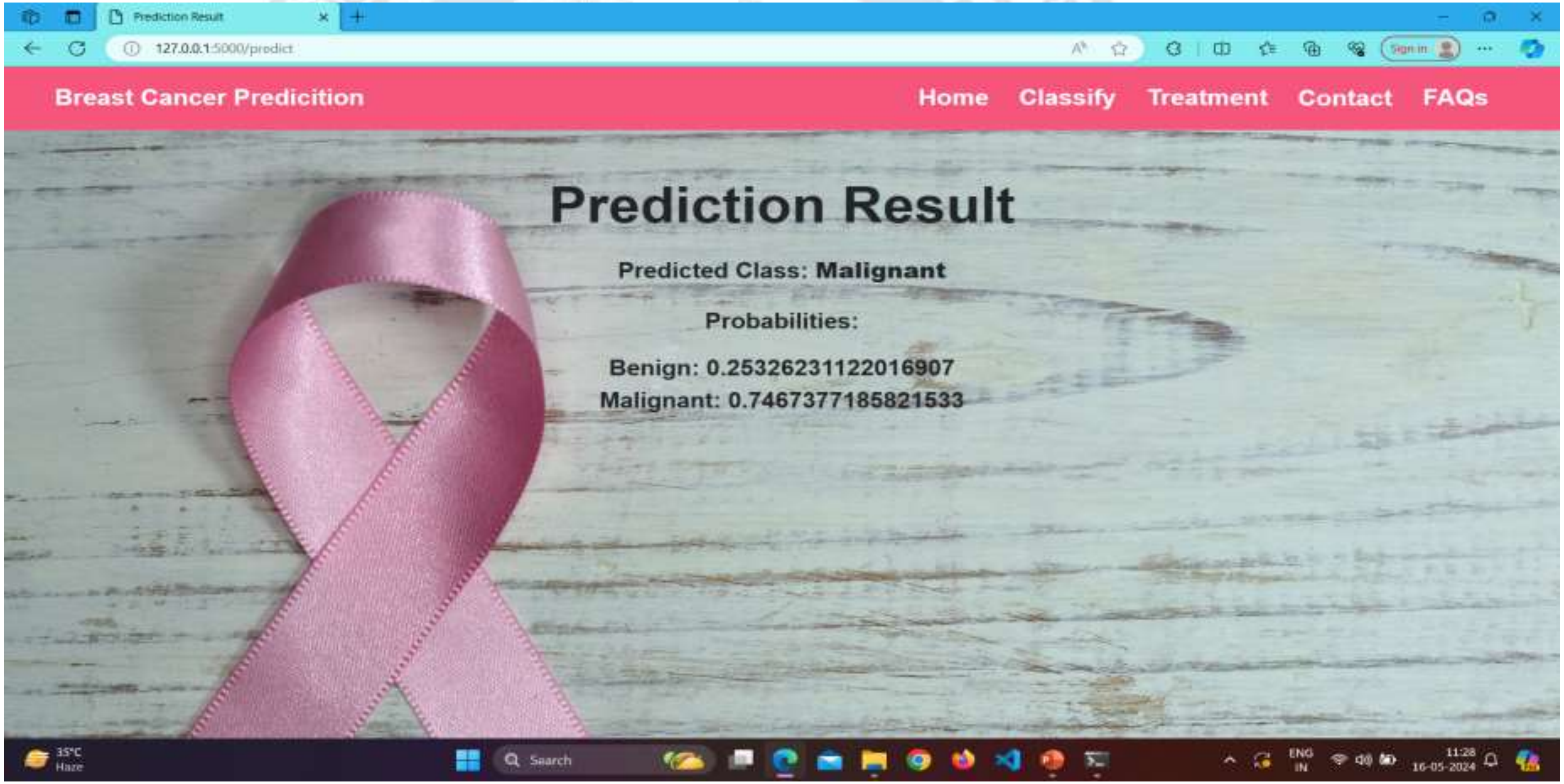
Website Design

- **Classify Page :**



Website Design

- **Result Page :**



Website Design

- **Treatment Page :**

Breast Cancer Prediction   Home   Classify   Treatment   Contact   FAQs

Treatment

Average Treatment Cost(All cost in USD)

PET Scan	Chemotherapy	Radiation Therapy	Surgery	Average Cost
$250	$300-$480	$2800	$2200-$3000	$4000

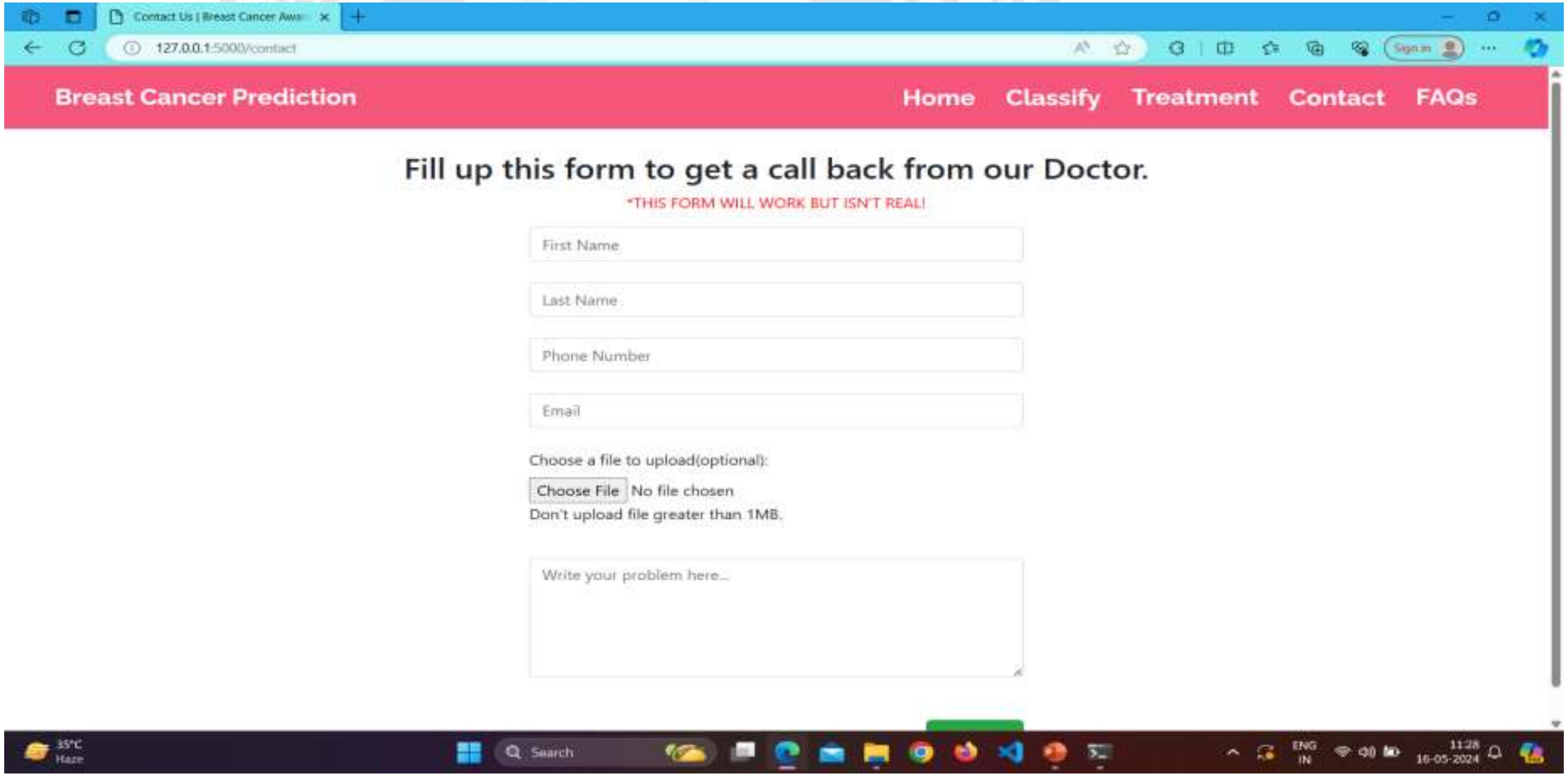
Accommodation Cost

Days in Hospital	Cost of Stay in Hotel	Cost of Service Apartments	Total duration of stay
3 Days	$40 – $50 per day	$35 per day	20 Days

Top Cancer Specialty Hospitals In India

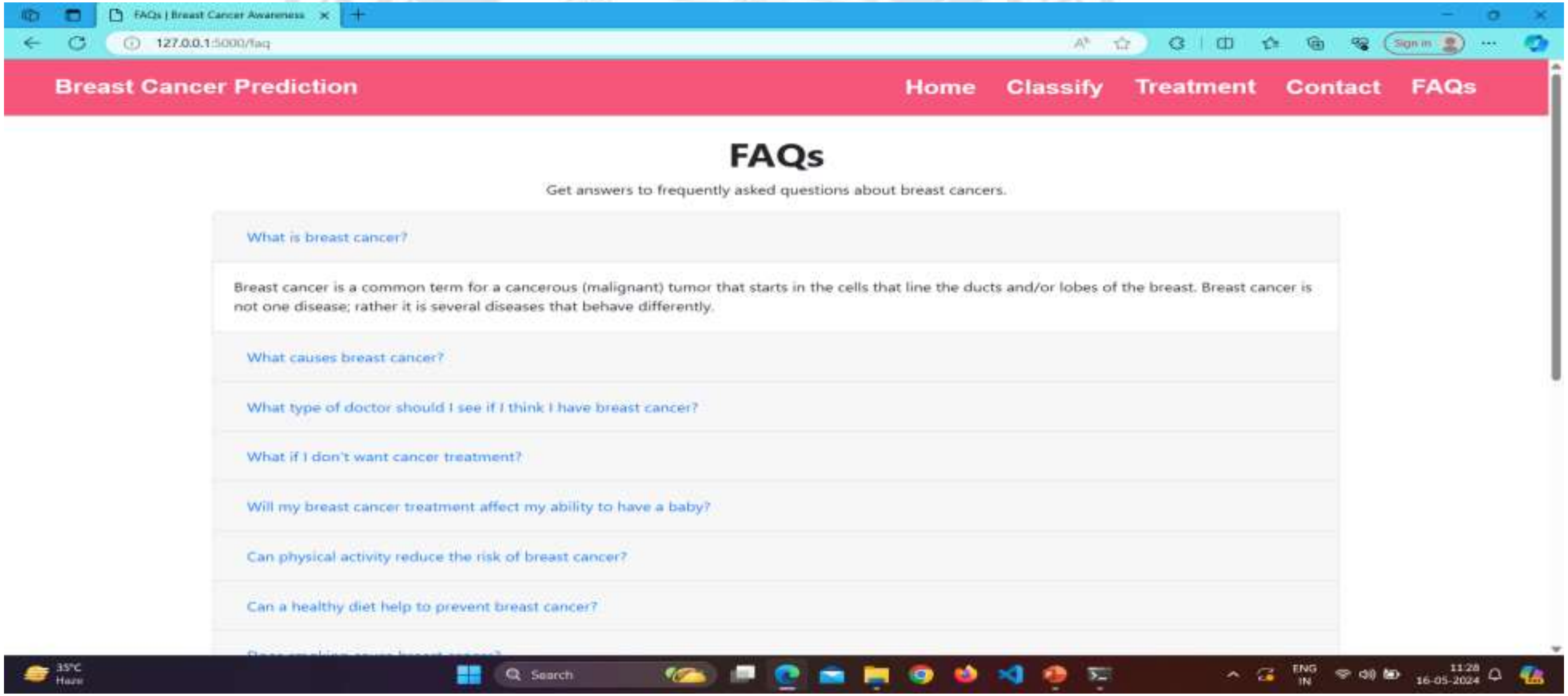
Website Design

- **Contact Page :**



Website Design

- **FAQs Page :**



Key Achievements

- **High Accuracy and AUC Score:** Achieving an accuracy of nearly 87.81% and an AUC score of 0.94 demonstrates the effectiveness of the deep learning model in accurately detecting breast cancer from histopathology images.
- **Generalization Performance:** The model exhibits strong generalization performance, as evidenced by the low validation loss of 0.30 and validation accuracy of 0.8765, indicating its ability to perform well on unseen data.
- **Clinical Relevance:** With precision values of 0.85 for cancerous cases and 0.91 for non-cancerous cases, the model demonstrates high precision in both classes, making it a valuable tool for aiding clinicians in diagnosing breast cancer with fewer false positives and negatives.

Key Achievements

- **F1 Score and Balanced Performance:** The F1 score of 0.81 reflects a balanced performance between precision and recall, indicating robust performance in detecting both cancerous and non-cancerous cases.

- **IDC Detection:** Successfully identifying 8506 cases of Invasive Ductal Carcinoma (IDC) and 7252 cases of non-IDC further underscores the model's ability to accurately detect cancerous tissue samples.

- **Clinical Impact:** The project has significant potential to impact clinical practice by providing clinicians with a reliable tool for diagnosing breast cancer from histopathology images, leading to earlier detection, better patient outcomes, and more informed treatment decisions.

Benefits

- Histopathology involves examining tissue samples under a microscope to identify cancerous cells. It provides a definitive diagnosis and detailed information about the tumor's characteristics, such as type, grade, and stage.
- Histopathology is highly accurate but is typically performed after a biopsy, which may require invasive procedures and time for processing and analysis.
- It is often used for confirming the presence of cancer and guiding treatment decisions, such as determining the need for surgery or chemotherapy.
- Histopathology is essential for research purposes and biomarker discovery. Analysis of tissue samples can identify novel biomarkers associated with cancer progression, drug response, and prognosis, leading to the development of targeted therapies and medicine .

Scope

- **Clinical Diagnosis Support:** The model can serve as a valuable tool to aid pathologists and clinicians in interpreting histopathology images and making accurate diagnoses of breast cancer.
- **Personalized Medicine:** The model can contribute to the development of personalized treatment plans by analyzing tumor characteristics and predicting patient prognosis.
- **Research and Development:** The model serves as a valuable tool for research purposes, facilitating the discovery of novel biomarkers, elucidation of disease mechanisms, and development of targeted therapies.
- **Resource Optimization:** By automating aspects of the diagnostic process, such as image analysis and feature extraction, the model enhances operational efficiency in healthcare settings

Limitations

- Requires tissue biopsy, which is invasive and may cause discomfort for the patient.
- Time-consuming process, as it involves tissue processing, staining, and microscopic examination by pathologists.
- **Interpretability:** Deep learning models are often considered black-box models, making it challenging to interpret their decisions.
- **Overfitting:** Deep learning models are prone to overfitting, particularly when trained on small datasets or complex architectures. Overfitting occurs when the model memorizes noise or irrelevant patterns in the training data, leading to poor generalization performance on unseen data.
- **Resource Requirements:** Training and deploying deep learning models require significant computational resources, including processing power, memory, and storage.

Conclusion

- The development of a deep learning model for breast cancer detection using histopathology images represents a significant advancement in the field of medical imaging and diagnostic technology.
- Despite these achievements, the model has limitations, including challenges related to data quality, interpretability, generalization, and ethical considerations. Addressing these limitations will be crucial for the model's successful deployment and adoption in clinical practice.
- Moving forward, further research and development efforts are needed to improve the model's performance, enhance interpretability, validate its clinical utility, and ensure compliance with regulatory and ethical standards.
- Ultimately, the deep learning model for breast cancer detection holds great promise for improving diagnostic accuracy, facilitating early detection, and ultimately saving lives.

Refrences

- https://data-flair.training/blogs/project-in-python-breast-cancer-classification/
- https://ieeexplore.ieee.org/document/7727519
- https://www.kaggle.com/datasets/paultimothymooney/breast-histopathology-images
- https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10217159/

THANK YOU