Introduction

Breast Cancer is the most commonly diagnosed cancer and second leading cause of mortality among women. In 2022, the World Health Organization reported that breast cancer resulted in the deaths of 670,000 people worldwide, with 2.3 million women diagnosed with the disease. It is the most common cancer globally, with 7.8 million women diagnosed in the past five years. Applying Machine Learning/Deep Learning techniques can help Detect breast cancer early.

Literature Review

Automatic MRI Breast tumor Detection using Discrete Wavelet Transform and Support Vector Machines:

In this work, a new method is done to detect breast cancer using the MRI images that is preprocessed using a 2D Median filter. The features are extracted from the images using discrete wavelet transform (DWT). These features are reduced to 13 features. Then, a support vector machine (SVM) is used to detect if there is a tumor or not. Simulation results have been accomplished using the MRI images datasets. These datasets are extracted from the standard Breast MRI database known as the "Reference Image Database to Evaluate Response (RIDER)". The proposed method has achieved an accuracy of 98.03 % using the available MRIs database. The processing time for all processes was recorded as 0.894 seconds. The obtained results have demonstrated the superiority of the proposed system over the available ones in the literature. [1]

Deep Learning Approach for Breast Cancer Diagnosis from Microscopy Biopsy Images:

The main contribution of this paper is proposing a deep learning approach to diagnose breast cancer from biopsy microscopy images. Several types of Deep Convolutional nets such as Vgg16, Alexnet, Inception, Resnet50, Resnet101, and Densenet169 are used. The three best models that achieved the highest accuracy without data preprocessing techniques are Densenet169, Resnet50, and Resnet101 with accuracy 62%, 68%, and 85% respectively. Additionally, the paper shows the impact of various data preprocessing techniques on the performance of the best models. The experimental results indicate that data augmentation and segmentation increase the accuracy of the best models by 20%, 17%, and 6% respectively. To further boost the accuracy of the models, the paper aggregates the best models using an ensemble learning technique. The results reveal that the best-achieved accuracy is 92.5%.[2]

A Systematic Literature Review of Breast Cancer Diagnosis Using Machine Intelligence Techniques:

In this paper, different machine intelligence techniques [machine learning (ML), and deep learning (DL)] were analyzed in the context of breast cancer. In addition, the classification of breast cancer into malignant and benign using different breast cancer image modalities were discussed. Furthermore, the diagnosis of breast cancer using various publicly and privately available image datasets, pre-processing techniques, feature extraction techniques, comparison between conventional ML and different convolutional neural network (CNN) architectures, and transfer learning techniques were discussed in detail. It also correlates the parameters and attributes impact in case of different methods applied. Advantages and the limitations of the machine intelligence approaches were highlighted based on the discussion and analysis. [3]

Ethical Considerations
The use of AI and ML in healthcare also raises ethical considerations, particularly around patient privacy, data security, and the potential for bias in algorithms. It is essential to develop and implement models that are transparent, explainable, and fair, ensuring that they do not inadvertently exacerbate existing health disparities.

Conclusion
The main aim of this project is to train machine learning models on breast cancer structured metadata, and compare the result to deep learning models using CNN and transfer learning trained using image data . and try to improve the accuracy and performance of the models compared to past research.

Data Review

Data Sources:
The primary data for the project involves mammography images, which are essential for training and validating the CNN models. Mammography data is typically obtained from public datasets like the Digital Database for Screening Mammography (DDSM) or private clinical sources. These datasets contain images labeled as benign or malignant, often accompanied by additional metadata such as patient age, tumor size, and biopsy results.

Data Preprocessing:
Data preprocessing is a crucial step in preparing mammography images for model training. This involves resizing images to a consistent dimension, normalizing pixel values, and augmenting the data to artificially increase the size of the dataset. Techniques like rotation, flipping, and zooming can help create a more robust model by simulating real-world variations in mammography images.

Given the size and complexity of mammography data, handling memory constraints is vital. It may be necessary to downsample images or use techniques like mini-batch training to manage computational resources effectively.

Data Challenges:
One of the key challenges in working with mammography data is the class imbalance, where the number of benign cases typically far exceeds the number of malignant ones. This imbalance can lead to a model that is biased towards predicting the majority class, reducing its

effectiveness in identifying malignant cases. Techniques like oversampling, undersampling, and the use of specialized loss functions can be employed to address this issue.

Technology Review

Convolutional Neural Networks (CNNs):
CNNs are the cornerstone of the project, providing the architecture for analyzing and classifying mammography images. CNNs are particularly well-suited for image analysis due to their ability to capture spatial hierarchies in data through the use of convolutional layers, pooling layers, and fully connected layers. Transfer learning, where a pre-trained CNN model (e.g., VGG16, ResNet) is fine-tuned on mammography data, can also be employed to improve performance, especially when working with limited datasets.

TensorFlow and Keras:
TensorFlow and Keras are the primary libraries used for implementing CNNs. TensorFlow is a comprehensive open-source platform for ML, providing tools for building and deploying ML models. Keras, which is now integrated into TensorFlow, offers a high-level API for quickly prototyping and building deep learning models. Together, these tools allow for efficient model training, evaluation, and deployment.

Django and REST Framework:
Django, coupled with Django REST Framework (DRF), provides the backend infrastructure for the project. Django is a high-level Python web framework that encourages rapid development and clean, pragmatic design. DRF extends Django's capabilities, making it easier to build and manage RESTful APIs, which are essential for serving predictions from trained models to a web or mobile application.

React.js:
React.js is used for the frontend of the application, providing a responsive and dynamic user interface. React is a popular JavaScript library for building user interfaces, particularly single-page applications, where data needs to be dynamically updated without reloading the page. Its component-based architecture aligns well with the modular approach of modern web development.

Deployment Considerations
Deploying ML models in a clinical setting requires careful consideration of scalability, reliability, and security. Containerization technologies like Docker can be used to package the application and its dependencies, ensuring consistent performance across different environments. Additionally, cloud platforms like AWS or Google Cloud can provide scalable infrastructure for hosting the application and managing large datasets.

References

1- A. M. Ibraheem, K. H. Rahouma and H. F. A. Hamed, "Automatic MRI Breast tumor Detection using Discrete Wavelet Transform and Support Vector Machines," 2019 Novel Intelligent and Leading Emerging Sciences Conference (NILES), Giza, Egypt, 2019, pp. 88-91, doi: 10.1109/NILES.2019.8909345. keywords: {Feature extraction;Magnetic resonance imaging;Support vector machines;Breast cancer;Breast tumors;Discrete wavelet transforms;Breast cancer;CAD system;Images detection;MRI images;Machine Learning.}

2 - S. N. Eldin, J. K. Hamdy, G. T. Adnan, M. Hossam, N. Elmasry and A. Mohammed, "Deep Learning Approach for Breast Cancer Diagnosis from Microscopy Biopsy Images," 2021 International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC), Cairo, Egypt, 2021, pp. 216-222, doi: 10.1109/MIUCC52538.2021.9447653. keywords: {Deep learning;Ultrasonic imaging;Biological system modeling;Microscopy;Computational modeling;Biopsy;Data preprocessing}.

3 - Nemade, V., Pathak, S. & Dubey, A.K. A Systematic Literature Review of Breast Cancer Diagnosis Using Machine Intelligence Techniques. *Arch Comput Methods Eng* 29, 4401–4430 (2022). https://doi.org/10.1007/s11831-022-09738-3