ABSTRACT:

Breast cancer is a leading cause of mortality among women worldwide, emphasizing the need for effective early detection techniques. Mammography is widely regarded as the standard imaging modality for breast cancer screening. However, the manual interpretation of mammograms is prone to subjectivity and may result in misdiagnoses. Machine learning algorithms have emerged as powerful tools to improve diagnostic accuracy and reduce human error in medical imaging. This study focuses on the application of machine learning techniques for breast cancer detection from mammograms. By leveraging advanced image processing and feature extraction methods, combined with supervised and unsupervised learning algorithms, the system can automatically identify abnormalities indicative of malignancies. A comprehensive dataset of mammogram images is used to train and validate the models, ensuring robustness and reliability. The results demonstrate the potential of machine learning algorithms in achieving high sensitivity and specificity, outperforming traditional diagnostic methods. This approach highlights the transformative role of artificial intelligence in enhancing early breast cancer detection, ultimately contributing to improved patient outcomes

CHAPTER 1

INTRODUCTION

1.1 Introduction

Breast cancer remains one of the leading causes of morbidity and mortality among women globally, with early detection being critical for improving survival rates. Mammography is widely regarded as one of the most effective imaging techniques for the early detection of breast cancer, enabling the identification of Tumors and abnormalities that may not yet be clinically visible. However, interpreting mammograms requires expert knowledge, and even experienced radiologists can sometimes miss subtle signs of cancer, leading to false negatives or false positives. This challenge highlights the need for advanced computational methods to assist in the detection process.

This project focuses on using machine learning algorithms for the detection and classification of breast cancer in mammograms. The goal is to improve diagnostic accuracy, reduce human error, and enhance early detection, thus contributing to better patient outcomes.

1.2 Motivation of work

A project on breast cancer detection using Machine learning's can serve as a crucial contribution to healthcare by harnessing the power of machine learning and image processing to aid early diagnosis. Breast cancer is one of the most prevalent cancers among women, and early detection significantly increases the chances of successful treatment. By working on this project, you not only deepen your technical skills but also contribute towards developing innovative solutions that can save lives by enabling faster and more accurate detection methods. Several problems occur in segmentation and classification of MRI images. The segmentation is useful in partitioning the image in meaningful regions. The misclassification or wrong segmentation leads to several problems in diagnosis of Tumour. The accurate classification of the breast Tumour is necessary to save the patient's life. The exact treatment can be given on the basis of segmentation and classification of Tumour. The existing systems provided significant improvement in Tumour detection but the accuracy level is low and high level of noise was presented. These issues motivate us to provide a segmentation and classification algorithm which will improve the accuracy. Here, the breast Tumour is detected by the use of medical imaging techniques. The main motivation for an engineer would be to propose a work to develop the system of breast Tumour segmentation and detection which would provide better performance parameters. Identifying different Tumour classes or subclasses with similar morphological appearances present is an interesting problem and has an important implication in Tumour diagnosis and treatment. Present technique includes 'biopsy' procedure which is operative in manner. Classification based on the imaging techniques is not acceptable by the radiologist and oncologist due to required accuracy. This motivates the engineers to develop an easy, quick and reliable algorithms to be implemented in the device which acts as a substitute for biopsy and produce accurate results.

1.3 Literature survey

Study/Article	Year	Authors
Breast Cancer Diagnosis Using Adaptive Voting Ensemble Machine Learning Algorithm	2018	Naresh Khuriwal , Nidhi Mishra
Performance Comparison Of Different Machine Learning For Early Prediction Of Breast Cancer Using Wisconsin Breast Cancer Dataset.	2022	Atajan Rovshenov , Serhat Peker
Using Supervised Learning For Breast Cancer Detection Using AI&ML	2023	: PratyakshSingh,jaideep Nagill,dr. Kavita Saini
Breast Cancer Detection Using Nanoparticle Sensor With Machine Learning Algorithms	2024	J.N.V.R. Swarup Kumar,charishma Karri, Naga Srihitha Vatsavayi,harshitha Lekkala, Saaketh Choudarapu
Breast Cancer Prediction System Utilizing Machine Learning Algorithms	2024	Chirayou Bista, Asreetha M ,Salahuddin Slimanzay ,Md Solaiman Sheikh , Dr . P Srinivasa Rao
Exploring Machine Learning Techniques For Enhanced Breast Cancer Detection	2024	Nagesh Sharma, Sandeep Singh Kang
An Analysis Of Ensemble Machine Learning Algorithms For Breast Cancer Detection: Performance And Generalization	2024	Rakesh Kumar ,Meeta Chaudhry ,H. K. Patel ,Navin Prakash ,Abhinav Dogra,sunil Kumar

Fig 1.3.1

1.4 Objectives of work

- **Early and Accurate Detection**
  Develop machine learning models capable of identifying breast cancer at its earliest stages with high sensitivity and specificity, reducing false negatives and false positives.
- **Enhancement of Diagnostic Accuracy**
  Use machine learning to assist radiologists in interpreting mammograms, minimizing subjectivity and improving overall diagnostic precision.
- **Feature Extraction and Analysis**
  Automate the process of extracting and analyzing relevant features (e.g., texture, shape, and density) from mammogram images for better classification of malignant and benign tissues.

- **Integration of Diverse Data**

  Combine imaging data with patient-specific information (e.g., age, genetic predisposition, and medical history) to create personalized risk prediction models.
- **Scalable and Cost-Effective Solutions**

  Design machine learning-based systems that are scalable and can be implemented in low-resource settings, making advanced diagnostic tools accessible worldwide.
- **Real-Time Application**

  Enable real-time processing of mammograms to provide immediate diagnostic support and reduce the workload on healthcare professionals.
- **Model Validation and Robustness**

  Validate the machine learning models on diverse and large datasets to ensure their generalizability and robustness across different populations.
- **Reducing Human Error**

  Use automation to minimize the variability in human interpretation and enhance the reliability of breast cancer detection.
- **Support for Decision-Making**

  Provide actionable insights and decision-support tools for oncologists and radiologists, aiding in effective treatment planning and monitoring.
- **Contribution to Research**

  Advance the field of medical imaging and machine learning by contributing novel methodologies and algorithms tailored for breast cancer detection.

CHAPTER 2

LITERATURE SURVEY

2.1 CHALLENGES AND PROBLEMS

1) “Performance Comparison of Different Machine Learning Techniques For Early Prediction of Breast Cancer using Wisconsin Breast Cancer Dataset” by Atajan Rovshenov and Serhat Peker (IEEE-2022)

A significant health issue, cancer is becoming more prevalent globally and is a leading cause of mortality. Recent studies have shown that breast cancer is one of the most prevalent cancer type, particularly among women. Early detection can increase the chances of survival for those with breast cancer and lower treatment cost. However, there are drawbacks to the early diagnosis methods utilized in today's healthcare systems. These include the need for substantial human resources, long-term effects, and difficult access to these services for everybody. For early breast cancer diagnosis, technologies that are simple to use, yield reliable findings
compared to scientific methodologies, and are available to everyone are required. Artificial Intelligence techniques enable the early diagnosis of breast cancer. This study aims to classify benign and malignant breast cancer image features. Artificial Neural Network, Support Vector Machine and Random Forest algorithms were used to classify features obtained from images. Experiments were performed on the Wisconsin Breast Cancer dataset. Experimental evaluation shows that 99% of the most successful results were achieved with the Artificial Neural Network algorithm. According to experimental findings, the classification technique can identify breast cancer in its early stages. The findings of the study are expected to shed on light new researches for investigation into breast cancer early detection.

2)“Breast Cancer Diagnosis Using Adaptive Voting Ensemble Machine Learning Algorithm” by Naresh Khuriwal and Nidhi Mishra(IEEE-2018)

According to Breast Cancer Institute (BCI), Breast Cancer is one of the most dangerous type of diseases that is very effective for women in the world. As per clinical expert detecting this cancer in its first stage helps in saving lives. As per
cancer.net offers individualized guides for more than 120 types of cancer and related hereditary syndromes. For detecting breast cancer mostly machine learning techniques are used. In this paper we proposed adaptive ensemble voting method for diagnosed breast cancer using Wisconsin Breast Cancer

database. The aim of this work is to compare and explain how ANN and logistic algorithm provide better solution when its work with ensemble machine learning algorithms for diagnosing breast cancer even the variables are reduced. In this paper we used the Wisconsin Diagnosis Breast Cancer dataset. When compared to related work from the literature. It is shown that the ANN approach with logistic algorithm is achieved 98.50% accuracy from another machine learning algorithm.

3)"Breast Detection Using Nanoparticle Sensor with Machine Learning Algorithms" by J.N.V.R. Swarup Kumar, Charishma Karri, Naga Srihitha Vatsavayi, Harshitha Lekkala and Saaketh Choudarapu. (IEEE-2024)

This research is mainly focused on the early detection of breast cancer in women by their urine samples with nanoparticle sensors. It detects certain enzymes and proteins that can be the main cause of cancer by machine learning algorithms, correlation analysis, and logical regression methods and created a web-based breast cancer prediction website using started nanoparticle urine analysis data, which contains the DNA barcode sequences that match with DNA of patient urine which is user friendly to use for every person. This study mainly focuses on non-invasive cancer. We analyse non-invasive cancer by using urine sample test data collected from the patients, utilizing advanced technology to access DNA signatures associated with breast cancer biomarkers. Our approach involves the barcode of nanoparticle sensors, which matches the urine samples. After the urine samples match, they are applied to the sensors, translated into digital data, and transmitted to a centralized system. Now, the centralized system collaborates with datasets derived from previous breast cancer cases. The algorithm now analyses the urine data and identifies the patterns. Then, it correlates with different stages of breast cancer.

4) "Using Supervised Learning for Breast Cancer Detection using AI&ML" by Pratyaksh Singh, Jaideep Nagill and Dr. Kavita Saini (IEEE -2023)

Breast cancer is the leading cause of cancer related deaths among women worldwide. A significant amount of research has been conducted to improve early detection of breast cancer, which is crucial for effective treatment and increased chances of survival. While mammograms have been the most reliable detection method, there is a need to explore alternatives that are cost-effective, safe, and accurate across different datasets. A hybrid paradigm of machine learning methods is presented in this work, proposed for effective breast cancer detection. The model combines several machine learning algorithms, including ANN, SVM, KNN and Decision Tree (DT), and can be applied to various data types, including images and blood tests. The proposed model aims to provide accurate results that are close to perfection.

5)“Breast Cancer Prediction System Utilizing Machine Learning Algorithms” By Chirayou Bista, Asreetha M, Salahuddin Slimanzay, Md Solaiman Sheikh and Dr. P Srinivasa Rao (IEEE-2024)

Breast cancer remains a significant global health concern, necessitating advanced predictive tools for early detection and effective prognosis. In response to this challenge, this research introduces a Breast Cancer Prediction System utilizing state-of-the-art machine learning algorithms. The system leverages the power of Random Forest, Support Vector Machine (SVM), and Gradient Boosting Ensemble to enhance accuracy and reliability. The Random Forest algorithm efficiently captures complex relationships within the breast cancer dataset, creating an ensemble of decision trees for robust predictions. Meanwhile, the Support Vector Machine optimally classifies data points by identifying hyperplanes, thereby enhancing the model's ability to discriminate between benign and malignant cases. Furthermore, the Gradient Boosting Ensemble technique synthesizes weak predictive models into a strong learner, boosting the overall predictive performance of the system. The Breast Cancer Prediction System offers a thorough and precise evaluation of the likelihood of breast cancer by combining the advantages of all three algorithms. By utilizing the discriminative power of the SVM, the boosting mechanism of the Gradient Boosting Ensemble, and the Random Forest's capacity to handle complex datasets, the system attains increased accuracy in early detection and prognosis. The suggested system has great promise for useful application in clinical settings, resulting in prompt interventions and better patient outcomes.

6) “Exploring Machine Learning Techniques for Enhanced Breast Cancer Detection” by Nagesh Sharma and Sandeep Singh Kang (IEEE-2024)

Breast cancer is the second-greatest cause of death for women worldwide, affecting the majority of them. On the other hand, if cancer is identified early and adequately treated, it may be cured. Patients' chances of survival and prognosis can be greatly improved by early identification of breast cancer and prompt treatment intervention. Additionally, accurate benign tumor classification might assist patients in avoiding unnecessary therapy. This study provides a thorough overview of several studies that examined how ML algorithms may be used to find breast cancer. The main goal is to evaluate these algorithms' performance in terms of precision, accuracy, recall, and overall effectiveness. This evaluation tries to identify the most promising approaches and indicate areas for further development by looking at a wide variety of algorithms

7) “An Analysis of Ensemble Machine Learning Algorithms for Breast Cancer Detection: Performance and Generalization” by Rakesh Kumar, Meeta Chaudhry, H. K. Patel, Navin Prakash, Abhinav Dogra and Sunil Kumar(IEEE-2024)

Breast cancer is an explorative area, now a days, it is common dieses in females. So, the diagnosis of breast cancer for a patient at early stage can help to prevent their lives. The prediction of breast cancer by using machine learning can be done after applying the various machine learning algorithms. Here, we applied ensemble machine learning algorithms and got better results in terms of performance and generalization. In this paper, the comparative analysis of Light Gradient Boosting (Light GBM) and Gradient Boosting have been done and experiments done by using labelled dataset of breast cancer. The Light GBM algorithm has been found to be less accurate in comparison to Extreme Gradient Boosting (XG Boost) algorithm.

2.2 MOTIVATION

- **Growing Prevalence of Breast Cancer**
  Breast cancer remains one of the most common and deadly cancers worldwide. Literature reveals an alarming rise in incidence, with early detection being critical for improved survival rates. This highlights the need for innovative approaches like machine learning to enhance diagnostic capabilities.
- **Limitations of Conventional Methods**
  Traditional mammogram interpretation is often prone to human error, including false positives and negatives, leading to delayed diagnosis or unnecessary interventions. Literature shows that machine learning algorithms can significantly improve accuracy, efficiency, and reliability in breast cancer detection.
- **Success of Machine Learning in Medical Imaging**
  A review of studies demonstrates the successful application of machine learning in various medical imaging domains, including the detection of tumors, classification of malignancies, and prediction of patient outcomes. This underscores its potential for transformative impact in breast cancer detection.
- **Personalized Healthcare Opportunities**
  Research highlights the ability of machine learning to integrate diverse data sources, such as imaging, genetics, and patient history, enabling personalized risk assessments and tailored treatment plans, a critical need in modern oncology.

- **Gap in Current Practices**

  The literature reveals gaps in the widespread adoption of AI in low-resource settings due to the high costs of traditional diagnostic tools and limited availability of trained radiologists. Developing cost-effective, scalable machine learning solutions can address this disparity.
- **Encouraging Results from Existing Studies**

  Recent studies have demonstrated machine learning models achieving high sensitivity and specificity in detecting breast cancer, sometimes surpassing human performance. This progress motivates further exploration and refinement of these models.
- **Potential to Save Lives**

  Early detection of breast cancer is directly linked to improved survival rates. By addressing the challenges highlighted in the literature, machine learning solutions can play a pivotal role in saving lives through timely and accurate diagnoses.

- **Inter disciplinary Collaboration**

  The project fosters collaboration between computer science, medicine, and public health, as indicated in multiple studies. This interdisciplinary nature makes the work both challenging and rewarding.
- **Global Impact**

  The scalability of machine learning tools, as highlighted in the literature, makes it possible to address global health challenges, particularly in underserved regions where access to healthcare is limited.
- **Contributions to the Field**

  Literature surveys reveal that while significant progress has been made, there is ample room for innovation in model development, feature extraction, and integration of new data modalities, motivating further contributions to the field.

CHAPTER 3

DESIGN AND IMPLEMENTION

3.1 BLOCK DIAGRAM

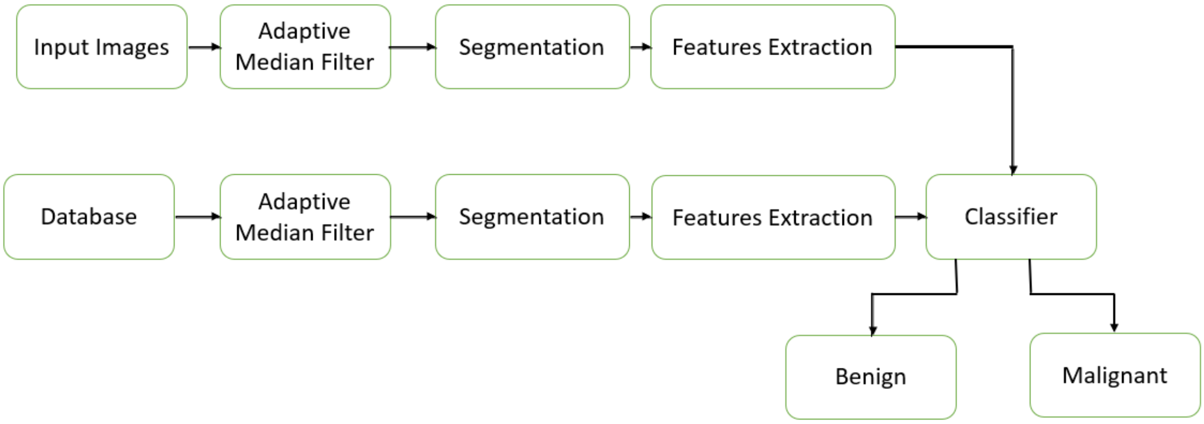

Fig 3.1.1

Input Images: Images are collected from a source for analysis.

Adaptive Median Filter: Both input images and database images go through an adaptive median filter, likely to reduce noise while preserving edges.

Segmentation: The filtered images undergo segmentation to isolate regions of interest.

Feature Extraction: Features are extracted from the segmented regions to create a feature set that represents the data.

Classifier: Extracted features are fed into a classifier that differentiates between two outcomes: Benign or Malignant.

3.2 IMAGE PROCESSING

Image processing is a technique that converts the ordinary image into digital form in order to enhance the quality of an image which gives useful information. The Medical image processing can be defined as picturing of body parts, tissues or organs for clinical analysis and treatment. It is one of the techniques used to create an image of the human body. Medical image processing is a highly challenging research area. The internal parts of the human body are diagnosed through medical imaging technique. Medical imaging has high importance because of correct diagnosis and treatment of diseases in health care system. The image of internal body parts where

produced by the equipment's like CT scanner, MRI, etc. The Medical image processing consists of medical signal assembly, image developing, processing the image and image display to medical analysis based on feature extraction. The outlining, noise cleaning, search, filtering, de-blurring and texture analysis are some of the basic techniques of image processing. Image processing covers four main areas namely image analyzation, visualization, information management, and image formation. The noise in the images may produce inaccurate data which can be rectified in the image processing. The clear view of the image is very much useful in diagnosis of the disease. The image analyzation produces detail information about the image which help in noise reduction. The various steps in image processing is shown in Fig.3.1.1

3.3 PRE-PROCESSING

The pre-processing phase has a great importance in the applications of image processing and specially segmentation. Generally in the pre-processing phase, the main goal is to remove the noise from the images. Undoubtedly MRI images have noises which have to be removed. But the noise deletion shouldn't destroy the edges of the image and decrease the clarity and quality of it. There are several methods for removing noise, including: Gaussian filter, contourlet transform approach and wavelet thresholding approach, median filter, anisotropic
diffusion filter. Anisotropic diffusion filter is a method for removing noise which is proposed by Persona and Malik. This method is for smoothing the image by preserving needed edges and structures. Fundamental idea is to adjust the smoothing level in a region based on the edge structure in the neighbourhood. Homogenous regions are highly smoothed and strong edge regions are barely smoothed (to preserve the structure).

ALGORITHM USED: Gaussian method in supervise machine learning

The Gaussian method in supervised machine learning typically refers to the use of Gaussian (normal) distributions in algorithms for tasks like classification, regression, or clustering.

1. Histogram Equalization

Algorithm: Histogram Equalization

Purpose: Enhance the contrast of the image by redistributing the pixel intensities, making features like tumors more visible.

Details:

The histeq function performs histogram equalization, adjusting the intensity distribution of the image to enhance contrast. It spreads the pixel values across the full range of intensity values.

matlab

Copy code

```
img_eq = histeq(img);
```

2. Median Filtering

Algorithm: Median Filter

Purpose: Reduce noise (e.g., salt-and-pepper noise) while preserving edges in the image.

Details:

The medfilt2 function applies a 2D median filter to the image using a 3x3 neighbourhood. This is a non-linear filter that replaces each pixel's value with the median of its neighbouring pixels, smoothing the image while preserving the structure of edges.

matlab

Copy code

```
img_filtered = medfilt2(img_eq, [3 3]);
```

3.Edge Detection

Algorithm: Canny Edge Detection

Purpose: Detect edges in the image, which can highlight boundaries of regions of interest (such as tumors).

Details:

The edge function with the Canny method is used. The Canny edge detection algorithm works in several stages:

Gaussian Smoothing: Reduces noise.

Gradient Calculation: Computes the gradient magnitude and direction.

Non-Maximum Suppression: Thin out edges by removing pixels that are not part of the edge.

Edge Tracking by Hysteresis: Uses high and low thresholds to identify strong and weak edges.

matlab

Copy code

```
edges = edge(img_filtered, 'Canny');
```

4. Morphological Operations

Algorithm: Dilation and Hole Filling (Morphological Operations)

Purpose: Refine the edge-detected image by closing small gaps and filling holes within regions.

Details:

Dilation (imdilate): Expands the boundaries of white regions in a binary image (edges), helping to fill small gaps between detected edges.

Hole Filling (imfill): Fills small gaps (holes) inside the detected regions, ensuring that connected regions are fully closed.

Structuring Element: A disk-shaped structuring element (strel('disk', 5)) is used for the dilation.

matlab

Copy code

```
se = strel('disk', 5);
dilated = imdilate(edges, se);
filled = imfill(dilated, 'holes');
```

5. Connected Component Labeling

Algorithm: Connected Component Labeling

Purpose: Label distinct connected regions (objects) in the binary image so that each region can be processed individually.

Details:

The bwlabel function is used to assign a unique label to each connected component (region) in the binary image (filled), where each connected region represents a potential tumor or abnormality.

matlab

Copy code

[labeledImage, numRegions] = bwlabel(filled);

6. Region Properties Calculation

Algorithm: Region Properties Analysis

Purpose: Extract properties of each labelled region, such as Area, Centroid, Bounding Box, and Perimeter, which are useful for feature extraction and classification.

Details:

The region props function calculates the following properties for each connected region:

Area: The total number of pixels within the region.

Centroid: The centre of mass of the region.

Bounding Box: The smallest rectangle that contains the region.

Perimeter: The total length of the region's boundary.

matlab

Copy code

```
stats = regionprops(labeledImage, 'Area', 'Centroid', 'BoundingBox', 'Perimeter');
```

7. Feature Extraction for Classification

Algorithm: Circularity (Shape Descriptor) and Feature Extraction

Purpose: Extract features like Area, Perimeter, and Circularity from each labelled region, which are used to classify the regions as Normal, Benign, or Malignant.

Circularity: Circularity is a shape descriptor that measures how round or irregular a shape is. It is calculated as:

Circularity = 4π×AreaPerimeter2

Circularity= Perimeter 24π×Area

Circularity values close to 1 indicate round shapes (e.g., benign tumors), and lower values (close to 0) suggest irregular shapes (e.g., malignant tumors).

matlab

Copy code

```
circularity = (4 * pi * features(k, 1)) / (perimeter^2);
```

8. Threshold-based Classification

Algorithm: Threshold-based Classification

Purpose: Classify detected regions into Normal, Benign, or Malignant based on their Area and Circularity values.

Details:The regions are classified based on two thresholds:

Area: A region's size is compared to a minimum threshold to determine if it's worth analysing.

Circularity: The shape is classified based on how circular or irregular it is.

The classification logic is:

Normal: Small area and low circularity.

Benign: Larger area and relatively high circularity.

Malignant: Larger area but low circularity, indicating irregular and possibly dangerous shapes.

matlab

Copy code

```
if area >threshold_area&& circularity >threshold_circularity
   if area >malignant_area_threshold&& circularity <malignant_circularity_threshold
      labels(k) = 2; % Malignant
   else
      labels(k) = 1; % Benign
   end
else
   labels(k) = 0; % Normal
end
```

9. Bounding Box Visualization

Algorithm: Bounding Box Drawing

Purpose: Visualize the classification results by drawing bounding boxes around the detected regions, with different colors representing different classifications (Normal, Benign, Malignant).

Details:

The rectangle function is used to draw bounding boxes around each region. The color of the bounding box is determined by the classification (blue for normal, yellow for benign, and red for malignant).

matlab

Copy code

```
rectangle('Position', stats(k).BoundingBox, 'EdgeColor', 'r', 'LineWidth', 2)
```

CHAPTER 4

STEPS INVOLVED IN DETECTION

STEP 1

```
% Step 1: Load the mammogram image
img = imread('file:///MATLAB Drive/s3.jpeg'); % Replace with your image path
imshow(img);
title('Original Mammogram Image');
```

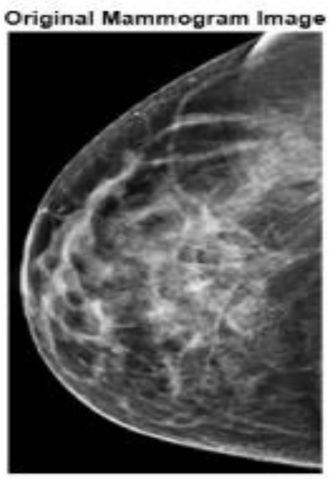

Fig 4.1.1

STEP 2-3:

```
% Step 2: Convert to grayscale if the image is RGB
if size(img, 3) == 3
img = rgb2gray(img);
end

% Step 3: Enhance Image (Histogram Equalization)
img_eq = histeq(img); % Enhance image contrast
figure;
imshow(img_eq);
title('Enhanced Mammogram Image');
```

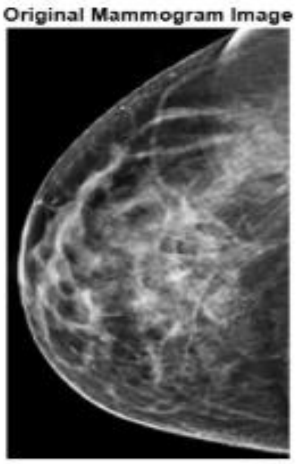

Fig 4.2.1

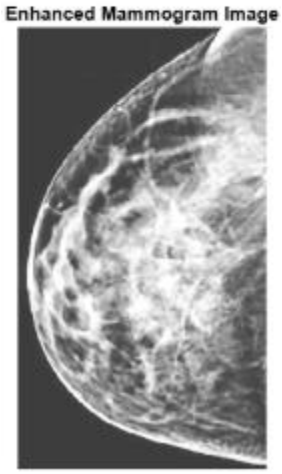

Fig 4.2.2

STEP 4:

```
% Step 4: Apply Median Filter for Noise Removal
img_filtered = medfilt2(img_eq, [3 3]); % Median filter
figure;
imshow(img_filtered);
title('Filtered Image');
```

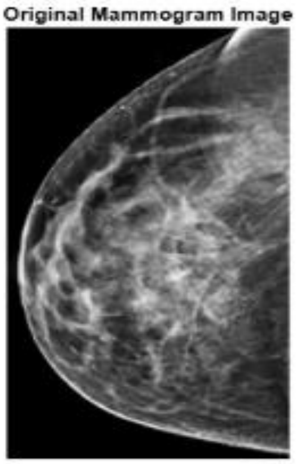

Fig 4.3.1

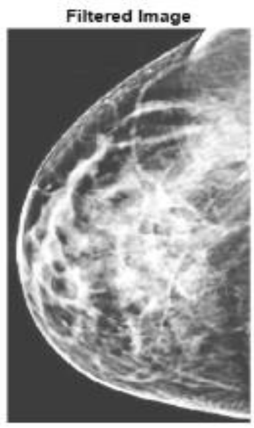

Fig 4.3.2

STEP 5-9:

```
% Step 5: Perform Edge Detection (Canny)
edges = edge(img_filtered, 'Canny');
figure;
imshow(edges);% Step 4: Apply Median Filter for Noise Removal
img_filtered = medfilt2(img_eq, [3 3]); % Median filter
```

```
figure;
imshow(img_filtered);
title('Filtered Image');
title('Canny Edge Detection');


% Step 6: Apply Morphological Operations (Dilate to close gaps)
se = strel('disk', 5); % Structuring element
dilated = imdilate(edges, se); % Dilation to fill gaps in edges
filled = imfill(dilated, 'holes'); % Fill holes in detected regions

% Step 7: Label Connected Components
[labeledImage, numRegions] = bwlabel(filled); % Label connected components
stats = regionprops(labeledImage, 'Area', 'Centroid', 'BoundingBox', 'Perimeter');

% Step 8: Extract Features (Area, Circularity, Perimeter)
features = zeros(numRegions, 3); % Feature matrix: [Area, Circularity, Perimeter]
for k = 1:numRegions
% Area of the detected region
features(k, 1) = stats(k).Area;

% Circularity: (4 * pi * Area) / Perimeter^2
   perimeter = stats(k).Perimeter;
   circularity = (4 * pi * features(k, 1)) / (perimeter^2);
features(k, 2) = circularity;

% Perimeter of the detected region
features(k, 3) = perimeter;
end
% Step 9: Classification (Benign vs Malignant Differentiation)
threshold_area = 100; % Minimum area for detection
threshold_circularity = 0.5; % Circularity threshold for tumors
malignant_area_threshold = 500; % Larger area indicates potential malignancy
malignant_circularity_threshold = 0.3; % Lower circularity indicates malignancy
```

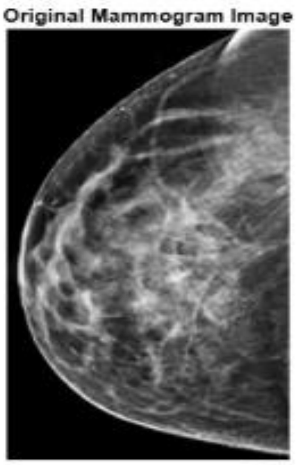

Fig 4.4.1

Fig 4.4.2

STEP 10-11:

```
% Step10:Classify regions based on area and circularity
labels = zeros(numRegions, 1); % 0 = Normal, 1 = Benign, 2 = Malignant
for k = 1:numRegions
   area = features(k, 1);
   circularity = features(k, 2);

if area >threshold_area&& circularity >threshold_circularity
if area >malignant_area_threshold&& circularity <malignant_circularity_threshold
          labels(k) = 2; % Malignant
else
          labels(k) = 1; % Benign
end
else
      labels(k) = 0; % Normal
end
end


% Step 11: Visualize the Classification
figure;
imshow(img);
hold on;
for k = 1:numRegions
if labels(k) == 1
% Benign (Yellow)
rectangle('Position', stats(k).BoundingBox, 'EdgeColor', 'y', 'LineWidth', 2);
elseif labels(k) == 2
% Malignant (Red)
rectangle('Position', stats(k).BoundingBox, 'EdgeColor', 'r', 'LineWidth', 2);
else
% Normal (Blue)
rectangle('Position', stats(k).BoundingBox, 'EdgeColor', 'b', 'LineWidth', 2);
end
end
title('Classification Results (Blue = Normal, Yellow = Benign, Red = Malignant)');
hold off;
```

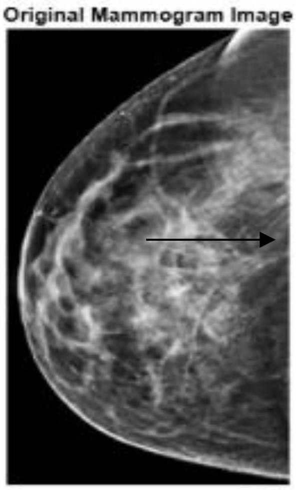

Fig 4.5.1

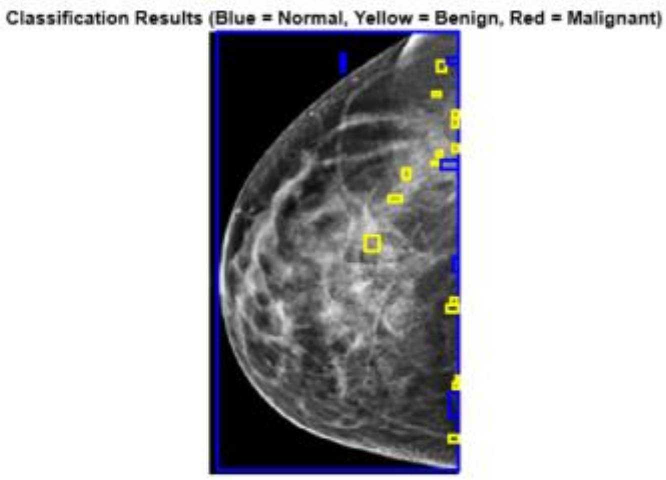

Fig 4.5.2

CHAPTER 5

RESULTS

Machine learning (ML) algorithms have revolutionized breast cancer detection using mammograms by significantly enhancing diagnostic accuracy and efficiency. Traditional ML algorithms such as Support Vector Machines (SVM), Random Forest (RF), and K-Nearest Neighbors (KNN) have been widely applied to classify breast tissues as benign or malignant. For example, SVM is known for its robust performance in classification tasks, achieving accuracy

levels between 85-95% when paired with effective feature extraction methods like texture or shape analysis. Random Forest, leveraging ensemble learning, offers comparable accuracy with improved specificity, effectively minimizing false positives. However, simpler algorithms like KNN, while effective for small datasets, are sensitive to noise and may not perform as well on larger datasets.

In recent years, deep learning models, particularly Convolutional Neural Networks (CNNs), have become the preferred choice for breast cancer detection due to their ability to learn hierarchical features directly from raw mammogram images. CNN-based models such as ResNet, VGG, and Inception have achieved remarkable accuracy, often exceeding 90%. Transfer learning, where pre-trained models like ResNet-50 or Inception-V3 are fine-tuned on mammogram datasets, has further improved performance, reaching accuracy levels as high as 98-99%. These models not only enhance accuracy but also significantly reduce false negatives, making them highly effective in early cancer detection.

The success of ML models depends heavily on the quality of training datasets. Commonly used datasets include the Digital Database for Screening Mammography (DDSM), Mammographic Image Analysis Society (MIAS) dataset, and the Curated Breast Imaging Subset of DDSM (CBIS-DDSM). These datasets provide annotated mammogram images that facilitate robust model training and evaluation. Performance is typically measured using metrics such as accuracy, sensitivity (recall), specificity, and the Area Under the Receiver Operating Characteristic Curve (AUC-ROC). State-of-the-art deep learning models frequently achieve AUC-ROC values above 0.90, demonstrating their effectiveness in distinguishing between malignant and benign cases.

Despite these advancements, challenges remain. Data imbalance, where malignant cases are significantly fewer than benign ones, can skew model performance. Techniques like oversampling, under sampling, or synthetic data generation (e.g., SMOTE) are often employed to address this issue. Additionally, the interpretability of deep learning models is a concern, as they often function as "black-box" systems. Research in explainable AI aims to make these models more transparent and trustworthy for clinical use. Moreover, ensuring that models generalize well across different populations and imaging systems is crucial for real-world deployment.

In clinical practice, ML models are increasingly integrated as decision-support tools, aiding radiologists in analyzing mammograms. These systems help streamline large-scale screening programs, enabling faster and more accurate diagnoses. While ML models are not yet a replacement for radiologists, their ability to enhance diagnostic efficiency and reduce workload makes them an invaluable asset in the fight against breast cancer.

Conclusion

The application of machine learning algorithms in breast cancer detection using mammograms has significantly improved diagnostic accuracy, sensitivity, and efficiency. Traditional algorithms like SVM and Random Forest provide reliable classification, while advanced deep learning models, particularly CNNs and transfer learning approaches, have set new benchmarks in performance, achieving accuracies as high as 99%. Despite challenges such as data imbalance, model interpretability, and generalizability, ongoing research and the development of explainable AI are addressing these limitations. When integrated into clinical workflows, these models serve as powerful decision-support tools, complementing radiologists and enhancing the overall effectiveness of breast cancer screening programs. As the field continues to evolve, machine learning is poised to play an increasingly vital role in early cancer detection and improving patient outcomes.