SOMAIYA VIDYAVIHAR UNIVERSITY

K. J. Somaiya College of Engineering, Mumbai -77
(A Constituent College of Somaiya Vidyavihar University)

Breast Cancer Detection System

Submitted at the end of semester VI in partial fulfilment of requirements of

TY Bachelors in Technology in Computer Engineering

by

Kavya Vasa

Roll No: 16010122812

Krisha Rawal

Roll No: 16010122817

Sakshi Bhosale

Roll No: 16010122826

Guide

Dr. Prasanna Shete

Department of Computer Engineering

K. J. Somaiya College of Engineering, Mumbai-77

(A Constituent College of Somaiya Vidyavihar University)

Batch 2023-24

SOMAIYA VIDYAVIHAR UNIVERSITY

K. J. Somaiya College of Engineering, Mumbai -77
(A Constituent College of Somaiya Vidyavihar University)

Certificate

This is to certify that the MINI PROJECT report entitled **Breast Cancer Detection System** submitted by Krisha Rawal, Kavya Vasa and Sakshi Bhosale at the end of semester VI of TY B. Tech are bona fide record for partial fulfilment of requirements for the degree of Bachelors in Computer Engineering of University of Mumbai

Guide

Head of the Department

Examiner

Date: 24/4/2024

Place: Mumbai-77

Certificate of Approval of Examiners

We certify that this Mini Project report entitled **Breast Cancer Detection System** is bona fide record of Mini project work done by Sakshi Bhosale, Krisha Rawal & Kavya Vasa during semester VI. This Mini project work is submitted at the end of semester VI in partial fulfilment of requirements for the degree of Bachelors in Technology in Computer Engineering of University of Mumbai.

Internal Examiner 1

Internal Examiner 2

Date:

Place: Mumbai-77

Declaration

We declare that this written report submission represents the work done based on our and / or others' ideas with adequately cited and referenced the original source. We also declare that we have adhered to all principles of intellectual property, academic honestyand integrity as we have not misinterpreted or fabricated or falsified any idea/data/fact/source/original work/ matter in my submission.

We understand that any violation of the above will be cause for disciplinary action by the college and may evoke the penal action from the sources which have not been properly cited or from whom proper permission is not sought.

Signature of the Student Roll No.	**Signature of the Student Roll No.**
Signature of the Student Roll No.	**Signature of the Student Roll No.**

Date:

Place: Mumbai-77

Acknowledgement

In the first place, we would like to thank K.J. Somaiya College of Engineering for giving us an opportunity to work on this project. It is a privilege for us to have been associated with Dr. Prasanna Shete, our guide, during this project work. We have been Greatly Benefited by her valuable suggestions and ideas. It is with great pleasure that we express our deep sense of gratitude to them for their valuable guidance, constant encouragement and patience throughout this work. We express our gratitude to Principal, Head of Department for their constant encouragement, cooperation, and support. We take this opportunity to thank all our classmates for their company during the course work and for useful discussion we had with them. We would be failing in our duties if we do not acknowledge with a deep sense of gratitude the sacrifices made by our parents and thus have helped us in completing all the project work successfully.

Sakshi Bhosale
Kavya Vasa
Krisha Rawal

Abstract

Breast cancer is a significant public health concern globally, with early detection being critical for successful treatment and improved patient outcomes. Mammography, as a widely utilized screening tool, plays a pivotal role in early detection. However, the interpretation of mammograms is challenging due to the subtle and complex patterns associated with breast abnormalities. In recent years, machine learning (ML) techniques have shown promise in enhancing the accuracy and efficiency of breast cancer detection from mammographic images.

This review aims to provide a comprehensive overview of recent advancements in the application of ML algorithms for breast cancer detection using mammograms. We begin by discussing the fundamentals of mammography and the challenges associated with traditional interpretation methods. Subsequently, we delve into various ML approaches, including supervised, unsupervised, and deep learning techniques, employed for feature extraction, classification, and segmentation tasks in mammographic images.

Overall, the integration of machine learning with mammography is highly useful for study purposes, enabling researchers to advance our understanding of breast cancer and providing clinicians with valuable tools for diagnosis and treatment. By harnessing the power of technology, we can improve outcomes for breast cancer patients and enhance the efficiency of healthcare delivery.

Index

CHAPTER 1:INTRODUCTION

1.1 Background

The project focuses on detecting breast cancer using mammograms, with the aim of aiding in the early diagnosis of the disease. Breast cancer is a significant health concern globally, with increasing cases prompting the need for more effective detection methods. The project aims to automate the detection process to facilitate early diagnosis, which can significantly improve the chances of successful treatment.

The project's methodology involves several steps. Initially, adaptive mean filtering is applied to the mammogram images to remove noise and enhance image quality. This filtering technique is chosen for its effectiveness in distinguishing fine details from noise. Following this preprocessing step, Gaussian Mixture Model (GMM) segmentation is performed to classify regions in the image. Additionally, k-means segmentation is employed to further analyze the image data.

A notable aspect of the project is the implementation of the Hidden Markov Random Field (HMRF) model and its Expectation-Maximization (EM) algorithm. These techniques are utilized to refine the segmentation results and improve the accuracy of cancer detection.

The project is implemented in MATLAB and provides a user-friendly interface (GUI) for easy interaction. Users can input mammogram images and follow the steps outlined in the GUI to analyze the images for the presence of cancer. The project draws from research literature, citing relevant studies that contribute to the understanding and development of the detection methodology.

Using an adaptive mean filter for breast cancer detection in mammograms offers several advantages over other filters, primarily due to its ability to preserve edge information while effectively reducing noise. Here's why adaptive mean filtering may be preferred :

The adaptive mean filter calculates the mean value in the neighborhood of each pixel and replaces the pixel value with this mean value. However, the size of the neighborhood changes depending on the local variance of the image. This allows the filter to adapt to the local image characteristics, providing better noise reduction while preserving image details, including edges.

The formula for the adaptive mean filter can be expressed as follows:

For each pixel $(x,y)(x,y)$ in the image:

1. Determine the local window size based on the local variance.
2. Calculate the mean value $M(x,y)M(x,y)$ within the local window.
3. Replace the pixel value at $(x,y)(x,y)$ with $M(x,y)M(x,y)$.

Advantages of Adaptive Mean Filter:

1. Preservation of Edge Information: The adaptive nature of the filter ensures that smaller neighborhoods are used in regions with high variance, preserving edges and fine details in the image. This is crucial in mammogram analysis, where edge information plays a significant role in detecting abnormalities.

2. Noise Reduction: By adjusting the neighborhood size according to local variance, the adaptive mean filter effectively reduces noise while maintaining image clarity. This is particularly important in medical imaging, where noise can obscure subtle features indicative of abnormalities.

3. Adaptability to Image Variations: Mammograms can exhibit variations in brightness, contrast, and noise levels across different regions. The adaptive nature of the filter allows it to adjust to these variations, resulting in more consistent and reliable noise reduction across the entire image.

Why we did not use Other Filters?

1. Linear Filters (e.g., Gaussian, Median): Linear filters, such as Gaussian or median filters, apply a fixed-size kernel to the entire image, which may not effectively adapt to local variations in noise and image characteristics. This can lead to either excessive smoothing, which may blur important details, or inadequate noise reduction.

2. Non-local Means Filter: While non-local means filtering can be effective in preserving fine details and reducing noise, it can be computationally expensive, especially for large images such as mammograms. In medical imaging applications where real-time or near-real-time processing is desired, computational efficiency is a critical factor.

3. Edge-Preserving Filters (e.g., Bilateral Filter): Edge-preserving filters are designed to preserve edges while reducing noise. However, they may not always adapt well to local variations in noise levels, especially in images with complex structures or heterogeneous noise characteristics like mammograms.

GMM (Gaussian Mixture Model) segmentation is a powerful technique commonly used in image processing and computer vision tasks, including medical image analysis such as breast cancer detection using mammograms. Here's why GMM segmentation can be a suitable choice:

1. Flexibility in Modeling Diverse Data Distributions: GMM can effectively model complex data distributions by representing them as a mixture of several Gaussian distributions. In the context of breast cancer detection, mammograms often contain regions with varying pixel intensities and textures. GMM can flexibly model these variations, making it suitable for segmenting different tissue types and abnormalities.

2. Probabilistic Framework for Uncertainty Estimation: GMM assigns a probability to each pixel belonging to different tissue types or classes. This probabilistic approach provides a measure of uncertainty in the segmentation, which can be valuable in medical image analysis where accurate delineation of boundaries is crucial.

3. Adaptability to Varying Image Characteristics: Mammograms can exhibit significant variations in contrast, noise levels, and anatomical structures across different patients and imaging conditions. GMM segmentation can adapt to these variations by learning the parameters of the Gaussian distributions from the image data itself, without relying on predefined thresholds or assumptions.

HMRF-EM introduces spatial regularization by incorporating spatial dependencies among neighboring pixels. Here's how HMRF-EM enhances GMM segmentation for breast cancer detection:

1. Spatial Smoothing: HMRF-EM imposes constraints on the segmentation result by penalizing abrupt changes in tissue classes within local neighborhoods. This helps smooth out segmentation boundaries and reduce noise sensitivity, which is particularly beneficial in medical images where smooth transitions between tissue types are expected.

2. Preservation of Local Image Structure: By considering spatial context, HMRF-EM ensures that the segmented regions adhere to the underlying anatomical structures present in the mammograms. This helps in preserving important features and details relevant to identifying abnormalities such as tumors or lesions.

The formulas for GMM segmentation and HMRF-EM regularization:

1. Gaussian Mixture Model (GMM) Segmentation:

The probability density function (PDF) of a Gaussian distribution is given by:

$$f(x|\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

Where:

- x is the pixel intensity,
- μ is the mean of the distribution,
- σ is the standard deviation of the distribution.

For a mixture of K Gaussian distributions in a GMM, the overall PDF is given by the weighted sum of individual Gaussian distributions:

$$p(x) = \sum_{i=1}^{K} \pi_i \cdot f(x|\mu_i, \sigma_i)$$

Where:

- $p(x)$ is the probability density function of the GMM,
- π_i is the weight (or mixing coefficient) of the i-th Gaussian component.

The parameters $\mu i \mu_i$, $\sigma i \sigma_i$, and $\pi i \pi_i$ are typically estimated using the Expectation-Maximization (EM) algorithm.

2. Hidden Markov Random Field Expectation-Maximization (HMRF-EM):

HMRF-EM adds spatial regularization to GMM segmentation by incorporating a Markov Random Field (MRF) prior over the segmentation labels. The energy function for HMRF is defined as:

$$E(\mathbf{y}, \mathbf{x}) = \sum_{i} U(y_i, x_i) + \sum_{i,j} V(y_i, y_j)$$

Where:

- $E(\mathbf{y}, \mathbf{x})$ is the total energy function,
- $\mathbf{y}$ represents the label field (segmentation result),
- $\mathbf{x}$ represents the observed image data,
- $U(y_i, x_i)$ is the unary potential function, measuring the compatibility between the label y_i and the observed data x_i,
- $V(y_i, y_j)$ is the pairwise potential function, measuring the smoothness constraint between neighboring labels y_i and y_j.

The HMRF-EM algorithm iteratively updates the label field (y) and the model parameters θ (including the parameters of the GMM) to minimize the energy function $E(\mathbf{y},\mathbf{x})$. This is achieved using an expectation-maximization framework, where the expectation step estimates the posterior probabilities of the labels given the observed data and the current model parameters, and the maximization step updates the model parameters based on these posterior probabilities.

In summary, GMM segmentation with HMRF-EM regularization combines the probabilistic modeling of pixel intensities using GMM with the spatial smoothing constraints provided by the HMRF, resulting in accurate and smooth segmentation of mammogram images for breast cancer detection.
Now, regarding why other segmentation methods might not be as suitable for breast cancer detection:

1. Threshold-based Methods: Simple thresholding techniques may struggle to accurately segment mammograms due to variations in image intensity and noise levels. They often require manual tuning of thresholds, which can be subjective and may not generalize well across different datasets.

2. Region-based Methods: While region-based segmentation methods consider homogeneity within regions, they may not effectively capture the spatial relationships between different tissue types. This could lead to fragmented or inaccurate segmentations, especially in complex mammogram images with overlapping structures.

3. Edge-based Methods: Edge detection algorithms may highlight boundaries between tissues but may not provide sufficient information about the internal texture and composition of the segmented regions. This could limit their effectiveness in detecting subtle abnormalities characteristic of early-stage breast cancer.

Using Probabilistic Neural Networks (PNN) for this project can offer several advantages. PNN is a type of neural network particularly suitable for classification tasks where probabilistic outputs are desired. Here's why PNN can be a good choice for breast cancer detection:

1. Probabilistic Output: PNN provides probabilistic outputs, which can be beneficial for medical diagnoses like breast cancer detection. Instead of just predicting a binary outcome (cancer or non-cancer), PNN can give the probability of an image being cancerous. This can help doctors in making more informed decisions about further diagnostic procedures or treatments.

2. Non-linearity Handling: PNN can capture complex non-linear relationships between features extracted from mammograms and the presence of cancer. Mammograms contain intricate patterns and textures that may not be easily captured by linear classifiers like Support Vector Machines (SVM).

3. Fast Training and Testing: PNN typically requires less computational resources for training compared to other complex neural network architectures like Convolutional Neural Networks (CNN) or Recurrent Neural Networks (RNN). This can be advantageous in medical settings where computational resources might be limited.

4. Interpretability: PNN provides a clear probabilistic interpretation of its outputs, making it easier for medical professionals to understand and trust the model's predictions. This interpretability is crucial in medical applications where decisions can have significant consequences.

Now, let's discuss the formula for PNN:

The formula for the output of a PNN for a given input x and class j can be expressed as:

$$P(j|x) = \frac{1}{\sigma\sqrt{2\pi}} \sum_{i=1}^{N_j} e^{-\frac{1}{2\sigma^2}||x-x_{ij}||^2}$$

Where:

- $P(j|x)$ is the probability that input x belongs to class j.
- N_j is the number of training samples in class j.
- x_{ij} is the ith training sample in class j.
- σ is a smoothing parameter that controls the width of the Gaussian kernel.

Now, let's address why other classification algorithms like QNN (Quantum Neural Networks), CNN, SVM, and RNN may not be selected for breast cancer detection:

1. QNN: Quantum Neural Networks are still in the early stages of development and may not yet offer significant advantages over classical neural networks for medical image analysis tasks like breast cancer detection. Additionally, the hardware requirements and expertise needed for implementing QNNs may be prohibitive.

2. CNN: While CNNs are powerful for image recognition tasks, they often require a large amount of labeled data and computational resources for training, which may not be readily available in medical imaging datasets. Additionally, interpreting the decisions made by CNNs can be challenging due to their black-box nature.

3.SVM: SVMs are effective for binary classification tasks, but they may struggle with capturing complex non-linear relationships in high-dimensional feature spaces, which are common in medical imaging data. SVMs also do not provide probabilistic outputs inherently, which may be desirable for medical diagnoses.

4. RNN: Recurrent Neural Networks are primarily used for sequential data where the order of inputs matters, such as natural language processing or time-series analysis. Mammograms are static images, and the sequential nature of RNNs may not be well-suited for this type of data.

Using the K-means algorithm for clustering in breast cancer detection systems using mammograms offers several advantages over other clustering algorithms, primarily due to its simplicity, efficiency, and effectiveness in identifying clusters within the data. Here's why K-means might be preferred and why other clustering algorithms might not be suitable:

The K-means algorithm partitions a dataset into K clusters by iteratively assigning each data point to the cluster with the nearest mean (centroid) and updating the centroids based on the mean of the data points assigned to each cluster. The algorithm aims to minimize the within-cluster variance, often measured using squared Euclidean distance.

The formula for the K-means clustering algorithm can be summarized as follows:

Given a dataset \(X\) consisting of \(n\) data points and a predetermined number of clusters \(k\), the K-means algorithm aims to partition the data into \(k\) clusters such that the within-cluster sum of squares (WCSS) is minimized. The steps of the algorithm can be summarized as follows:

1. Initialization: Randomly initialize \(k\) cluster centroids in the feature space.
2. Assignment: Assign each data point to the nearest cluster centroid based on a distance metric (usually Euclidean distance).
3. Update Centroids: Recalculate the centroids of the clusters by taking the mean of all data points assigned to each cluster.
4. Repeat: Iterate steps 2 and 3 until convergence criteria are met (e.g., centroids do not change significantly, or a maximum number of iterations is reached).

The objective function \(J\) of K-means, which represents the total within-cluster variance, can be expressed as:

$$J = \sum_{i=1}^{k} \sum_{x \in C_i} ||x - \mu_i||^2$$

Where:

- C_i represents the i^{th} cluster.
- μ_i is the centroid of cluster C_i.
- $||x - \mu_i||$ denotes the Euclidean distance between data point x and the centroid μ_i.

Minimizing J effectively optimizes the clustering such that the data points within each cluster are close to their respective centroids while maximizing the separation between clusters. This objective aligns with the goal of identifying distinct patterns or structures in the mammogram images that may correspond to different tissue types or abnormalities.
Advantages of K-means:

1. Simplicity: K-means is easy to implement and understand, making it accessible even to users with limited experience in machine learning and clustering techniques.

2. Efficiency: The algorithm's computational complexity is relatively low, especially compared to more complex clustering algorithms. It is suitable for large datasets and real-time applications.

3. Effectiveness in Clustering: K-means is effective in identifying compact, spherical clusters within the data. In breast cancer detection systems, where distinct patterns or clusters of abnormal tissue may indicate the presence of cancer, K-means can help identify these patterns efficiently.

Why Not Other Clustering Algorithms?:

1. Hierarchical Clustering: While hierarchical clustering can be useful for exploring hierarchical structures in the data, it may not be as efficient or scalable as K-means, especially for large datasets. Additionally, hierarchical clustering may produce dendrograms that are not as straightforward to interpret in the context of breast cancer detection.

2. Density-Based Clustering (e.g., DBSCAN): Density-based clustering algorithms like DBSCAN are effective in identifying clusters of arbitrary shapes and sizes. However, they may require tuning of parameters such as minimum cluster size and neighborhood density, which can be challenging in medical imaging datasets where the characteristics of normal and abnormal tissue may vary widely.

3. Gaussian Mixture Models (GMM): GMMs model the data distribution as a mixture of Gaussian distributions and can capture complex cluster shapes and densities. However, GMMs may be computationally more expensive than K-means, especially for high-dimensional data like mammograms, and may require careful initialization and regularization to avoid overfitting.

The training process of a Probabilistic Neural Network (PNN) involves storing the input patterns along with their corresponding class labels and estimating the probability density functions (PDFs) for each class. Here's overview of how PNN training typically works in this project:

1. Initialization: Initialize the PNN model with random weights and parameters.

2. Input Pattern Processing: For each input pattern in the training dataset:
   - Calculate the activation of each neuron in the input layer based on the input pattern.
   - Propagate the activations forward through the network.

3. Pattern Storage: Store the processed input pattern and corresponding class label in the PNN.

4. Kernel Density Estimation (KDE): For each class in the dataset:
   - Estimate the probability density function (PDF) of the input patterns belonging to that class using a kernel function.
   - Calculate the mean and standard deviation of the input patterns for each class.

The PDF represents the probability distribution of the input patterns belonging to each class. The formula for KDE generally involves calculating the weighted sum of kernel functions centered at each training data point:

$$p(x|C_i)=\frac{1}{N}\sum_{j=1}^{N}K_h(x-x_j)$$

Where:

- $p(x|C_i)$ is the estimated PDF for class C_i.
- N is the number of training patterns in class C_i.
- x_j represents the jth training pattern in class C_i.
- K_h is the kernel function with bandwidth parameter h.

5. Training Complete: Once all input patterns have been processed and KDE is performed for each class, the training of the PNN is complete.

After training, the PNN is ready to classify new input patterns. During classification, the PNN calculates the probability of the input pattern belonging to each class using the stored PDFs and class statistics.

Overall, the project represents a valuable contribution to the field of medical image processing, offering a promising approach to automated breast cancer detection using mammograms. It showcases the potential of advanced image processing techniques and machine learning algorithms in improving healthcare outcomes, particularly in the early diagnosis of cancer.

1.2 Problem Statement

1. Cancer, particularly breast cancer, is a significant health concern globally and in India, with an increasing number of individuals affected by the disease.
2. To address the need for an automated breast cancer detection system due to the difficulty in detecting subtle abnormalities in mammographic images.
3. Early detection of cancer is crucial for effective treatment and improved survival rates. However, the subtle signs of breast cancer, such as masses and microcalcification clusters, pose challenges for accurate diagnosis.
4. To analyze mammograms much faster than humans.
5. Assisting radiologists in identifying suspicious areas on mammograms, detection systems can optimize the use of healthcare resources, ensuring that patients receive appropriate follow-up care based on the severity of their condition.
6. In regions with limited access to healthcare professionals, automated detection systems can serve as a valuable tool for providing screening and early detection services, particularly in underserved communities.
7. Breast cancer screening programs often involve large numbers of mammograms, making manual interpretation impractical. Automated systems can help in processing these volumes efficiently, ensuring timely screening for at-risk individuals.
8. By automating the detection process, the project seeks to improve accessibility to early diagnosis, thereby increasing the chances of successful treatment outcomes and reducing mortality rates associated with breast cancer.

1.3 Scope of the Project

This code seems to be a MATLAB implementation of a system for detecting breast cancer from mammogram images. Here's a breakdown of the steps involved:

1. Preprocessing:
   - Adaptive mean filtering is applied to remove noise from the image.
   - The image is converted to grayscale and then to an unsigned 8-bit integer format.

2. Segmentation:
   - Gaussian Mixture Model (GMM) segmentation is performed on the preprocessed image.
   - K-means segmentation is also applied.

3. Feature Extraction:
   - Hidden Markov Random Field (HMRF) model is implemented for feature extraction.

4. Classification:
   - The extracted features are compared with features from a database of known cases (benign, malignant, normal).
   - A neural network classifier is trained using the features from the database.

5. User Interface:
   - A GUI is provided for users to browse images, apply filters, perform segmentation, and classify.

6. References:
   - The project is based on research papers referenced in the document, indicating the methods used and their effectiveness.

7. Operation:
   - The project is designed to be run as an application, where users can open it and follow the steps outlined in the GUI.

CHAPTER 2: LITERATURE SURVEY

2.1 Survey

The paper titled "A Novel Approach for Breast Cancer Detection and Segmentation in a Mammogram" by Anuj Kumar Singh and Bhupendra Gupta addresses the critical issue of breast cancer detection using mammogram images. Breast cancer remains a leading cause of death among women worldwide, emphasizing the importance of early detection for improved survival rates. Mammography, as a widely used diagnostic technique, plays a crucial role in identifying suspicious lesions indicative of breast cancer. However, accurately distinguishing cancerous tissues from normal dense tissues in mammogram images presents a significant challenge.

Mammography is currently one of the important methods to detect breast cancer early. The magnetic resonance imaging (MRI) is the most attractive alternative to mammogram. However, the MRI test is done when the radiologists want to confirm about the existence of the tumor [5]. Bethapudi et al. introduced a method involving thresholding, median filtering, and morphological operations to detect mass structures in digital mammogram images. While effective, these methods may not always provide satisfactory results due to variations in image characteristics. Basheer and Mohammed introduced a segmentation method based on adaptive median filtering and texture analysis for breast mass segmentation. This approach aims to contour the image based on texture properties to identify regions of interest[1].

Dalmiya et al. proposed a segmentation method using wavelet-based k-means clustering for mammogram segmentation. By extracting high-level details using wavelet transform and clustering, this method aims to locate tumor regions accurately. Cascio et al. proposed a segmentation method using contour searching and neural network classification for mass lesions. By employing supervised neural networks, this approach aims to classify regions as pathological or non-pathological based on extracted features. Guliato et al. introduced segmentation methods using fuzzy sets and region growing techniques. These methods aim to determine tumor boundaries based on fuzzy logic principles and region growing algorithms.

Along with segmentation, pixels of cancer region are also identified. The method is simple and fast because of using basic image processing techniques. The method can also be helpful in other medical imaging applications, pattern matching, feature extraction[1].

Many breast conditions mimic the symptoms of cancer and need tests and sometimes a biopsy for diagnosis. False positive results occur when mammogram finds something that looks like cancer, but turns out to be benign (not cancer). Depending on the density of the breasts radiologists may miss up to 30% of breast cancers [2].

Most pre-trained neural network models in deep learning can be broadly grouped into artificial neural networks, convolutional neural networks and recurrent neural networks. However, there are also neural networks that do not fall directly into any of these classes because they operate using an integration of multiple network types such as generative adversarial networks and deep belief network [4].

Experimental results are presented, showcasing the classification accuracy of the PSOWNN classifier compared to traditional classifiers like Feedforward Neural Network (FFNN). The results demonstrate the effectiveness of the PSOWNN approach in detecting various types of abnormalities in mammograms, with a high sensitivity and specificity.

We presented the SLR to systematically review the existing research works on breast cancer diagnosis based on the deep learning-based methods involving genetic sequencing or the histopathological imaging process [3]. It mentions various aspects covered in these reviews, such as image modalities, classification techniques, and imaging methods like mammography and MRI. This contextualizes the current study within the broader landscape of breast cancer research.

In terms of results and findings, the survey identifies 98 research articles for review after screening and quality evaluation. It highlights CNNs as the most accurate and extensively used model for breast cancer detection, with accuracy metrics being the preferred evaluation method. This synthesis of findings provides valuable insights into the effectiveness of deep learning approaches in breast cancer diagnosis.

More studies should emphasize deriving important aspects from gene expression data to improve outcomes and improve accuracy by employing confusion matrix parameters. This opens up the prospect for future studies to address a range of related issues, including determining risk levels and predicting the likelihood of recurrence. Breast cancer detection and survival likelihood were the main focuses of the majority of studies, which only used genetic sequencing data with binary categorization. Furthermore, large-scale, thorough and fully labeled WSI datasets are currently lacking. Consequently, the creation of sizable public databases is crucial for future research [3].

2.2 Objective of the Project

1. Data Preprocessing: Utilize adaptive mean filtering to remove noise from mammographic images while preserving fine details crucial for accurate analysis. Convert the preprocessed images into grayscale and adjust the pixel intensity to ensure compatibility with machine learning and deep learning models.

2. Feature Extraction: Extract relevant features from preprocessed mammographic images that are indicative of potential cancerous regions. Explore techniques such as texture analysis, edge detection, and shape descriptors to capture key characteristics associated with breast cancer.

3. Model Development: Develop machine learning and deep learning models capable of accurately detecting breast cancer from mammographic images. Experiment with various architectures, including convolutional neural networks (CNNs), support vector machines (SVMs), and ensemble methods, to identify the most effective approach.

4. Training and Validation: Train the developed models using annotated mammographic datasets to learn the patterns associated with both cancerous and non-cancerous regions. Validate the trained models using cross-validation techniques to ensure robustness and generalizability.

5. Segmentation: Explore segmentation algorithms such as Gaussian Mixture Model (GMM) and k-means clustering to partition mammographic images into distinct regions corresponding to potential tumors. Evaluate the performance of segmentation algorithms in accurately delineating cancerous areas from background tissue.

6. Integration of HMRF-EM: Implement Hidden Markov Random Field with Expectation-Maximization (HMRF-EM) algorithm to refine segmentation results and improve the delineation of tumor boundaries. Assess the effectiveness of HMRF-EM in enhancing the accuracy of tumor localization and characterization.

7. Deployment and Integration: Develop a user-friendly interface for the breast cancer detection system, allowing healthcare professionals to easily upload mammographic images for analysis. Ensure seamless integration of the system into existing clinical workflows to facilitate early detection and treatment planning for breast cancer patients.

CHAPTER 3: PROJECT DESIGN

3.1 Proposed System Design

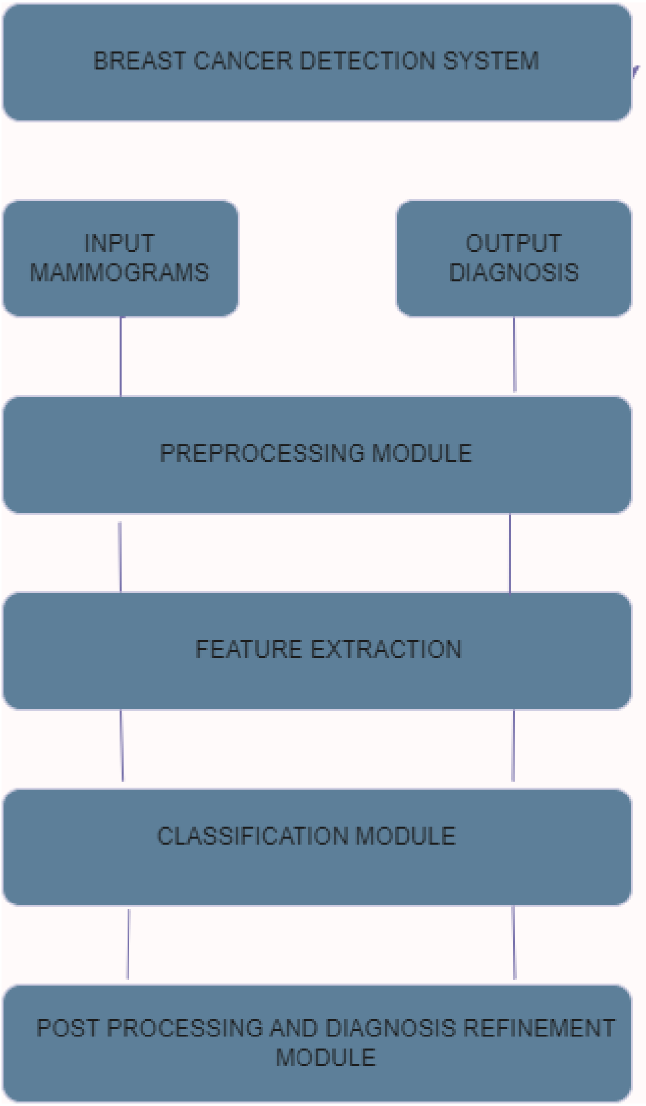


Fig 3.1 :Context Diagram

3.2 System Architecture

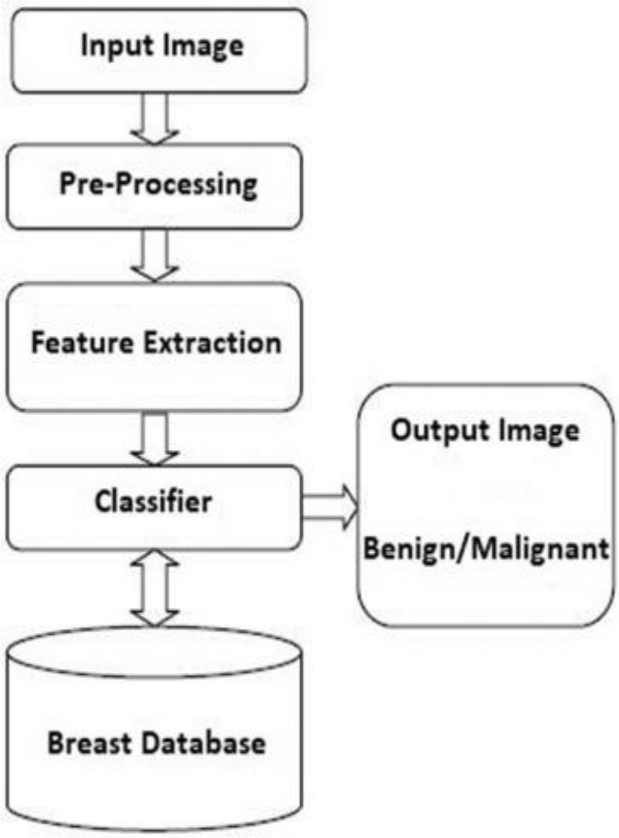


Fig 3.2 :System Architecture

3.3 Software Design

3.3.1Frontend Design

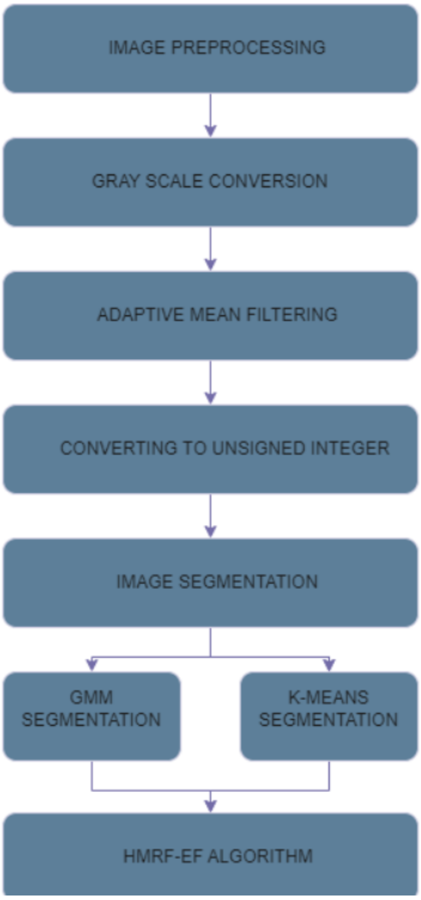


Fig 3.3.1 :frontend Diagram

3.3.2 Backend Design

Fig 3.3.2: Backend Diagram

3.3.3 System Architecture Design

Fig 3.3.3: System Architecture

3.3.4 UML Diagram

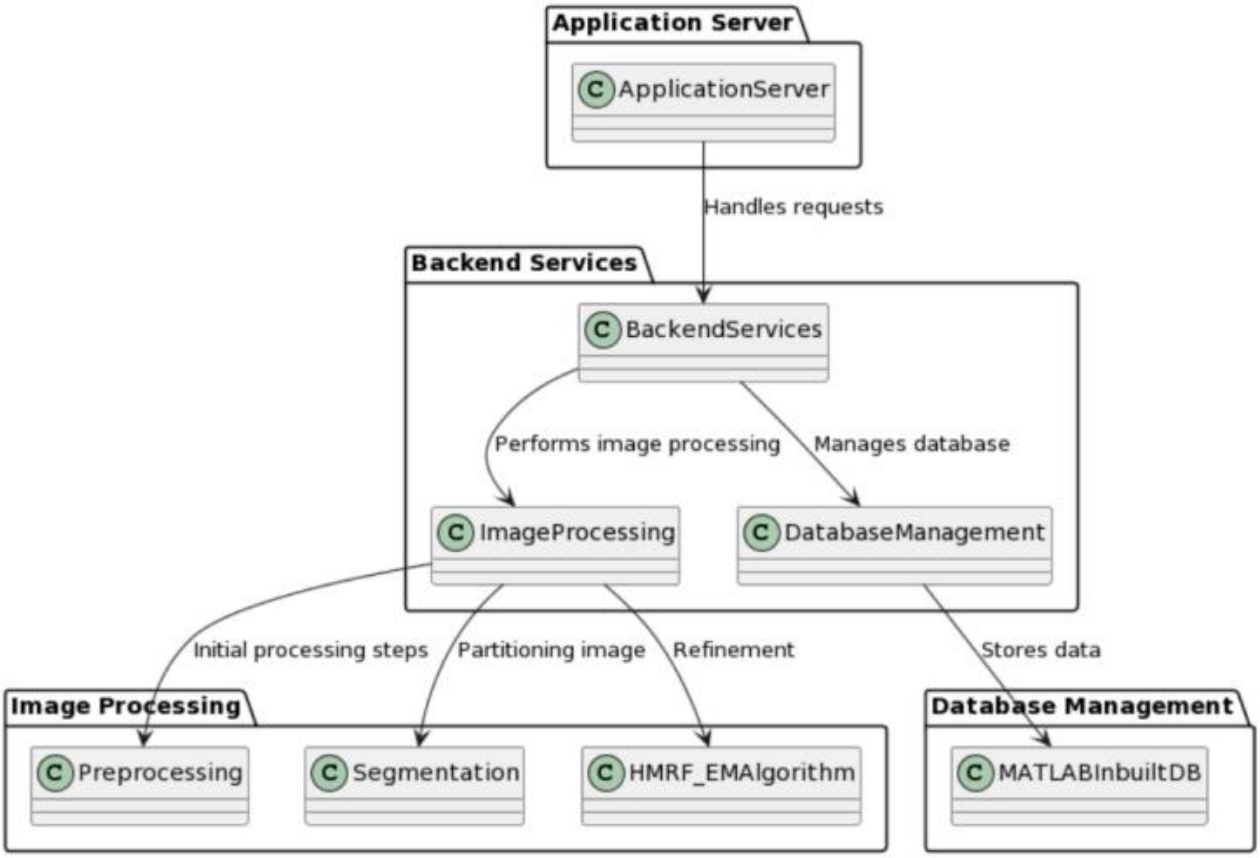


Fig 3.3.4: UML Diagram

3.3.5 Algorithm Design

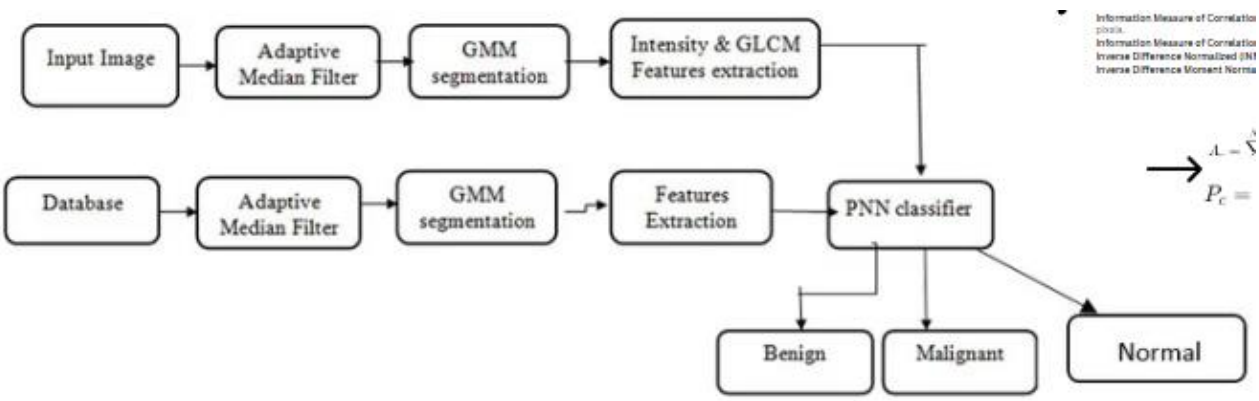


Fig 3.3.5: Algorithm Design

3.4 Software Project Management Plan

SR NO	TASK	MEMBER
1.	GUI components creation	Krisha Rawal
2.	HMRF-EM implementation	Kavya Vasa
3.	k-means segmentation	Kavya Vasa
4.	Adaptive Median Filter	Sakshi Bhosale
5.	PNN classifier & training	Kavya Vasa
6.	Acquire the datasets	Krisha Rawal
7.	GMM segmentation	Sakshi Bhosale
8.	Downloading the toolboxes	Krisha Rawal
9.	GLCM	Sakshi Bhosale

3.5 Software and Hardware Requirements

3.5.1 Software Requirements

1. MATLAB: A high-level programming language and interactive environment for numerical computation, visualization, and algorithm development.

2. M language: The programming language used within MATLAB for scripting and programming tasks.

3. Image Processing Toolbox: Provides a comprehensive set of tools for image analysis, processing, visualization, and algorithm development.

4. Statistics and Machine Learning Toolbox: Offers functions and tools for statistical analysis, machine learning, and predictive modeling tasks.

5. Signal Processing Toolbox: Contains functions and tools for analyzing, visualizing, and processing signals, such as audio and biomedical signals.

6. Neural Network Toolbox: Facilitates the design, simulation, and implementation of neural networks for pattern recognition, classification, and regression tasks.

7. Parallel Computing Toolbox: Enables parallel computing and distributed computing tasks to speed up computations by utilizing multiple processors or computing resources.

8. Operating System (Windows): The platform on which MATLAB runs, providing the underlying environment for executing MATLAB code and interacting with hardware and other software components.

3.5.2 Hardware Requirements

1. At least 8GB of RAM is recommended, but having more would be beneficial for smoother performance.
2. A multi-core processor (e.g., Intel Core i5/i7 or AMD Ryzen) would be beneficial for parallel processing tasks.
3. A high-resolution display would be beneficial for visualizing images and results effectively.
4. Digital mammography systems that capture breast images digitally. They require X-ray tubes, detectors, and image processing capabilities.

CHAPTER 4: IMPLEMENTATION AND TESTING

4.1 Functions Implemented

Major Functions:

1. guidemo: This is the main function that initializes the GUI and sets up its structure, including defining callbacks for various events.
2. guidemo_OpeningFcn: Executes just before the GUI is made visible. It initializes some default settings and data for the GUI.
3. guidemo_OutputFcn: Defines the output arguments when the GUI is invoked.
4. Browse_Callback: Callback function executed when the "Browse" button is pressed. It allows the user to select an image file.
5. AdaptiveMedianFilter_Callback: Callback function executed when the "Adaptive Median Filter" button is pressed. It applies an adaptive median filter to the selected image.
6. GMMSegmentation_Callback: Callback function executed when the "GMM Segmentation" button is pressed. It performs Gaussian Mixture Model (GMM) segmentation on the processed image.
7. Classifier_Callback: Callback function executed when the "Classifier" button is pressed. It performs classification based on features extracted from the segmented image.
8. loaddatabase_Callback: Callback function executed when the "Load Database" button is pressed. It loads a database of images for training or testing purposes.
9. TrainPNN_Callback: Callback function executed when the "Train PNN" button is pressed. It trains a Probabilistic Neural Network (PNN) using the loaded database.

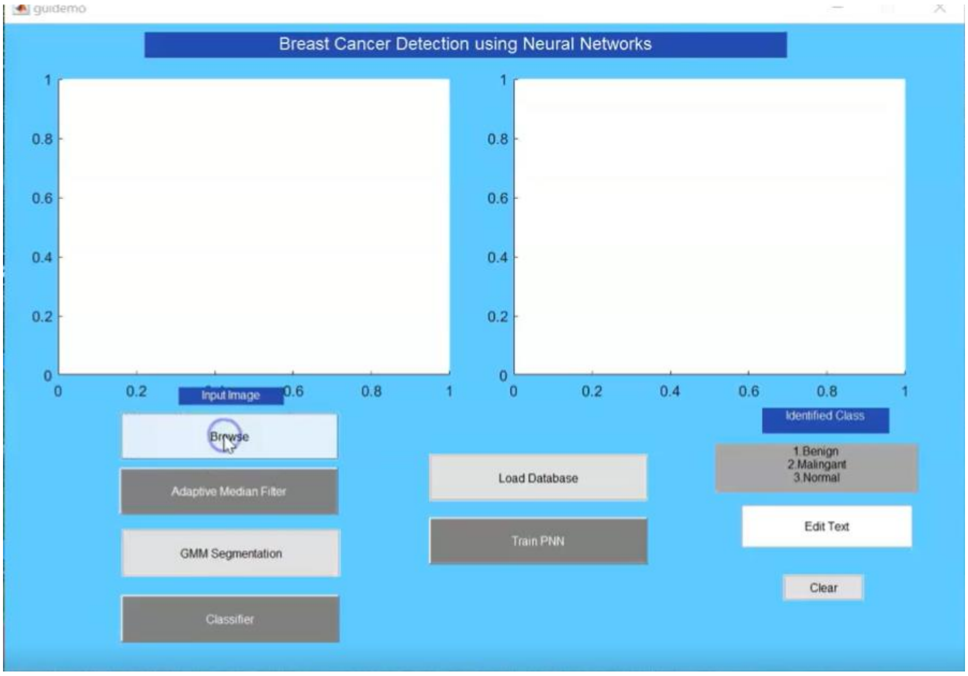
Minor Functions:

1. Various callback functions for specific buttons (`Browse`, `AdaptiveMedianFilter`, `GMMSegmentation`, `Classifier`, `loaddatabase`, `TrainPNN`).
2. Image processing functions:
   - `uigetfile`: Opens a dialog box for file selection.
   - `imread`: Reads an image file.
   - `imshow`: Displays an image.
   - `rgb2gray`: Converts an RGB image to grayscale.
   - Custom functions for adaptive median filtering, GMM segmentation, feature extraction, etc.
3. Handling of GUI components:
   - Updating handles structure using `guidata`.
   - Setting string values for `edit` components.

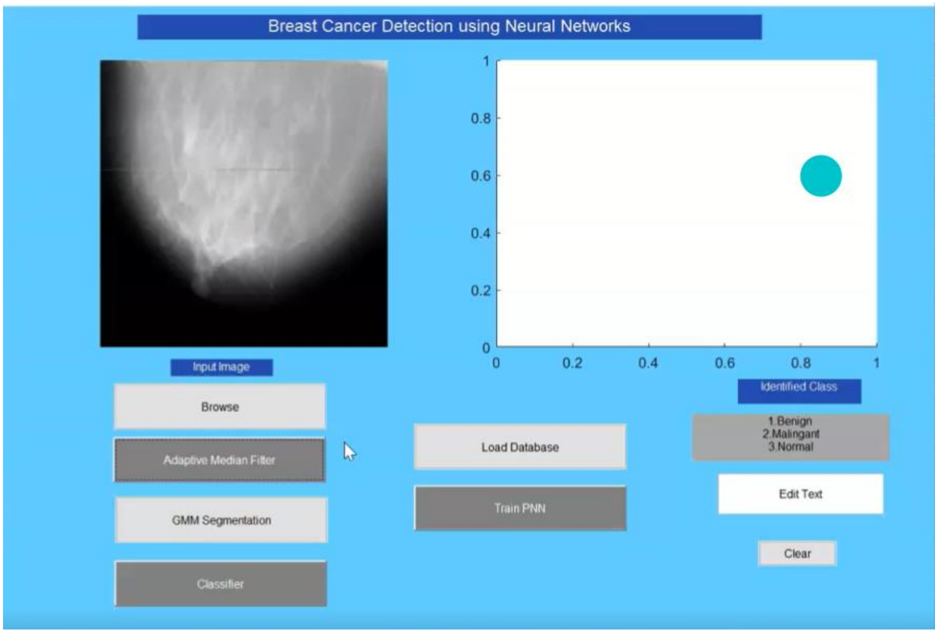
Dataset Link: https://data.mendeley.com/datasets/ywsbh3ndr8/2

4.2 OUTPUT SCREENSHOTS

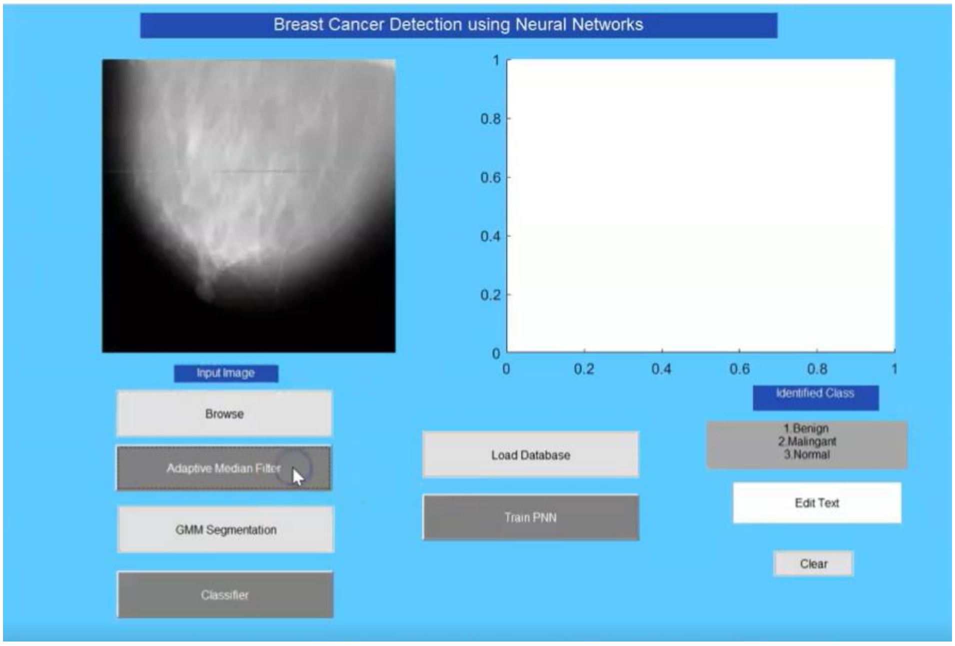
4.2.1 The front page

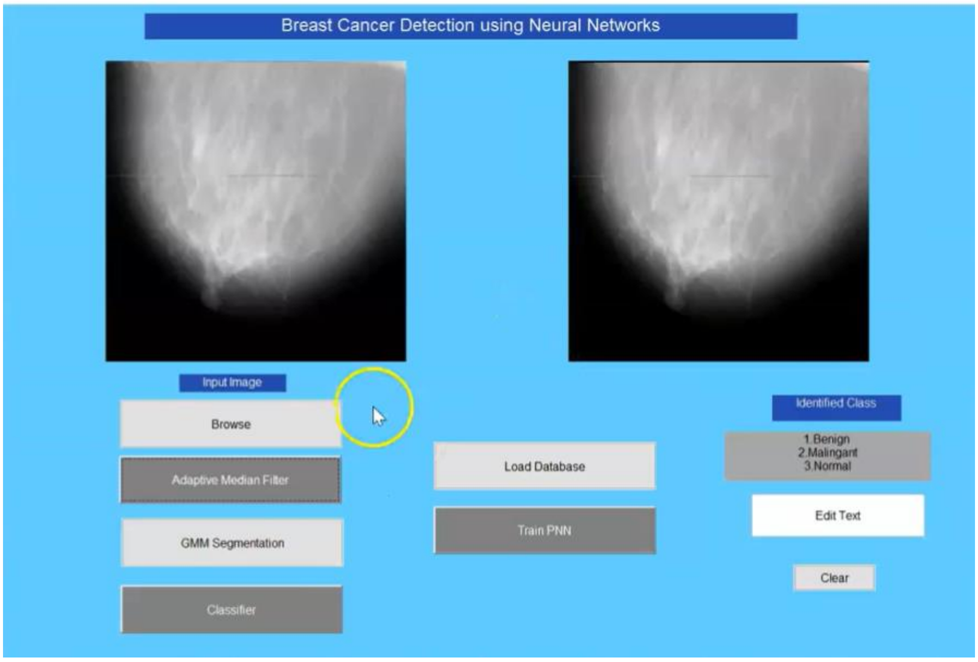


4.2.2 The mammogram is browsed

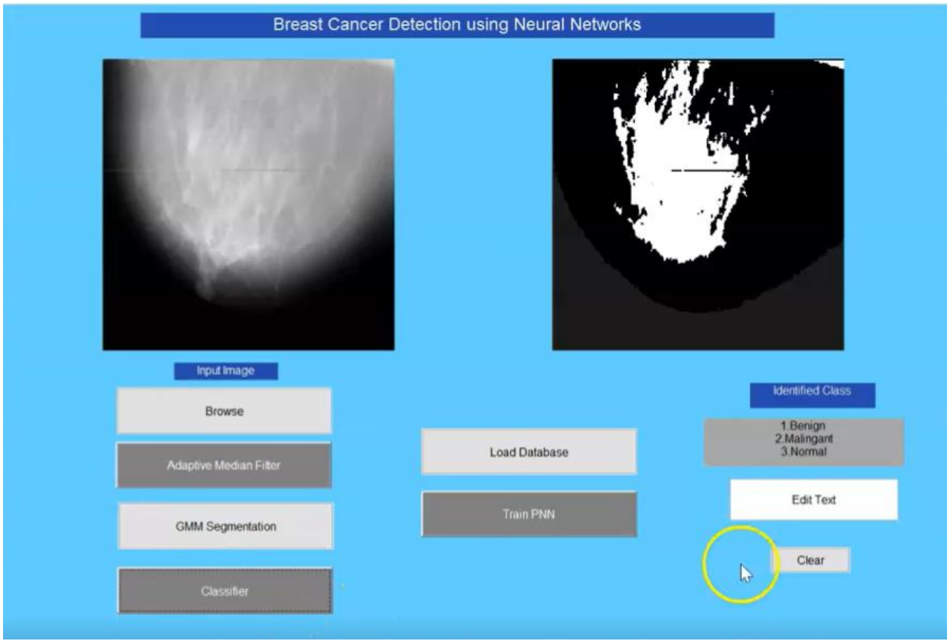


4.2.3 Adaptive Mean Filter is applied

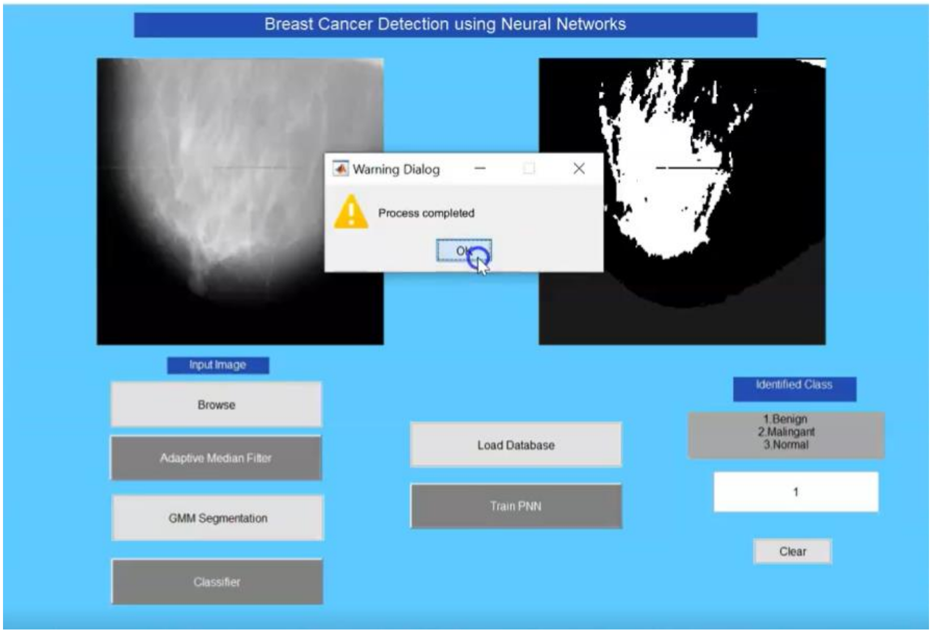




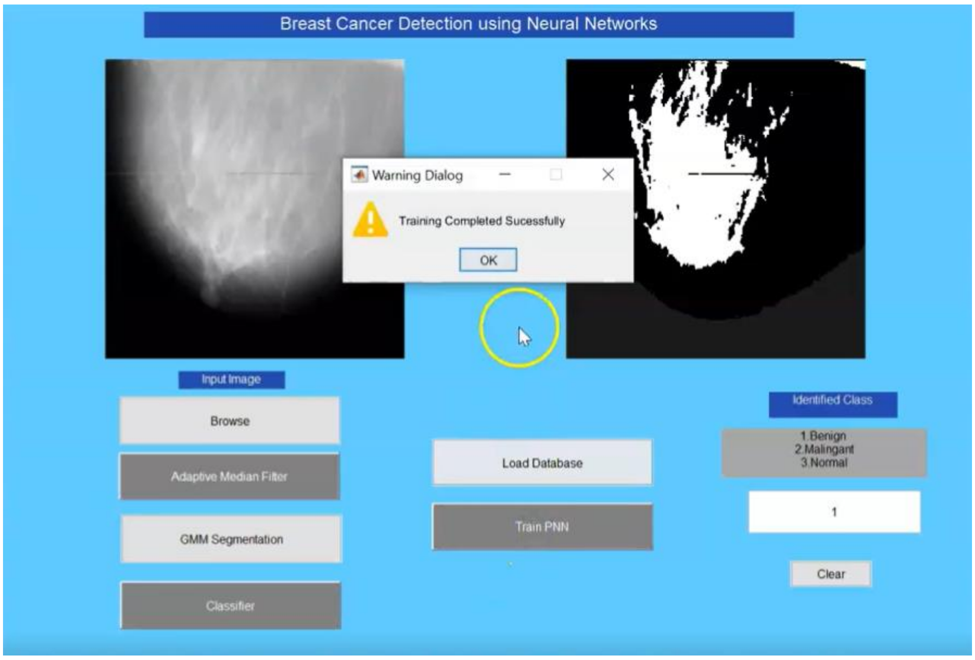
4.2.4 GMM Segmentation is applied



4.2.5 Classification of the mammogram is done



4.2.6 PNN is trained



CHAPTER 5: CONCLUSION AND FUTURE SCOPE

5.1 Conclusion

1. Automated breast cancer detection systems address the critical need for efficient and accurate interpretation of mammographic images, particularly in detecting subtle abnormalities.
2. These systems optimize healthcare resources by assisting radiologists in identifying suspicious areas on mammograms, thereby ensuring timely follow-up care for at-risk individuals.
3. In regions with limited access to healthcare professionals, automated detection systems serve as invaluable tools for providing screening and early detection services, especially in underserved communities.
4. The integration of automated systems into breast cancer screening programs enables the processing of large volumes of mammograms efficiently, facilitating timely screening for at-risk individuals.
5. By automating the detection process, these systems improve accessibility to early diagnosis, ultimately increasing the chances of successful treatment outcomes and reducing mortality rates associated with breast cancer.

5.2 Future Scope

1. Enhance Accuracy: Continuously refine algorithms to improve the accuracy of detection.
2. Integrate with Telemedicine: Incorporate systems into telemedicine platforms for remote screening and diagnosis.
3. Personalized Medicine: Utilize machine learning to tailor treatment plans based on individual patient data.
4. Multi-Modality Integration: Integrate with other imaging modalities for a comprehensive assessment of breast health.
5. Global Deployment: Scale up deployment globally to ensure equitable access to early diagnosis and improved outcomes.

REFERENCES

[1] Anuj Kumar Singh and Bhupendra Gupta “A novel approach for breast cancer detection and segmentation in mammography ” Expert System With Applications 42(2015)990-1002.
[2] Dheeba, N.Albert Singh, S. Tamil Selvi “Computer-aided detection of breast cancer on mammograms: A swarm intelligence optimized wavelet neural network approach” Journal of Biomedical Informatics (2014).
[3] “Systematic Review and Future Direction” by Maged Nasser and Umi Kalsom Yusof (2014).
[4] "Breast Cancer Diagnosis Using Deep Learning Neural Networks" by Khalid Saeed, Ghulam Muhammad, and Ajith Abraham. This paper explores the use of deep learning neural networks for breast cancer diagnosis.
[5] "Automated Breast Cancer Detection and Classification Using Deep Convolutional Neural Networks" by Anitha Julius and Bhagya Raghavan. This paper discusses the application of deep convolutional neural networks for automated breast cancer detection.
[6] "Breast Cancer Detection in Mammograms Using Convolutional Neural Networks" by R.N. Dharaskar and R. Thool. This paper investigates the use of convolutional neural networks for breast cancer detection in mammograms.
[7] "Breast Cancer Detection Using Principal Component Analysis and Probabilistic Neural Network" by A. Khokhar et al. This paper presents a method for breast cancer detection using principal component analysis and probabilistic neural network (PNN) classification.
[8] "Deep Learning for Medical Image Analysis" by S. Kevin Zhou, Hayit Greenspan, Dinggang Shen. This book covers various deep learning techniques applied to medical image analysis, including breast cancer detection.
[9] IEEE Xplore (https://ieeexplore.ieee.org/): This platform provides access to a vast collection of research papers on various topics, including medical imaging and neural networks.
[10] PubMed (https://pubmed.ncbi.nlm.nih.gov/): A resource for accessing biomedical literature, including research papers related to breast cancer detection and medical imaging.
[11] arXiv (https://arxiv.org/): A preprint repository where you can find recent research papers on breast cancer detection, neural networks, and related topics.