CHAPTER 1
INTRODUCTION

1.1 BREAST CANCER

The history of breast cancer detection is a remarkable journey from having basic methods to advanced technology. While ancient civilizations documented breast lumps, effective detection remained limited. The 19th century relied on physical examinations, but a true revolution arrived in the 20th century. X-rays, introduced in 1895, allowed doctors to see the nature of breasts for the first time. This paved the way for mammography, in the year 1976. Mammography uses low-dose X-rays to capture detailed pictures of breast tissue, revealing abnormalities potentially missed during a physical exam.

Breast cancer detection incorporates additional techniques. Ultrasound provides a closer look at suspicious areas identified on mammograms, while biopsies remain the definitive diagnostic tool. A biopsy involves removing a tissue sample for microscopic analysis to confirm the presence or absence of cancer. Mammography provides a more detailed view, reducing false positives. Emerging technologies like contrast-enhanced spectral mammography and breast MRI offer even greater accuracy, particularly for women with dense breast tissue. These advancements are leading to earlier detection, improved treatment outcomes, and ultimately, saving lives.

Mammography scans themselves are X-ray images of the breast tissue. This technology for early detection uses low-dose radiation to capture detailed pictures. While some discomfort is possible during compression, the scan itself is quick and relatively painless. Early detection through regular mammograms is crucial for successful treatment outcomes. Machine learning (ML) with Convolutional Neural Networks (CNNs) is another exciting development in breast cancer detection. Inspired by the structure of the visual cortex, CNNs excel at identifying patterns in images. By analysing mammograms, CNNs can learn to differentiate between normal and abnormal tissue with high accuracy. This technology holds significant promise for assisting radiologists in identifying subtle abnormalities they might otherwise miss, potentially leading to earlier diagnoses and improved patient outcomes for breast cancer patients.

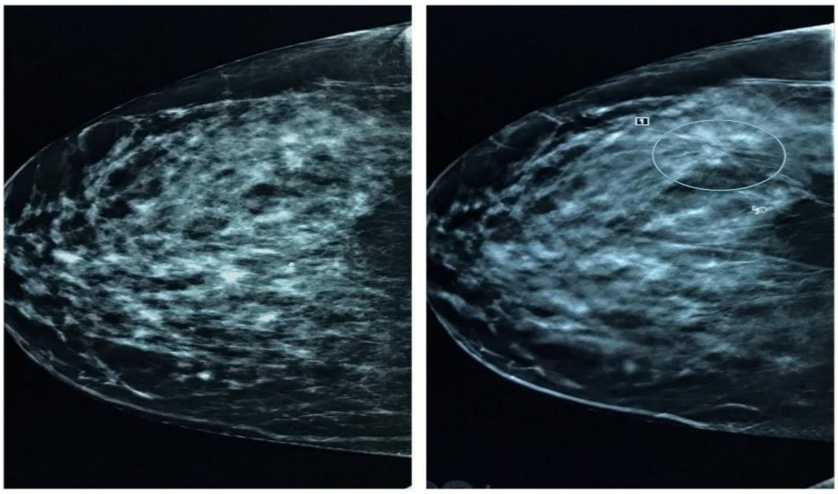

Figure 1.1 Breast Affected by Cancer

1.2 MOTIVATION

Breast cancer detection remains a critical area for improvement despite existing advancements. This motivates research into CNN-based detection projects. CNNs excel at pattern recognition in images, offering the potential to identify subtle abnormalities in mammograms that radiologists might miss. This could lead to earlier and more accurate diagnoses, a crucial factor in successful treatment. CNNs can act as a powerful secondary analysis tool, assisting in reviewing mammograms and freeing up valuable time for complex cases and patient interaction. CNNs trained on vast datasets hold promise for personalized medicine. By accounting for individual variations in breast tissue, they could pave the way for more tailored screening approaches and treatment plans. This project aligns with the ongoing pursuit of improved breast cancer detection for better patient outcomes.

1.3 IMPACT OF BREAST CANCER

Breast cancer primarily affects women and individuals assigned female at birth (AFAB) due to the presence of breast tissue. However, it's important to acknowledge that males also have the chance of getting breast cancer, although at a much lower rate. Risk factors for breast cancer can be broadly categorized as modifiable and non-modifiable. Non-modifiable factors include age, family history, and genetic mutations like Breast Cancer Gene1(BRCA1) and Breast Cancer Gene1 (BRCA2). Modifiable factors encompass lifestyle choices like diet, exercise, and hormonal influences. Understanding these factors and undergoing regular screenings are crucial for early detection and improved outcomes for everyone susceptible to breast cancer.

1.4 CONVOLUTIONAL NEURAL NETWORK

Convolutional Neural Networks (CNNs) are revolutionizing image detection tasks like breast cancer screening. Inspired by the human visual system's ability to recognize patterns, CNNs excel at analyzing images.

- Feature extraction - CNNs consist of layers that automatically learn features from images. These features start simple, like edges and shapes and become progressively more complex in deeper layers.
- Focus and Efficiency - CNNs excel at analysing mammograms because they focus on specific image regions during feature extraction. This targeted approach makes them efficient and well-suited for the task.
- Powerful Classification -The final layers of a CNN act as a classifier. They determine whether an image contains a specific feature, like a potentially cancerous tumor.

1.5 CONTRIBUTION OF CNN IN BREAST CANCER DETECTION

Convolutional Neural Networks (CNNs) offer promising benefits for breast cancer detection, potentially leading to improved patient outcomes.

- CNNs excel at pattern recognition in mammograms, potentially identifying subtle abnormalities missed by human observation. This has the potential to lead to earlier and more accurate diagnoses, a crucial factor in successful treatment.
- CNNs can act as a powerful secondary tool, assisting in reviewing mammograms and freeing up valuable time for radiologists to focus on complex cases and patient interaction.

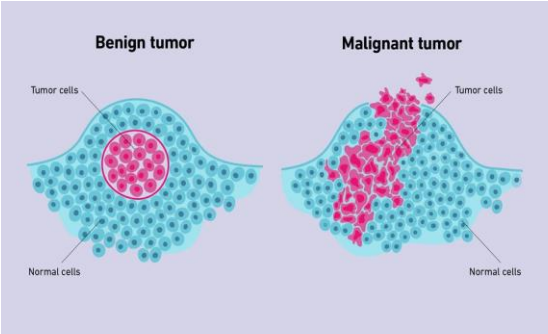

Figure 1.2 Malignant and benign

1.6 OBJECTIVES

- To develop a system that can analyze mammograms and identify signs of breast cancer with greater accuracy
- To minimize the number of mammograms flagged as suspicious when no cancer is present, reducing patient anxiety and unnecessary procedures.
- to contribute to the development of more effective breast cancer detection methods, leading to earlier diagnoses and improved patient outcomes.

CHAPTER 2
REVIEW OF LITERATURE

G. G. Naveen Kumar (2022), “Breast Cancer Detection Using Machine Learning”. The paper discussed that Breast cancer is one of the most contagious illnesses and the second leading cause of cancer mortality in women. Early breast cancer detection improves survival rates because better care may be given. Machine learning-based data categorization has been widely employed in breast cancer diagnosis and early detection. This literature review primarily focused in the categorization of accessible data using ML for breast cancer early pin-pointing and spotting. It was clear from reading multiple publications on artificial intelligence that there are several ways to detect cancer. This study aimed to compile reviews and technical publications on breast cancer diagnosis and prognosis. It provided an overview of the current research being done on various breast cancer datasets.

Devanshu Rathi, Chanchal Rathad, A. K. Raskar, Om Randhave, and Mukta Rasal (2023), “Breast Cancer Detection Using Machine Learning”. It discussed one of the biggest issues facing humanity in developing countries cancer-related mortality. Some cancer kinds still lack a cure, even though there are many strategies to stop it from occurring in the first place. Breast cancer is one of the most prevalent cancer kinds, and early detection is the key to the most crucial aspect of treatment. One of the most crucial steps in the treatment of breast cancer is an accurate diagnosis. There is a lot of research about predicting the type of breast cancer in the literature. In this study, predictions on the types of breast cancers were made using information on breast cancer tumors from Dr. William H. Walberg of the University of Wisconsin Hospital.

Yhosvany Soler Castillo (2023), “Breast Cancer Detection Technique using Machine Learning Classifiers”. The paper discussed that in the last decade, researchers working in the domain of computer vision and Artificial Intelligence (AI) have beefed up their efforts to come up with an automated framework that not only detected but also identified the stage of breast cancer. The reason for this surge in research activities in this direction is mainly due to the advent of robust AI algorithms (deep learning), the availability of hardware that can train those robust and complex AI algorithms, and the accessibility of large enough datasets required for training AI algorithms. Different imaging modalities that have been exploited by researchers to automate the task of breast cancer detection are mammograms, ultrasound, magnetic resonance imaging, histopathological images, or any combination of them.

Shahid Munir Shah, Rizwan Ahmad Khan, Sheeraz Arif, and Unaiza Sajid (2022), “Artificial Intelligence for Breast Cancer Detection: Trends & Directions.” by. This article analyzed that imaging modalities presented their strengths, and limitations, and enlists resources from where their datasets can be accessed for research purposes. This article then summarized AI and computer vision-based state-of-the-art methods proposed in the last decade, to detect breast cancer using various imaging modalities. Generally, in this article, we have focused on reviewing frameworks that have reported results using mammograms as it is the most widely used breast imaging modality that serves as the first test that medical practitioners usually prescribe for the detection of breast cancer.

Alex Krizhevsky, Ilya Sutskever, and Geoffrey E (2022) , “ImageNet Classification with Deep Convolutional Neural Networks”.The paper discussed about deep convolutional neural network to classify the 1.2 million high-resolution images in the ImageNet LSVRC-2010 contest into 1000 different classes. On the test data, we achieved top-1 and top-5 error rates of 37.5% and 17.0% which is considerably better than the previous state-of-the-art. The neural network, which has 60 million parameters and 650,000 neurons, consists of five convolutional layers, some of which are followed by max-pooling layers, and three fully-connected layers with a final 1000-way SoftMax. To make training faster, we used non-saturating neurons and a very efficient GPU implementation of the convolution operation.

María de la Luz Escobar. Ismael de la Rosa, Carlos E. Galván-Tejada, Jorge I. Galván-Tejada, Hammurabi Gamboa-Rosales, Daniel de la Rosa de la Rosa Gomez, Huitzilopoztli Luna-García, Jose M. Celaya-Padilla (2022), “Breast Cancer Detection Using Automated Segmentation and Genetic Algorithms” . The paper discussed that Breast cancer is the most common cancer among women worldwide, after lung cancer. However, early detection of breast cancer can help to reduce death rates in breast cancer patients and also prevent cancer from spreading to other parts of the body. This work proposes a new method to design a bio-marker integrating Bayesian predictive models, radiomics systems, and genetic algorithms to classify benign and malignant lesions. The method allows one to evaluate two types of images: The radiologist-segmented lesion, and a novel automated breast cancer detection by the analysis of the whole breast.

V Harvind Viswanath (2022),“Breast Cancer Diagnosis Using Machine Learning Techniques” . The paper discussed that Breast cancer is one of the most threatening diseases in women's lives; thus, early and accurate diagnosis plays a key role in reducing the risk of death in a patient's life. Mammography was the reference technique for breast cancer screening; nevertheless, many countries still lack access to mammograms due to economic, social, and cultural issues. Latest advances in computational tools, infrared cameras, and devices for bio-impedance quantification, have given a chance to emerge other reference techniques like thermography, infrared thermography, electrical impedance tomography, and biomarkers found in blood tests, therefore being faster, more reliable and cheaper than other methods.

V Harvind Viswanath, Lorena Guachi-Guachi, Saravana Prakash Thirumuruganandham (2022), “Breast Cancer Detection Using Image Processing Techniques and Classification Algorithms”. The paper discussed Breast cancer as the top cancer in women worldwide. Early detection of this disease and its classification into cases can improve the prognosis and even save lives by promoting timely clinical management to patients. An accurate diagnosis of breast cancer and its classification into benign, malignant, and normal cases is a challenge in cancer research. Because of the ability to enable the computer to learn from past samples to detect and classify patterns. However, many of them are focused on binary classification (cancer and noncancer; benign and malignant). In this work, the paper presented a Computer-Aided Diagnosis (CAD) approach for the diagnosis and classification of patients into three conditions (malignant, benign, and normal) from pixel mammogram images.

2.1 SUMMARY OF LITERATURE

Mammograms use Low-dose radiation exposure and may miss cancers in dense breast tissue. The AI algorithms such as KNN, SVM, and other rely heavily on the training data. This lack of transparency and imaging techniques used alongside physical exams, like mammograms, involve low doses of radiation, which may have long-term health risks.

CHAPTER 3
EXISTING SYSTEM

3.1 PHYSICAL EXAMINATION

During a physical breast exam for cancer detection, a healthcare professional will visually inspect and manually palpate your breasts and surrounding areas. They'll look for any changes in size, shape, symmetry, or dimpling of the breast tissue. Nipple inversion, redness, or discharge are also noted. Palpation involves using gentle pressure to feel for lumps or masses in the breast tissue and armpits (lymph nodes). The exam is typically performed with you sitting upright and then lying down, with arm positions changing to allow thorough examination. While a physical exam can help detect abnormalities, it's not a definitive test for breast cancer. If anything, suspicious is found, further imaging tests like mammograms or ultrasounds will likely be recommended.

3.2 MAMMOGRAM

The X-rays capture the internal structure of the breast, revealing potential abnormalities like lumps, masses, or calcifications. These findings don't automatically mean cancer, but they warrant further investigation. Sometimes additional mammogram views or an ultrasound are needed. Mammography is most effective for women with fatty breast tissue, as dense tissue can obscure abnormalities. While there's a small risk of radiation exposure, the benefits of early detection far outweigh this. Regular mammograms, typically starting at age 40 or earlier for high-risk women, are crucial for catching breast cancer early when it's most treatable. Mammography isn't perfect. Up to 20% of breast cancers, particularly in younger women with denser breast tissue, can be missed. Dense tissue appears white on mammograms, which can mask tumors that also show up white.

Mammograms show abnormalities that turn out not to be cancer. This can lead to additional testing, anxiety, and unnecessary biopsies. Mammography uses low-dose X-rays, but there is still a small risk of radiation exposure with repeated mammograms over time. Early detection can sometimes lead to finding slow-growing cancers that may not have caused any harm during a woman's lifetime. This can lead to unnecessary treatment and its associated side effects.

Figure 3.1 Mammography scan

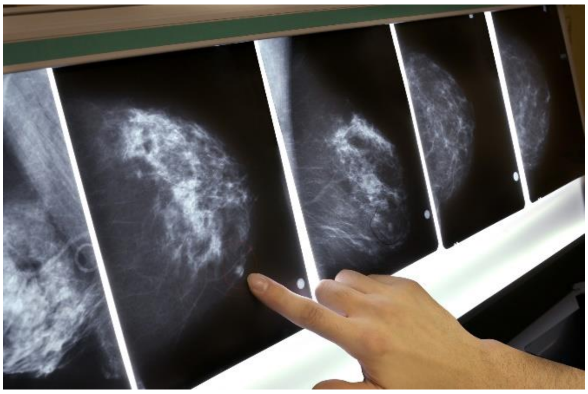

Figure 3.2 Mammography image of the breast

3.3 MAGNETIC RESONANCE IMAGING

Magnetic resonance imaging (MRI) is a detailed imaging technique used for breast cancer detection in specific situations. X-rays, MRI employs strong magnets and radio waves to create cross-sectional images of the breast tissue. Dense breast tissue can make mammograms difficult to interpret as both appear white. MRI can better differentiate between normal tissue and potential abnormalities. If a mammogram reveals suspicious findings, an MRI can offer a clearer picture to determine if a biopsy is necessary. Women with a strong family history or genetic predisposition may benefit from an MRI as a supplemental screening tool. Similar to mammograms, MRIs can reveal abnormalities that turn out to be benign. This can lead to additional testing and unnecessary anxiety. While great for overall structure, MRI might not provide the same level of detail for differentiating certain types of tissue compared to mammograms.

Figure 3.3 MRI scan

3.4 CLASSIFICATION ALGORITHMS

SUPPORT VECTOR MACHINE (SVM)

- This transforms raw image data (pixel matrices) into numerical features. This might involve extracting color information, edge details, or texture patterns.
- The SVM analyses labelled training images. It identifies a separation line (hyperplane) in the high-dimensional feature space that best divides these classes.

- The unlabelled image, the SVM uses the learned hyperplane to predict its class. Imagine the image landing on one side or the other of the decision boundary.
- SVMs can be difficult to interpret. We might not fully understand why an image is classified a certain way.
- SVM performance relies heavily on good feature selection. Poor features can significantly hurt accuracy.
- SVMs work well for moderate datasets, but training times can become lengthy for massive image collections.

K -NEAREST NEIGHBOUR (KNN)

- Images are first converted into numerical features that capture essential characteristics. This might involve extracting color histograms, edge features, or texture descriptors.
- The KNN algorithm is trained on a dataset of images that are already labelled with their corresponding classes. Each image in the training set is represented by its extracted features.
- KNN predicts the class of the new image by performing a majority vote among its k nearest neighbours. The class label that appears most frequently among the neighbours is assigned to the new image.
- As the number of features extracted from images increases, the effectiveness of KNN can decrease. Finding meaningful nearest neighbours becomes more challenging in high-dimensional spaces.

- KNN might struggle with complex image classification problems where subtle differences between classes exist.

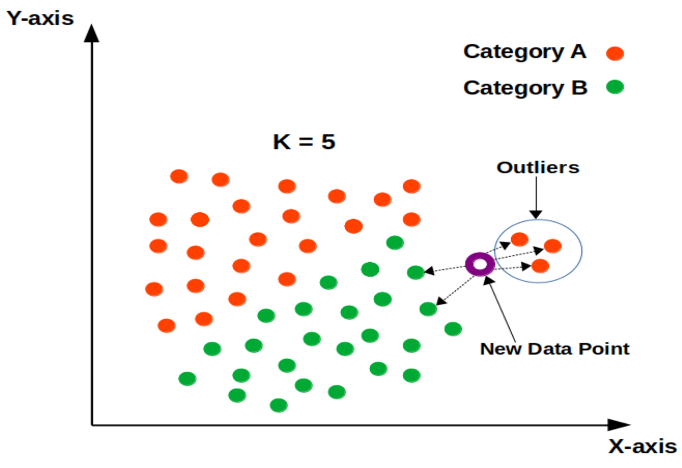

Figure 3.4 Graphical representation of KNN

CHAPTER 4
PROPOSED SYSTEM

4.1 BREAST CANCER

Breast cancer disrupts lives in many ways. Physical symptoms can include lumps, nipple changes, and skin dimpling. The emotional toll can be significant, with anxiety, fear, and uncertainty. Treatment side effects like fatigue, hair loss, and pain add another layer of suffering. Financially, medical bills and lost work can create hardship. Despite these challenges, early detection and treatment significantly improve outcomes. Support systems, counselling, and financial assistance can all help people with breast cancer navigate this difficult journey.

4.2 CONVOLUTIONAL NEURAL NETWORK (CNN)

Convolutional Neural Networks (CNNs) excel at image analysis, making them valuable tools in breast cancer detection. Imagine the CNN as a multi-layered filter. Each layer extracts features from mammograms, like shapes and textures. These features are then combined to form a higher-level understanding of the image. By analyzing millions of mammograms, CNN learns to differentiate between healthy and cancerous tissue patterns. This allows the CNN to flag suspicious areas for radiologists to review, potentially leading to earlier diagnoses and improved patient outcomes. It helps to bring better accuracy, early detection of cancer, and saving millions of lives.

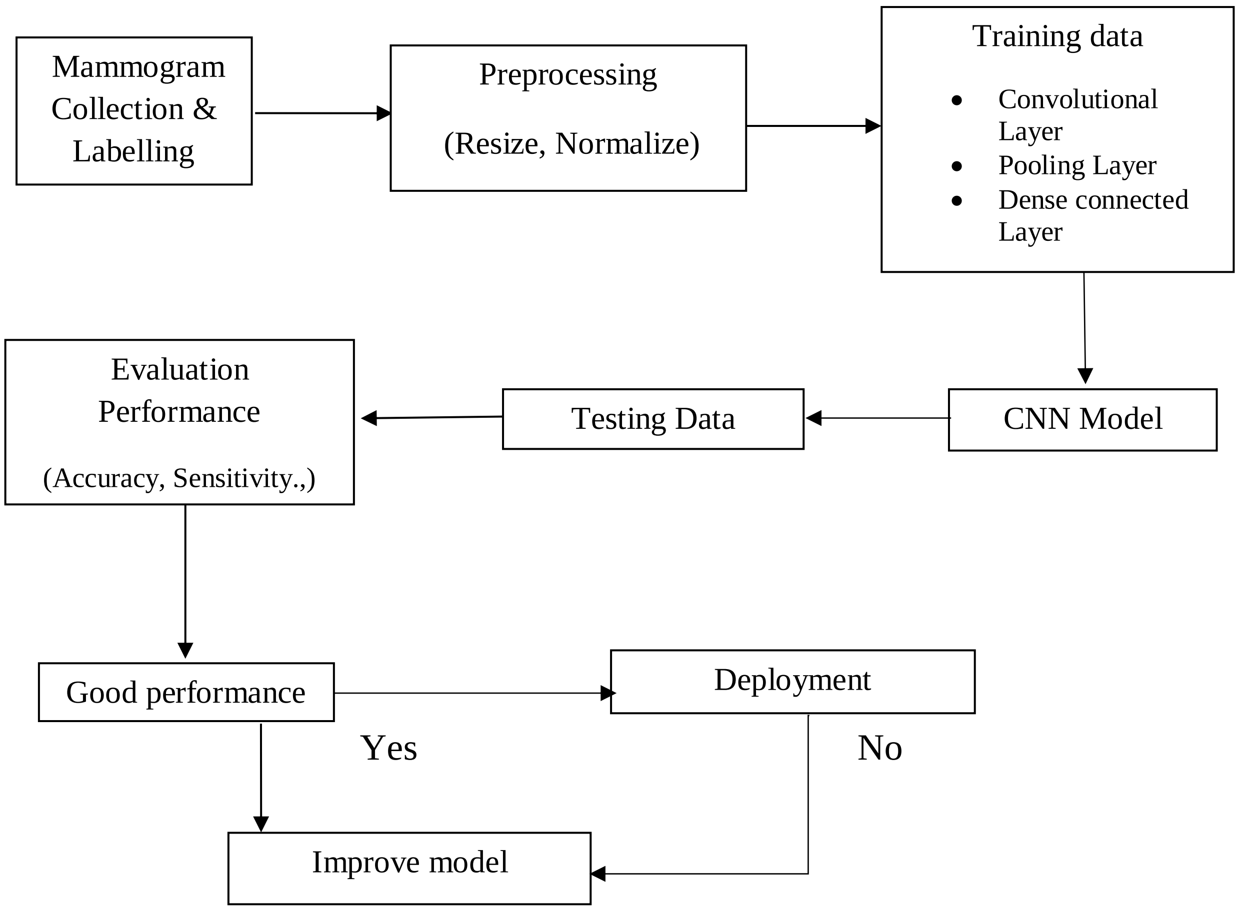

Figure 4.1 Working of CNN in Breast Cancer Detection

Images are resized, normalized, and potentially denoised for consistency. The pre-processed mammogram becomes the input layer. Convolutional layers with learnable filters analyze the image for patterns. Pooling layers summarize data and introduce some invariance. The data is transformed into a single, one-dimensional vector. Fully connected layers combine features. The model's output is compared to the known label (benign or malignant). The error is used to adjust weights and biases in the network. The CNN iteratively learns and improves its classification accuracy.

4.3 DATA COLLECTION

- Author - William H. Wolberg.
- Date- 1992.
- Title - Breast Cancer Wisconsin (Original).
- Publisher - UCI Machine Learning Repository.

4.4 PREPROCESSING

Preprocessing a breast cancer detection dataset is crucial for training effective machine learning models. Here's a breakdown of the common steps:

Data Cleaning

- Techniques like mean/median imputation or deletion can fill in missing data points.
- Identifying and potentially removing outliers that might skew the model's learning.
- Ensuring consistency in data types

Data Normalization

- Transforming features to a common scale for better comparison during training.

Data Augmentation

- Artificially creating variations of existing data points (flips, rotations) to increase dataset size and improve model generalizability.

Feature Engineering

- Choosing a subset of relevant features to improve model efficiency and potentially reduce overfitting.
- Deriving new features from existing ones to potentially capture additional information.

4.5 ALGORITHM

Data Preparation

- The mammogram images are collected and organized.
- Images are standardized for analysis (resizing, normalization).
- Techniques are applied to virtually increase the dataset size (augmentation)
- The data is then divided for training, validating, and testing the model.

Model Building

- The CNN acts like a multi-stage filter, analyzing the images layer by layer. Early layers identify basic patterns and textures.
- Later layers combine these features to detect more complex structures.
- The model learns to differentiate between healthy and cancerous tissue patterns.

Training and Refinement

- The model is trained using the labelled data (benign or malignant).
- It continuously adjusts its internal filters to improve accuracy.
- The validation set helps fine-tune the model to prevent overfitting.

Evaluation

- The model's performance is assessed on unseen data (testing set).
- Metrics like accuracy, precision, and recall measure its effectiveness

Optimization

- Different model configurations (hyperparameters) can be tested.
- This helps achieve the best possible detection accuracy.

4.6 BLOCK DIAGRAM

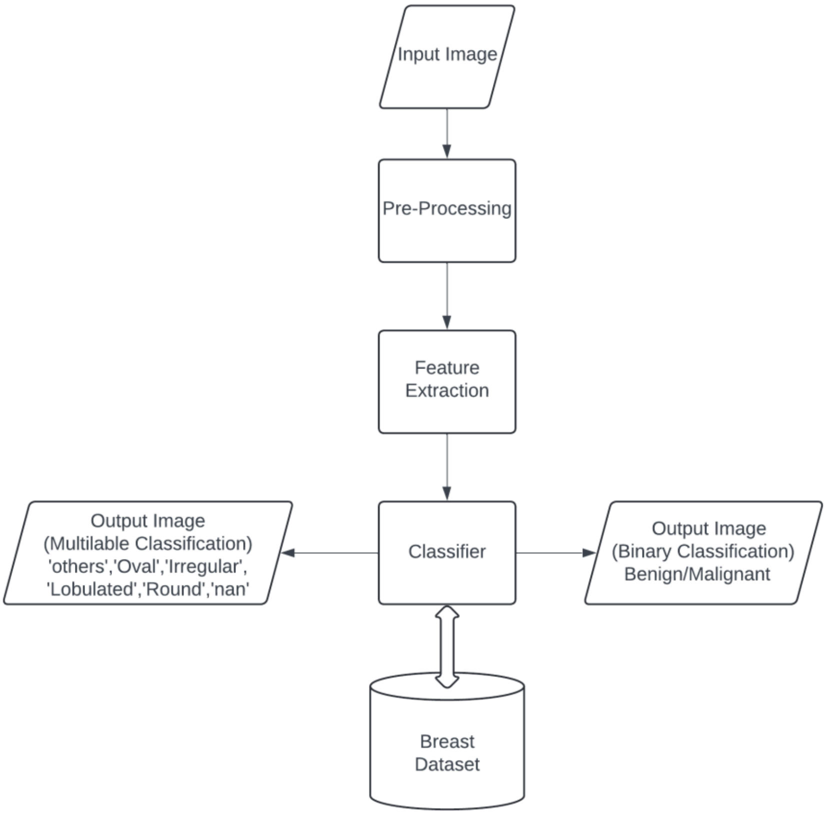

Figure 4.2 Block Diagram

Mammogram images are resized to a standard format, normalized for consistency, denoised to remove better analysis. This prepped image serves as the starting point for CNN. Convolutional layers, scan the image to identify patterns. Pooling layers then condense the information, making it more manageable and resistant to minor variations. The data is ultimately flattened into a single dimension. Fully connected layers take over, weaving together the extracted features. Finally, the model outputs a classification (benign or malignant) and compares it to the known truth. Errors guide the CNN, fine-tuning the internal connections to achieve better accuracy over repeated cycles of learning.

4.6 CODE

```
import os
import tensorflow as tf
import shutil
import pandas as pd
import sys
import matplotlib.pyplot as plt
import matplotlib.image as mpimg
import cv2
# Importing dependencies

import os
import pandas as pm

import numpy as np
import matplotlib.pyplot as plt
import cv2
from tqdm.notebook import tqdm

import tensorflow as tf
from tensorflow.keras import *
from tensorflow.keras.optimizers import AdamW
from tensorflow.keras.callbacks import *
import keras_cv


current_working_directory = os.getcwd()

# print output to the console
print(current_working_directory)

csv_path = 'C:/Users/Admin/project/dataset/csv/meta.csv'
df_meta = pd.read_csv(csv_path)
dicom_data = pd.read_csv('C:/Users/Admin/project/dataset/csv/dicom_info.csv')



image_dir = 'C:/Users/Admin/project/dataset/jpeg'
full_mammogram_images = dicom_data[dicom_data.SeriesDescription == 'full
mammogram images'].image_path
cropped_images = dicom_data[dicom_data.SeriesDescription == 'cropped
images'].image_path
```

```
roi_mask_images = dicom_data[dicom_data.SeriesDescription == 'ROI mask
images'].image_path

full_mammogram_images = full_mammogram_images.apply(lambda x:
x.replace('CBIS-DDSM/jpeg', image_dir))
cropped_images = cropped_images.apply(lambda x: x.replace('CBIS-DDSM/jpeg',
image_dir))
roi_mask_images = roi_mask_images.apply(lambda x: x.replace('CBIS-
DDSM/jpeg', image_dir))
full_mammogram_images.iloc[0]

full_mammogram_dict = dict()
cropped_dict = dict()
roi_mask_dict = dict()

for dicom in full_mammogram_images:
    # print(dicom)
    key = dicom.split("/")[6]
    # print(key)
    full_mammogram_dict[key] = dicom
for dicom in cropped_images:
    key = dicom.split("/")[6]
    cropped_dict[key] = dicom
for dicom in roi_mask_images:
    key = dicom.split("/")[6]
    roi_mask_dict[key] = dicom

mass_train_data=pd.read_csv('C:/Users/Admin/project/dataset/csv/mass_case_des
cription_train_set.csv')
mass_test_data=pd.read_csv('C:/Users/Admin/project/dataset/csv/mass_case_descr
iption_test_set.csv')
calc_train_data=pd.read_csv('C:/Users/Admin/project/dataset/csv/calc_case_descri
ption_train_set.csv')
calc_test_data=pd.read_csv('C:/Users/Admin/project/dataset/csv/calc_case_descrip
tion_test_set.csv')

def filter_dataframe_by_base_directory(df):
    base_directory = '/content/jpeg'

    # Check if all three columns start with the base directory
    mask = (
        df['image file path'].str.startswith(base_directory) &
        df['cropped image file path'].str.startswith(base_directory) &
```

```
df['ROI mask file path'].str.startswith(base_directory)
    )

    # Keep only the rows where all three columns start with the base directory
    filtered_df = df[mask]

    return filtered_df


def fix_image_path_mass(dataset):
    for i, img in enumerate(dataset.values):
        img_name = img[11].split("/")[2]
        if img_name in full_mammogram_dict:
            dataset.iloc[i, 11] = full_mammogram_dict[img_name]

        img_name = img[12].split("/")[2]
        if img_name in cropped_dict:
            dataset.iloc[i, 12] = cropped_dict[img_name]

        img_name = img[13].split("/")[2]
        if img_name in roi_mask_dict:
            dataset.iloc[i, 13] = roi_mask_dict[img_name]



fix_image_path_mass(mass_test_data)
fix_image_path_mass(mass_train_data)
mass_train = mass_train_data.rename(columns={'left or right breast':
'left_or_right_breast','image view': 'image_view','abnormality id':
'abnormality_id','abnormality type': 'abnormality_type,.'mass shape':
'mass_shape','mass margins': 'mass_margins','image file path':
'image_file_path,'cropped image file path': 'cropped_image_file_path','ROI mask
file path': 'ROI_mask_file_path'})


mass_test = mass_test_data.rename(columns={'left or right breast':
'left_or_right_breast','image view': 'image_view','abnormality id': 'abnormality_id',
'abnormality type': 'abnormality_type','mass shape': 'mass_shape','mass margins':
'mass_margins','image file path': 'image_file_path','cropped image file path':
'cropped_image_file_path','ROI mask file path': 'ROI_mask_file_path'})

shape=mass_train.mass_shape
pd.value_counts(shape)
```

```
class_counts = mass_train['pathology'].value_counts()
least = class_counts[class_counts<81].index.tolist()
mass_train['pathology']=mass_train['pathology'].apply(lambda x:'others' if x in
least else x)
print(mass_train['pathology'].value_counts())
mass_train['pathology'] = mass_train['pathology'].astype(str)

import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import MultiLabelBinarizer
from tensorflow.keras.preprocessing.image import ImageDataGenerator
from tensorflow.keras.applications import EfficientNetB0
from tensorflow.keras.layers import Dense, GlobalAveragePooling2D
from tensorflow.keras.models import Model
from tensorflow.keras.optimizers import Adam
from tensorflow.keras.callbacks import EarlyStopping

datagen = ImageDataGenerator(rescale=1./255,)

# Create training and validation data generators
train_generator =
datagen.flow_from_dataframe(dataframe=data,  x_col='image_file_path',y_col='pa
thology',target_size=(512, 512),color_mode='grayscale',  # Set color_mode to
'grayscale' for single-channel
images,batch_size=32,class_mode='categorical',subset='training')
# Model Building
num._classes = mass_test_combined['pathology'].nunique()

base_model = EfficientNetB0(weights=None, include_top=False,
input_shape=(512, 512, 1))  # Set input_shape to (512, 512, 1)

# Add custom classification layers on top of the base model
x = GlobalAveragePooling2D()(base_model.output)
x = Dense(1024, activation='relu')(x)
predictions = Dense(len(mass_train['pathology'].unique()),
activation='sigmoid')(x)  # Use sigmoid for multi-label classification

# Combine base model and custom layers into a new model
model = Model(inputs=base_model.input, outputs=predictions)

# Compile the model
```

```
model.compile(optimizer=Adam(), loss='binary_crossentropy',
metrics=['accuracy'])

history=model.fit(train_generator,epochs=20,callbacks=[early_stopping]

# Take only 5 images and their true labels
images = images[:5]
true_labels = true_labels[:5]

# Make predictions on the batch of images
predictions = model.predict(images)

# Decode predicted classes
decoded_predictions = decode_predictions(predictions)

# Convert true labels from numerical to class names
true_labels = [label_to_class[label[0]] for label in true_labels]

# Display the images along with predicted and true labels in rows
num_images = len(images)
fig, axes = plt.subplots(num_images, 1, figsize=(5, 5 * num_images))

for i in range(num_images):
    axes[i].imshow(images[i])
    axes[i].set_title(f"Predicted: {decoded_predictions[i]}\nTrue: {true_labels[i]}")
    axes[i].axis('off')

plt.tight_layout()
plt.show()
```

CHAPTER 5

RESULTS AND DISCUSSION

The Screening and diagnosis of breast cancer detection using convolutional neural networks algorithm can identify subtle patterns and abnormalities in medical images, enabling the detection of breast cancer at its earliest stages when treatment is most effective. It can achieve high levels of accuracy in classifying breast cancer lesions from mammograms, ultrasound images, or MRI scans. They can distinguish between benign and malignant tumors with great precision, reducing the chances of misdiagnosis.

- The input image dataset is provided in the form of mammogram image that is full images or masked images or cropped images.

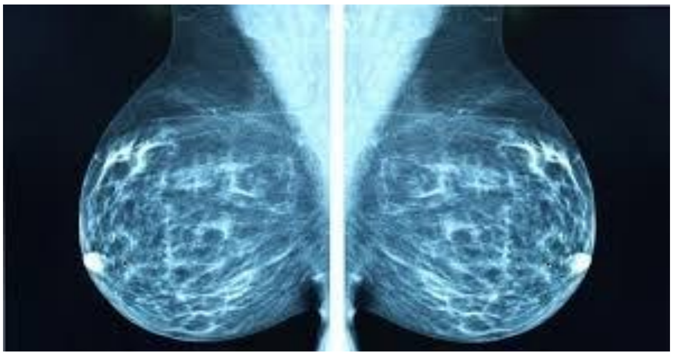

Figure 5.1 Mammogram image of the breast

The full image gives no specificity or accurate knowledge of the tumor, cropped images and masked image shows the place of the tumor Dataset provided is divided into training and testing which is 70 percent for training the model and 30 percent for testing the model.

These images undergo pre-processing, and feature extraction is carried out it compares the value with the threshold set up by the optimizer called Adam.

```
shape=mass_train.mass_shape
pd.value_counts(shape)
import pandas as pd
class_counts = mass_train['pathology'].value_counts()
least = class_counts[class_counts<81].index.tolist()
mass_train['pathology']=mass_train['pathology'].apply(lambda x:'others' if x in least else x)
print(mass_train['pathology'].value_counts())
mass_train['pathology'] = mass_train['pathology'].astype(str)
```

```
pathology
MALIGNANT                  637
BENIGN                     577
BENIGN_WITHOUT_CALLBACK    104
Name: count, dtype: int64
```

Figure 5.2 Training dataset

This line accesses the column named 'mass shape' in the DataFrame mass_train assigns it to the variable shape and counts the occurrences of each unique value in the Series shape. Occurrences of each unique value in the 'pathology' column of the DataFrame mass_train and assigns it to the variable class_counts.

The data is trained, and feature and data extraction is carried out.

	patient_id	breast_density	left_or_right_breast	image_view	abnormality_id	abnormality_type	mass_shape	mass_margins
0	P_00016	4	LEFT	CC	1	mass	IRREGULAR	SPICULATED
1	P_00016	4	LEFT	MLO	1	mass	IRREGULAR	SPICULATED
2	P_00017	2	LEFT	CC	1	mass	ROUND	CIRCUMSCRIBED
3	P_00017	2	LEFT	MLO	1	mass	ROUND	ILL_DEFINED
4	P_00032	3	RIGHT	CC	1	mass	ROUND	OBSCURED

Figure 5.3 Feature and data extraction of the patient

Breast density is a term used in breast cancer detection to describe the proportion of different types of tissue present in the breast, as visualized on mammograms.

Table 5.1 Parameter description

Levels	Breast density name	Description
1	Fatty	The breast has a high proportion of fatty tissue, making it easier to visualize abnormalities on mammograms
2	Scattered fibroglandular	There is some fibrous and glandular tissue present, but it is scattered throughout the breast.
3	Heterogeneously dense	The breast has a significant amount of fibrous and glandular tissue, which appears as dense areas on the mammogram. This density can make it more challenging to detect abnormalities, as they may be obscured by the dense tissue.
4	Extremely dense	The breast has a very high proportion of fibrous and glandular tissue, making it difficult to distinguish abnormalities from normal tissue. Mammograms of extremely dense breasts have the highest likelihood of missing small tumors

- left_or_right_breast: Indicates whether the abnormality is located in the left or right breast
- image_view: Describes the view of the mammogram is CC for cranio-caudal or MLO for medio-lateral oblique.
- abnormality_id: Identifier for each detected abnormality in the mammogram. abnormality_type: Type of abnormality detected is mass.
- mass_shape: Describes the shape of the detected mass is irregular, round, lobulated, oval or not identified (others , nan).

Table 5.2 Malignant Classification

Shape	Description
Oval	The tumour is in oval shape, it denotes a benign stage of breast cancer.
Lobulated	The tumour that starts in the milk-producing gland, or lobules, of your breast and has spread into surrounding breast tissue.
Irregular	The tumour is spread across the breast, irregular shape suggests a greater likelihood of malignancy.
Round	A round lump in tumour with a well-defined outline is, in most cases, benign.

mass margins: Describes the margins or edges of the detected mass is spiculated, circumscribed, ill defined, obscured.

Assessment score or rating of the abnormality, likely indicating suspicion level or likelihood of malignancy.
Pathology results associated with the abnormality, indicating whether it is malignant or benign.

Table 5.3 Cancer Types

Pathology	Description
Malignant	Malignant describes invasive breast cancer that has spread beyond the breast and nearby lymph nodes to other organs of the body, such as the lungs, distant lymph nodes, skin, bones, liver, or brain.
Benign	Benign describes non-invasive breast cancers or precancers. This includes the most common form of non-invasive cancer, called ductal carcinoma in situ (DCIS)

.

EfficientNetB0, GlobalAveragePooling along with sigmoid activation is carried out using 'relu' activator which is threshold-based activator used for multilabel classification. Optimizer compares the threshold with the input images, determining the loss.

Accuracy measures the proportion of correctly classified instances (both true positives and true negatives) out of the total number of instances in the dataset.

The formula for accuracy is:
Accuracy = (Number of Correct Predictions) / (Total Number of Predictions)
Mathematically, it can be expressed as:

$$\boxed{\text{Accuracy} = (TP+TN)/(TN+TP+FP+FN)}$$

where
TP (True Positives) is the number of instances correctly classified as positive (e.g., correctly identifying malignant cases as malignant).

TN (True Negatives) is the number of instances correctly classified as negative (e.g., correctly identifying benign cases as benign).

FP (False Positives) is the number of instances incorrectly classified as positive (e.g., incorrectly identifying benign cases as malignant).

FN (False Negatives) is the number of instances incorrectly classified as negative (e.g., incorrectly identifying malignant cases as benign).

Accuracy provides a general measure of the model's correctness across all classes.

- During training, the CNN processes batches of input data, computes the loss (error) between the predicted output and the actual output, and updates the network's parameters (weights and biases) using optimization algorithms Adam.
- After processing all batches of the training dataset once, the training process enters the next epoch, where it repeats the same process with the same dataset.

```
Epoch 17/20
42/42 ━━━━━━━━━━━━━━━━━━━━ 734s 17s/step - accuracy: 0.8970 - loss: 0.1761
Epoch 18/20
42/42 ━━━━━━━━━━━━━━━━━━━━ 734s 17s/step - accuracy: 0.9105 - loss: 0.1687
Epoch 19/20
42/42 ━━━━━━━━━━━━━━━━━━━━ 729s 17s/step - accuracy: 0.9343 - loss: 0.1227
```

Figure 5.4 Accuracy and loss

- This repetition continues for a specified number of epochs or until a convergence criterion is met, such as reaching a satisfactory level of accuracy or loss. The accuracy of the project is 93%

  Similar processes are carried out for model testing, thereby producing an output

Outputs

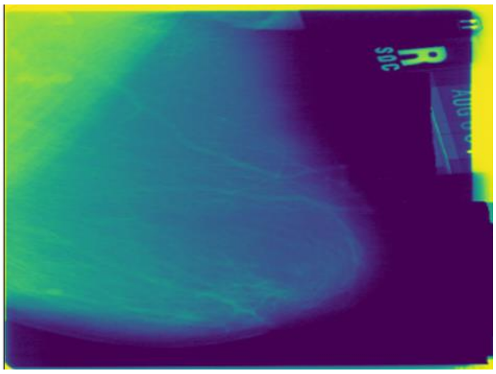

Figure 5.5 Malignant

The final predicted output is obtained "Malignant" by the model classified based on pathology, the out (Malignant / Benign) is obtained by the functions of actuator, optimizer and loss factor.

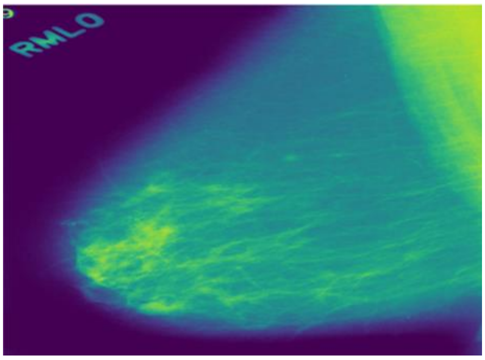

Figure 5.6 Benign

The predicted is "BENIGN" by the model comparing based on pathology ,that is threshold is compared with the trained set using optimizers the approximations are done and the final predicted output is obtained.

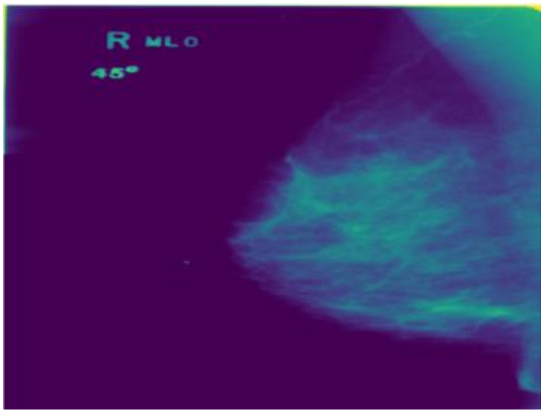

Figure 5.7 Oval

The image is compared with the reference model trained on comparing the threshold, accuracy is increased and losses are decreased providing an output as “oval”.This denotes the breast cancer is oval in shape.

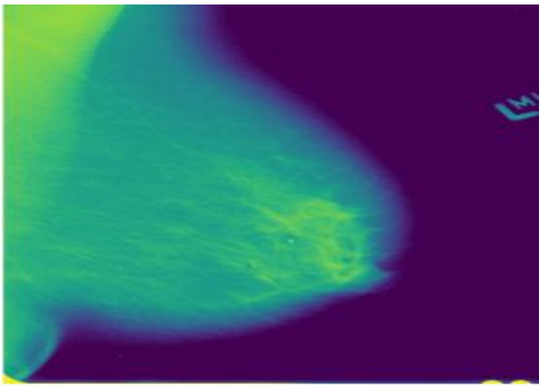

Figure 5.8 Irregular

The image is compared with the reference model trained based on mass shape, accuracy is increased and losses are decreased , providing an output as “Irregular”. This denotes the tumour is irregular in shape.

CHAPTER 6

CONCLUSION

Breast Cancer represents one of the diseases that makes highest number of deaths every year. At present, only a few accurate prognostic and predictive factors are used clinically for managing patients with breast cancer. Here, by making use of the CNN approach, high accuracy can be achieved in the detection of affected cell shapes with exact markings on detected contours. The proposed system helps to enhance the performance of mammogram retrieval by selecting optimal features. These techniques improve accuracy in tracking breast cancer cells. To assess the correctness in classifying data concerning the efficiency and effectiveness of each algorithm in terms of accuracy, precision, sensitivity, and specificity. It provides better accuracy and efficiency of cancer and protects people. "It has an accuracy of 93% with early detection."

The proposed system can be enhanced by using ensemble learning techniques for improving model robustness and generalization. Ensemble methods, such as bagging, boosting, or model averaging, can combine multiple diverse models to produce more reliable predictions and better classification of cancer.

REFERENCES

[1] Begum, H. D, A. D. S, A, and Hemanth, P. (2023), ‘Early Detection of Breast Cancer with IoT: A Promising Solution,’ Proceedings of the International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI), Chennai, India, pp. 1-6.

[2] Dandekar, A.R., Sharma, A. and Mishra, R. (2024), ‘A Deep Learning and Feature Optimization-Based Approach for Early Breast Cancer Detection,’ Proceedings of the International Conference on Electrical, Electronics and Computer Science (SCEECS), Bhopal, India, pp. 1-7.

[3] Gamal, G., Atef, H. D., Youssef, TIsmail., El-Azab (2023), ‘Early Breast Cancer Screening from Thermography via Deep Pre-Trained Edge Detection with Extreme Gradient Boosting’, Proceedings of the International Conference Intelligent Methods, Systems, and Applications (IMSA), Giza, Egypt, pp. 430-433.

[4] Gengtian. S, Bing. B and Guoyou. Z (2023), ‘EfficientNet-Based Deep Learning Approach for Breast Cancer Detection with Mammography Images’, Proceedings of the International Conference on Computer and Communication Systems (ICCCS), Guangzhou, China, pp. 972-977.

[5] Jiang *et al.*. G, (2022) ‘Integrated Photoacoustic Pen for Breast Cancer Sentinel Lymph Node Detection’, Proceedings of the International Ultrasonics Symposium (IUS), Venice, Italy, pp. 1-3.

[6] Khatri. P. V. N, Sharma. H and Shukla. P. K (2023), ‘Empowering Early Diagnosis: Optimizers Revolutionizing Breast Cancer Detection with Dense Net’, Proceedings of the International Conference on Technological Advancements in Computational Sciences (ICTACS), Tashkent, Uzbekistan, pp. 582-588.

[7] Manjunath, N., Gomathi, N., and Muthulingam, S. (2023), ‘Early Detection of Breast Cancer using Machine Learning’, Proceedings of the International Conference on Sustainable Computing and Smart Systems (ICSCSS), Coimbatore, India, pp. 165-169.

[8] Melek , A., Fakhry, S., and Basha, T., (2023), ‘Spatiotemporal Mammography-based Deep Learning Model for Improved Breast Cancer Risk Prediction’,Proceedings of the International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Sydney, Australia, pp. 1-4.

[9] Meena, L .C., Joe Prathap, P .M., and Sankara Narayanan, S., (2023), ‘Detection of Breast Cancer using Curvelet Transform and Adaptive Particle Swarm Optimization Technique’,Proceedings of the International Conference on Advanced Computing (ICoAC), Chennai, India , pp. 1-5.

[10] Neffati, S., and Machhout,M. (2023), ‘Optimized CAD System for Breast Cancer Detection with Tabu Search and RNN’, Proceedings of the International Conference on Systems and Control (ICSC), Sousse, Tunisia, pp. 20-25.

[11] Khatri, P. V. N., Sharma, H., and Shukla, P. K., (2023), ‘Empowering Early Diagnosis: Optimizers Revolutionizing Breast Cancer Detection with Dense Net’, Proceedings of the International Conference on Technological Advancements in Computational Sciences (ICTACS), Tashkent, Uzbekistan, ,pp. 582-588.

[12] Manjunath, N. Gomathi, N. and Muthulingam, S. (2023), ‘Early Detection of Breast Cancer using Machine Learning’, Proceedings of the International Conference on Sustainable Computing and Smart Systems (ICSCSS), Coimbatore, India, pp. 165-169.

[13] Melek, A., Fakhry, S., and Basha, T. (2023), ‘Spatiotemporal Mammography-based Deep Learning Model for Improved Breast Cancer Risk Prediction’, Proceedings of the International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Sydney, Australia, , pp. 1-4.

[14] Meena, L.C., Joe Prathap, P.M., and Sankara Narayanan, S. (2023), ‘Detection of Breast Cancer using Curvelet Transform and Adaptive Particle Swarm Optimization Technique’,Proceedings of the International Conference on Advanced Computing (ICoAC), Chennai, India, pp. 1-5.

[15] Neffati, S., and Machhout, M., (2023), ‘Optimized CAD System for Breast Cancer Detection with Tabu Search and RNN’, Proceedings of the International Conference on Systems and Control (ICSC), Sousse, Tunisia, pp. 20-25.