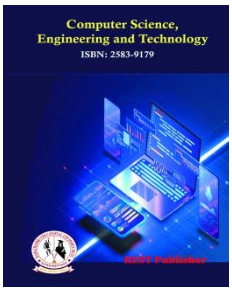
Computer Science, Engineering and Technology

Vol: 1(4), December 2023

REST Publisher; ISSN: 2583-9179 (Online)

Website: https://restpublisher.com/journals/cset/

DOI: https://doi.org/10.46632/cset/1/4/1



Understanding Breast Cancer Detection: A Comprehensive Survey

***Deepthi Rani S S, Kavya Madanan S**

Christ Nagar college Maranalloor, Arumaloor, Kerala India.

**Corresponding Author Email: deepthisiva.ss@gmail.com*

Abstract. *The potentialities of improving the penetration of millimeter waves for breast cancer imaging are here explored. Methods: A field shaping technique based on a convex optimization method is proposed, capable of increasing the field level inside a breast-emulating stratification. Results: The theoretical results are numerically validated via the design and simulation of two circularly polarized antennas. The experimental validation of the designed antennas, using tissue-mimicking phantoms, is provided, being in good agreement with the theoretical predictions. Conclusion: The possibility of focusing, within a lossy medium, the electromagnetic power at millimeter-wave frequencies is demonstrated. Significance: Field shaping can be a key for using millimeter waves for breast cancer detection. Index Terms—Breast Cancer, Microwave Imaging, Millimeter Waves, Radial Line Slot Arrays, Near-Field Optimization. A field focusing technique based on a convex optimization method is proposed, capable of increasing the field level inside a breast-emulating stratification. A focusing technique based on convex optimization has been used here to increase the penetration in breast cancer imaging scenarios at millimeter waves. Significant enhancements of the field level inside a breast model have been achieved by employing focused apertures. The key challenge in cancer detection is how to classify tumors into malignant or benign machine learning techniques can dramatically improves the accuracy of diagnosis. As the methods itself show that it's much complicated in designing and implementation. The data visualization process itself is a complicated process which can generate false results if the image has noise, so the result won't be correct. It usually contains two major steps but the further steps are much similar to the above methods ,which makes it further more complicated as it including converting patched images to real images which makes the result inaccurate in diagnosis.*

1. INTRODUCTION

The potentialities of improving the penetration of millimeter waves for breast cancer imaging are here explored. A field focusing technique based on a convex optimization method is proposed, capable of increasing the field level inside a breast-emulating stratification. The theoretical results are numerically validated via the design and simulation of two circularly polarized antennas. The experimental validation of the designed antennas, using tissue- mimicking phantoms, is provided, being in good agreement with the theoretical predictions. In this project, Field focusing can be a key for using millimeter waves for breast cancer detection As we all know that cancer emerged as one of the most dangerous and life threatening diseases in the world, it mostly effects women. So, our prime target is to use deep learning technique for diagnosis of cancer in patients especially women. The cancer or tumor in breast or any other body part starts when abnormal cells in effected body part begin to grow uncontrollably. A combination of genetic and environmental factors can be termed as the main factors for breast cancer. The genetic factors such as family historyare the factors that cannot be controlled or changed. Our main contribution of this process is image classification, tumor prediction and improvise in performance and analysis. In addition, different areas of image which exhibit variable and high appearance are characterized by various tissues. In order to extract information and also to get enhanced image the method used is commonly referred to as Image processing. Cancer is one of the dangerous disease known to humans, there are plenty of Causes for cancer it may be a gene defect or a carcinogen . Breastcancer is one of the main diseases that affect the women either she is a working woman or a housewife it doesn't matter. Detecting in the early stage and treating is the most effective way to cure it if cancer enters the second phase the only way to cure is radiotherapy or chemotherapy. Hence there is a need for low cost, portable device which can diagnose cancer effectively.

Even though there are various instruments available to diagnose breast cancer but the cost, power requirements, size and availability of such scientific instruments affecting patients with low income & patients in rural areas. Even though there are various instruments available to diagnose breast cancer but the cost, power requirements, size and availability of such scientific instruments affecting patients with low income & patients in rural areas. There are four methods for diagnosing breast cancer the first one involves a low dose of X-ray to detect the tumour it is called mammography. Ultrasonic scanning can also be used to detect the tumour, in some cases, MRI is used to diagnose the tumour and the last one is a biopsy in this method a sample of breast tissue is obtained and tested for cancer. Researchers showed that it is possible to diagnose breast cancerwith the help of an electromagnetic device [4-10]. The device measures the dielectric differencein the breast as shown in Fig. 1 and represent the difference as a series of electric field readings.But the difficulty is processing such complex data is mathematically challenging and it is harderto process the data by any conventional method known. Hence the machine learning was employed to predict the tumour size by using this complex curve data. In some works, have tried to address the same problem using image analysis techniques. The work uses the technique of noise removal and subsequent feature extraction. After the noise removal, the image is fed into classifier for further feature extraction process and finally the prediction of the cancer. Most of the earlier publications focused on feature extraction and then subsequent disease prediction was done. In some other papers, have used Artificial Neural Network for dealing with this complex problem while papers have machine learning algorithms for the task.

2. RESEARCH QUESTIONS

The project aims to find solution to these problems to the best and hopefully found out. The method used helps to generate the result with the best detection method which helps in the medical diagnosis of the disease to the best. The result had been compared using the CNN model and the result is generated. Apart from these some other research gaps like: does it can hold enormous dataset?, is it capable of storing these huge amount of data set?, whether it can hold a huge variety and varacity of dataset? Etc… These are the major concerns obtained while analysing the previous works. The aim is to find a solution and making the system capable of answering these questions.

The main objective of the project is to find answers for the following research questions:

1.How to increase the accuracy of cancer detection in humans?
2.How to increase the speed of the system?
3.How much accurate the result generated by the system?
4.Is there any chance of any other research gap?
5.What is the advantage of using convex optimization method than using other proposed methods or techniques?
6.How many types of dataset is available?
7.Does all the types of dataset can be tested?
8.Does it allow a large variety of data storage?

3. OBJECTIVES

A field focusing technique based on a convex optimization method is proposed, capable of increasing the field level inside a breast-emulating stratification. A focusing technique based on convex optimization has been used here to increase the penetration in breast cancer imaging scenarios at millimeter waves. Significant enhancements of the field level inside a breast model have been achieved by employing focused apertures. The key challenge in cancer detection is how to classify tumors into malignant or benign machine learning techniques can dramaticallyimproves the accuracy of diagnosis.

The main objective of the project is to find answers for the following research questions:

- How to increase the accuracy of cancer detection in humans?
- How to increase the speed of the system?

The project aims to find solution to these problems to the best and hopefully found out. The method used helps to generate the result with the best detection method which helps in the medical diagnosis of the disease to the best. The result had been compared using the CNN modeland the result is generated.

4. LITERATURE REVIEW

We have used the concept of transfer learning for the classification. Transfer learning (Millimeter wave detection for cancer cells[5] paper) is a popular method in computer vision that allows us to build accurate models faster. With transfer learning, instead of starting the learning process from scratch, we start from patterns that have been learned when solving a different problem.

- We split the dataset into two parts-training set and test set with 80% and 20% images respectively.
- We used data augmentation like shearing, zooming, flipping and brightness change to increase the dataset size to almost double the original dataset size.
- We tried with pre trained models like Inception v3, ResNet v2, by fine tuning the last few layers of the network
- We used 50% dropout and batch normalization layers in between to reduce over fitting.
- We used two dense layers with 64 neurons and 2 neurons respectively. The last layer is used for the classification with softmax as the activation function.
- We used binary cross entropy as the loss function.
- We trained the model for 20 epochs with a batch size of 32 by changing the hyper-parameters like learning rate, batch size, optimizer and pre-trained weights.

The methods (Millimeter wave detection for cancer cells[5] paepr) are as follows:

- Datapre-processing and image augmentation
- Feature extraction
- Image segmentation and classification
- Detection using deep learning

The Backpropagation algorithm looks for the minimum value of the error function in weight space using a technique called the delta rule or gradient descent. we are trying to get the value of weight such that the error becomes minimum. Basically, we need to figure out whether we need to increase or decrease the weight value. Once we know that, we keep on updating the weight value in that direction until error becomes minimum. You might reach a point, where if you further update the weight, the error will increase. At that time, you need to stop, and that is your final weight value. Similarly, we can select the test or query image from any of the category in an image data store. We then resize the selected input image as per pre-trained network model and features are extracted from images by using the pre-trained network and then it's corresponding category is predicted by classifier and trained features, test features and trained labels. Finally, the classifier predicts the category and the accuracy rate are calculated from confusion matrix by taking the mean value of diagonal elements of confusion matrix.

TABLE 1. CNN networks accuracy comparison

Model	**Accuracy**	**Accuracy (F- T*)**	**F-T* Effect %**	**Accuracy(max)**
Xception	0.809	0.809	±00.0	0.809
VGG16	0.809	0.809	±00.0	0.809
VGG19	0.813	0.813	±00.0	0.813
ResNet50	0.813	0.816	±00.4	0.816
InceptionV3	0.806	0.795	−01.3	0.806
InceptionResNetV2	0.812	0.582	−28.4	0.812
MobileNet	0.583	0.556	−04.7	0.583
MobileNetV2	0.781	0.656	−16.0	0.781
DenseNet121	0.795	0.778	−02.2	0.795
DenseNet169	0.569	0.642	12.8	0.612
DenseNet201	0.797	0.795	−00.2	0.797
NASNetMobile	0.799	0.802	0.4	0.802
NASNetLarge	0.813	0.809	−00.4	0.802
DenseNet201	0.797	0.795	−00.2	0.797
NASNetMobile	0.799	0.802	0.4	0.802
NASNetLarge	0.813	0.809	−00.4	0.813

Evaluation of Two relevant algorithms is as follows: In the paper Comparative Analysis of Convolution Neural Network[6] it says that it needs to be stressed that the advantage of using ML algorithms over traditional statistical models to predict a health outcome has not always been observed. For instance, a systematic review found no evidence that ML algorithms had better accuracy than logistic regression for clinical prediction modeling. Another study also found no clear difference in performance between regression models, including logistic regression and lasso regression, and ML algorithms for prognostication of traumatic brain injury. Similarly, a study demonstrated that logistic regression performed equally to ML algorithms in predicting the risk of multiple chronic diseases. Exploring potential applications of ML algorithms to the broad field of nutrition is therefore timely as we know little about their advantage over traditional statistical models.

In the paper Machine Learning Algorithms accuracy in prediction of values[7] Logistic regression and Lasso predicted adequate VF consumption with an accuracy of 0.64 (95%) confidence interval [CI]: 0.58–0.70) and 0.64 (95%CI: 0.60–0.68) respectively. Among the ML algorithms tested, the most accurate algorithms to predict adequate VF consumption were the support vector machine (SVM) with either a radial basis kernel or a sigmoid kernel, both with an accuracy of 0.65 (95%CI: 0.59–0.71). The least accurate ML algorithm was the SVM with a linear kernel with an accuracy of 0.55 (95%CI: 0.49–0.61). Using dietary intake data from the wFFQ and applying a feature selection algorithm had little to no impact onthe performance of the algorithms. In summary, ML algorithms and traditional statistical models predicted adequate VF consumption with similar accuracies among adults. These results suggest that additional research is needed to explore further the true potential of ML in predicting dietary behaviours that are determined by complex interactions among several individual, social and environmental factors.

The convolutional layers are the key component of a CNN, where filters are applied to the input image to extract features such as edges, textures, and shapes. The output of the convolutional layers is then passed through pooling layers, which are used to down-sample the feature maps, reducing the spatial dimensions while retaining the most important information. The output of the pooling layers is then passed through one or more fully connected layers, which are used to make a prediction or classify the image. CNNs are trained using a large dataset of labeled images, where the network learns to recognize patterns and features that are associated with specific objects or classes. Once trained, a CNN can be used to classify new images, or extract features for use in other applications such as object detection or image segmentation.

- By analysing the papers it’s clear that each paper differ from the methods. The methods chosen are unique and different as they use different approaches. Each process is unique in the way they had been used. The methods implemented are as follows: In the paper Deep Learning to Improve Breast Cancer Early Detection on Screening Mammography[1] it uses some deep learning techniques to analyse the cancer much more deeper. The following methods are implemented in this paper Deep Learning to Improve Breast Cancer Early Detection on Screening Mammography. To our knowledge, few papers have focused purelyon supporting the image search user experience through novel UIs. These authors described several techniques for presenting all images within a collection in a short time. Moreover, authors askedusers to think and perform browsing an image gallery and selectingan image from the gallery. These studies, showed us refinement techniques as complements in image systems with relevant user feedback. However, the presented works are limited to non-clinicalusers, making it impossible to do a generalization to our research. The result is it mainly focus on the analysis done by deep learning methods.It is much more accurate and useful and some instances but may become complicated as it goes further deeper in the analysis. Thus we focus on other methods also.

- In the second paper Breast Screening: On the Use of Multi-Modality in Medical Imaging Diagnosis[2] it uses screening method for the analysis. It uses some visualization methods so that the images can be learned and analyse much more deeper. The following methods are implemented in the paper Breast Screening: On the Use of Multi- Modality in Medical Imaging Diagnosis. The design of Breast Screening started with a qualitative study to understand radiology practices and workflow in the context of breast screening. Our study involved 31 clinicians, recruited on a volunteer basis from a large range of clinical scenarios (distinct health institutions in Portugal): 8 clinicians from Hospital Fernando

Fonseca; 12 clinicians from IPO-Lisboa; 1 clinician from Hospital de Santa Maria; 8 clinicians from IPO-Coimbra; 1 clinician from Madeira Medical Center; and 1 clinician from SAMS. Clinicians' experience ranged from 5 – 30 years of medical practice. The recruited specialists are in advanced career positions and were observed and interviewed in a semi-structured fashion. The result mainly depends on screening methods using imaging techniques. So the result is almost accurate and predictable as it focus o screening the images.

- In the paper Over-diagnosis and breast cancer screening: a case study (theconversation.com)[3] the methods uses some advanced form of visualization techniques so that the analysis would be much more deeper but it uses some complex methods. The methods implemented in the paper Over-diagnosis and breast cancer screening: a case study (theconversation.com) is as follows: From current medical imaging technologies,several issues were identified in the HCI design. Some works show the current medical imaging identification techniques for other clinical domains, where most of available systems fail to address the visual nature of the task. In these two works the authors create a visual approach to support the Mental Model development of the user. Medical imaging technologies are used to support physicians on the examination, diagnosis, and (in some cases) report. Others, study the effectiveness and performance of medical imaging systems, demonstrating how to design a user study for medical imaging experts. Further, van Schooten et al. measured user performance in terms of time taken and error rate, while interacting with the provided system. Executing it with several medical users, in this work, the authors show an experiment where their users have similar characteristics as ours.

The result is much more accurate as it focus on further more developed visualization technique but these techniques become complicated some times.

- In the paper The Utility of the Oncotype DX Test for Breast Cancer Patients in an Australian Multidisciplinary[4] Setting it focus on patients with the estrogen receptor- positive (ER+), human epidermal growth factor receptor 2-negative (HER2−) phenotype represent the most common variety of breast carcinoma, accounting for 75% of all breast cancers. The widespread use of adjuvant chemotherapy in this subtype has contributed to the reduction of breast cancer-related mortality. However, not all patients benefit from adjuvant chemotherapy. Adjuvant chemotherapy is associated with significant morbidity and expense, and its avoidance for patients identified at lower risk where there is minimal benefit is ideal. Traditional histological parameters are of variable reliability as predictors of risk in a significant proportion of ER+/HER2− patients. Genomic assays and algorithms have been developed to quantify expression of specific genes that play an important role in the recurrence risk of breast cancer and more accurately determine the likelihood of benefit of adjuvant chemotherapy.

However, the benefit of adjuvant chemotherapy remained unclear in the intermediate- risk group of patients. The prospective Trial Assigning Individualized Options for Treatment (TAILORx) involved 10,273 women with ER+/HER2−, node-negative (N0) breast cancer . Our study examined the adjuvant systemic therapy prescribing patterns and distant recurrence rates for ER+/HER2− nonmetastatic breast cancer patients in a multidisciplinary Australian institutional setting based on the Oncotype DX recurrence risk groups. While analysing the above papers it is clear that each of them differ in their techniques and way of working. So our research focus on reducing such complications to the most by implementing some techniques that reduce these complications. Our research focus on including CNN networking methods along with the visualization techniques so that the analysis and result generations are much more faster and accurate. The network technology helps in easy transformation and accessing of the required data for the comparison so that the result is more accurate to predict the medication to the most. While comparing each papers it is more clear that the techniques are implemented more complicated while compare with the above methods in our research. The above methods are easy in computing and evaluation of the results. The result can be generated much faster than them using other methods.

TABLE 2. Paper Comparisons

Sl No	Paper Name	Advantages	Disadvantages
1.	Breast Screening :On the Use of Multi-Modality in Medical Imaging Diagnosis[1]	In this paper they use a differenttypes of methods like: • Data Visualization • Clinical Workflow • Medical Imaging • Diagnostic Systems	As the methods itselfshow that it's much complicated in designing and implementation. The data visualization process itself is a complicated process which can generate false results if the image has noise, so the result won't be correct.
2.	Deep Learning to ImproveBreast Cancer Early Detection on Screening Mammography[2]	The methods in this paper include: • Converting a Classifier from recognising patches to whole images. • Network Design	It usually contains two major steps but the further steps are much similar to the above methods ,which makes it further more complicated as it including converting patched images to real images which makes the result inaccurate indiagnosis.
3.	Over-diagnosis and breast cancer screening: a case study[3]	In this paper it focus on over- diagnosis of breast cancer in women, ie; the pre-detection of cancer cells which enables in treatment of the cancer cells before they become in- treatable.	Since it focus on pre- detection techniques sometimes the post- detection of the cancer cells make them crucialin treatment process as the stage of the cancer has been changed.
4.	Artificial Intelligence(AI) for breast cancer population, Breast screen screening: based cohort study of cancer detection [4]	It focus on using the AI methods for cancer detection. Since it uses AI several algorithms so the detection of the cells is much more accurate and the treatment can be started with the proper diagnosis.	Since, it uses AI it needed to trained with the training datasets. Since the time changes the applied datasets and the detection of the cancer cells in future may be difficult and the may be difficult and the may be difficult and the result generated is not accurate. So the process is not applicable sometimes.
5.	The Utility of the Oncotype DX Test for Breast Cancer Patients in an	It focus on analyzing the results by using lab test result.The test result generation is faster and can be used anywhere and anytime of need.	The method uses testresult so sometimes it may time much more time in generating the result.Also the result can have some inaccuracy in result generation.
6.	Experimental Validation on Tissue-Mimicking Phantoms of Millimeter-Wave Imaging for Breast	It focus on how to mimic the cells more effectively.The result shows how efficiently the cells can be mimicked.	The method uses wave imaging for mimickingby identifying the damaged cell so far thisprocess cannot be accomplished successfully.

5. MACHINE LEARNING

Machine Learning is a branch of artificial intelligence that develops algorithms by learning the hidden patterns of the datasets used it to make predictions on new similar type data, without being explicitly programmed for each task. Traditional Machine Learning combines data with statistical tools to predict an output that can be used to make actionable insights. Machine learning is used in many different applications, from image and speech recognition to natural language processing, recommendation systems, fraud detection, portfolio optimization, automated task, and so on. Machine learning models are also used to power autonomous vehicles, drones, and robots, making them more intelligent and adaptable to changing environments.

Types of Machine Learning

1. Supervised Machine Learning:

Supervised learning is a type of machine learning in which the algorithm is trained on the labeled dataset. It learns to map input features to targets based on labeled training data. In supervised learning, the algorithm is provided with input features and corresponding output labels, and it learns to generalize from this data to make predictions on new, unseen data.

2. Unsupervised Machine Learning:

Unsupervised learning is a type of machine learning where the algorithm learns to recognize patterns in data without being explicitly trained using labeled examples. The goal of unsupervised learning is to discover the underlying structure or distribution in the data.

3. Reinforcement Machine Learning

Reinforcement learning is a type of machine learning where an agent learns to interact with an environment by performing actions and receiving rewards or penalties based on its actions. The goal of reinforcement learning is to learn a policy, which is a mapping from states to actions, that maximizes the expected cumulative reward over time.

6. PROPOSED SYSTEM

The potentialities of improving the penetration of millimeter waves for breast cancer imaging are here explored. A field focusing technique based on a convex optimization method is proposed, capable of increasing the field level inside a breast-emulating stratification. The theoretical results are numerically validated via the design and simulation of two circularly polarized antennas. The experimental validation of the designed antennas, using tissue- mimicking phantoms, is provided, being in good agreement with the theoretical predictions. In this project, Field focusing can be a key for using millimeter waves for breast cancer detection As we all know that cancer emerged as one of the most dangerous and life threatening diseases in the world, it mostly effects women. So, our prime target is to use deep learning technique for diagnosis of cancer in patients especially women. We have used the concept of transfer learning for the classification. Transfer learning is a popular method in computer vision that allows us to build accurate models faster. With transfer learning, instead of starting the learning process from scratch, we start from patterns that have been learned when solving a different problem.We split the dataset into two parts-training set and test set with 80% and 20% images respectively. We used binary cross entropy as the loss function. We trained the model for 20 epochs with a batch size of 32 by changing the hyper- parameters like learning rate, batch size, optimizer and pre-trained weights.

Advantage of Proposed System

- The feature extraction method is automated because of introduction of the CNN.
- A deeper neural network allows the model to learn irrespective of the data limitations.
- Use of residual networks eliminates the vanishing gradient problem.
- Pre-processing methods are very better.

- The input image beginning from the top-left corner within the borders and then count the number of cells in which the feature detector matches the input image.
- The number of matching cells is then inserted in the top-left cell of the feature map.
- Then move the feature detector one cell to the right and do the same thing. since we are moving the feature detector one cell at time, that would be called a stride of one pixel.
- The feature detector's middle-left cell with the number 1 inside it matches the cell that it is standing over inside the input image. That's the only matching cell, and so you write “1” in the next cell in the feature map, and so on and so forth.
- After you have gone through the whole first row, you can then move it over to the next row and go through the same process.
- The process of filling in a pooled feature map differs from the one we used to come up with the regular feature map.
- The neuron in the fully-connected layer detects a certain feature, thus generate a model

Existing System

In some works, have tried to address the same problem using image analysis techniques. The work uses the technique of noise removal and subsequent feature extraction. After the noise removal, the image is fed into classifier for further feature extraction process and finally the prediction of the cancer. Most of the earlier publications focused on feature extraction and then subsequent disease prediction was done. In some other papers, have used Artificial Neural Network for dealing with this complex problem while papers have machine learning algorithms for the task. Computer vision techniques have played a major role in many previous literatures.

As is evident, the publishers have utilized the image processing techniques to accomplish the pre-processing task. In the similar way we also try to implement the computer vision techniques, but out implementation mainly focuses for dataset augmentation.

Disadvantages of Existing System

- Data Limitation is big concern for training the models.
- Because of the shallower neural networks used in the model cannot learn very effective distinctions.
- Lack of the Data variety.
- Although deep neural networks give very good performance, they need lots of computational power to train networks.

Proposed System Architecture

- We split the dataset into two parts-training set and test set with 80% and 20% images respectively.
- We used data augmentation like shearing, zooming, flipping and brightness change to increase the dataset size to almost double the original dataset size.
- We tried with pre trained models like Inception v3, ResNet v2, by fine tuning the last few layers of the network
- We used 50% dropout and batch normalization layers in between to reduce overfitting.
- We used two dense layers with 64 neurons and 2 neurons respectively. The last layer is used for the classification with softmax as the activation function.
- We used binary cross entropy as the loss function.

We trained the model for 20 epochs with a batch size of 32 by changing the hyper-parameters like learning rate, batch size, optimizer and pre-trained weights.

7. MODULE DESCRIPTION

The modules that are used in the project is as follows. These modules are selected and used according to the main picture transformation for testing. The input image needed to be transformed in such a way that the resultant image is generated in such a way that it can be compared with the CNN model for identification.So for this purpose the following modules are made and selected:

Data Pre-Processing and Image Augmentation: The pre-processing of images is an important task or activity which helps in saving time for training as well as provides the clear enhancement for the further steps by increasing the efficiency of the model. Pre-processing includes the following:

- Collection of the dataset
- Shading removal
- Glare removal

The images that is taken from the dataset contains shade around the region of the cancer this shade for few images is dark and for few is light, removal of the shade in the region of cancer. Sometime the images are captured from camera the images will contain glare this glare is not visible to the naked eyes, we remove this glare using the Gaussian filters, this minute noise sometimes may affect the accuracy at the end.

Feature Extraction: The importance of feature selection in a machine learning model is inevitable. It turns the data to be free from ambiguity and reduces the complexity of the data. Also, it reduces the size of the data, so it is easy to train the model and reduces the training time. It avoids over fitting of data. Selecting the best feature subset from all the features increases the accuracy. Some feature selection methods are wrapper methods, filter methods, and embedded methods. From the relevant features for this classification, we chose the LBP color histogram and HOG color histogram. The LBP discriminator was first applied on a 3 × 3 neighborhood around a pixel and then enhanced in 2002 by extending the algorithm to any spatial neighborhood. The advantages of the LBP operator are the following: high discriminative power, invariance to lighting conditions, simplicity of calculation (allows real-time operation), and the ability to encode fine details. It is well known that the LBP algorithm encodes the local pixel information in a binary code. All the pixels from the neighborhood are compared with the central pixel. For the Histogram of Oriented Gradient descriptor (HOG), the absolute value of the gradient is calculated at the pixel level. By combining the horizontal and vertical components, it can extract the absolute value (magnitude) of the gradient and its orientation. The amplitude of the gradient indicates the intensity transitions between pixels. On the other hand, the gradient orientation represents the values of the angle at which the intensity transitions between pixels take. The LBP discriminator was first applied on a 3 × 3 neighborhood around a pixel and then enhanced in 2002 by extending the algorithm to any spatial neighborhood. The advantages of the LBP operator are the following: high discriminative power, invariance to lighting conditions, simplicity of calculation (allows real-time operation), and the ability to encode fine details. It is well known that the LBP algorithm encodes the local pixel information in a binary code. For the Histogram of Oriented Gradient descriptor (HOG), the absolute value of the gradient is calculated at the pixel level. By combining the horizontal and vertical components, it can extract the absolute value (magnitude) of the gradient and its orientation. The amplitude of the gradient indicates the intensity transitions between pixels. On the other hand, the gradient orientation represents the values of the angle at which the intensity transitions between pixels take.

Image Segmentation and Classification: Splitting operation performed on images in 2X2, 3x3 up to 10X10 patches we called it as segmentation. In this segmentation process we train to the system to identify the close regions of interest which are important to detect the BC.By eliminating unrelated data from the image, it's easy to identify the tumor as early as possible. K-mean clustering algorithm is a method of groups it means similar objects combine in same group. Segmentation operation rely on it for better results and it gives better results when similar objects present in one group. It processes fastly as compare scattered data. The objective of the support vector machine algorithm is to find a hyperplane in an N-dimensional space (N — the number of features) that distinctly classifies the data points. To separate the two classes of data points, there are many possible hyperplanes that could be chosen. The objective is to find a plane that has the maximum margin, i.e the maximum distance between data points of both classes. Splitting operation performed on images in 2X2, 3x3 up to 10X10 patches we called it as segmentation. In this segmentation process we train to the system to identify the close regions of interest which are important to detect the BC.By eliminating unrelated data from the image, it's easy to identify the tumor as early as possible. K-mean clustering algorithm is a method of groups it means similar objects combine in same group. Segmentation operation rely on it for better results and it gives better results when similar objects present in one group.

Detection Using Deep Learning: In the neural Networks we have used the Back Propagation Algorithm. The Back Propagation is a supervised learning algorithm, for training the multi-layer perceptron's. while designing the neural networks we initialize the weights with some random values as we do not know what exactly the weight can be, so we first give some random weight if the model provides an error with large values. so, we need to need to change the values to somehow minimize the error value. To generalize this, we can just say

- Calculate the error – How far is your model output from the actual output Minimum Error – Check whether the error is minimized or not.
- Update the parameters – If the error is huge then, update the parameters (weights and biases). After that again check the error. Repeat the process until the error becomes minimum.
- Model is ready to make a prediction – Once the error becomes minimum, you can feed some inputs to your model and it will produce the output

The Backpropagation algorithm looks for the minimum value of the error function in weight space using a technique called the delta rule or gradient descent. we are trying to get the value of weight such that the error becomes minimum. Basically, we need to figure out whether we need to increase or decrease the weight value. Once we know that, we keep on updating the weight value in that direction until error becomes minimum.

Admin Module: The admin view the users details and also add, delete, update both doctor's and technician's details. The admin is responsible for training the system with new available dataset so that CNNmodel matching can be done. It's not always necessary to do so. The admin add, update and delete the doctor's details based on the need.

Doctor Module: The doctor can view the details of the patient by giving his/her id details.The details include both cancer related details as well as daily routine details.Based on the details the doctor can add the prescription to the given column. The doctor can chat with other doctor's as well ass with technicians also.

Technician Module: The technician is responsible for adding the patient records.The technician is also responsible of uploading the cancer related details and images which is then accessed and viewed by the doctor.The technician add the images then the image goes under training process like segmentation of image,converting to feature extraction,generation HOG pattern etc.Then the resultant image is saved and uploaded so that the doctor can view the image and prescription is given based on it.

Use Case Diagram: In software and systems engineering, a use case is a list of actions or event steps, typically defining the interactions between a role (known in the Unified Modeling Language as an *actor*) and a system, to achieve a goal. The actor can be a human, an external system, or time. In systems engineering, use cases are used at a higher level than within software engineering, often representing missions or stakeholder goals. Another way to look at it is a use case describes a way in which a real-world actor interacts with the system. In a system use case you include high-level implementation decisions. System use cases can be written in both an informal manner and a formal manner.

Steps to create use case:

- Use Case: What is the main objective of this use case. For eg. Adding a software component, adding certain functionality etc.
- Primary Actor: Who will have the access to this use case. In the above examples, administrators will have the access.
- Scope: Scope of the use case
- Level: At what level the implementation of the use case be.
- Flow: What will be the flow of the functionality that needs to be there. More precisely, the work flow of the use case.

Some other things that can be included in the use cases are:

- Preconditions
- Postconditions
- Brief course of action
- Time Period

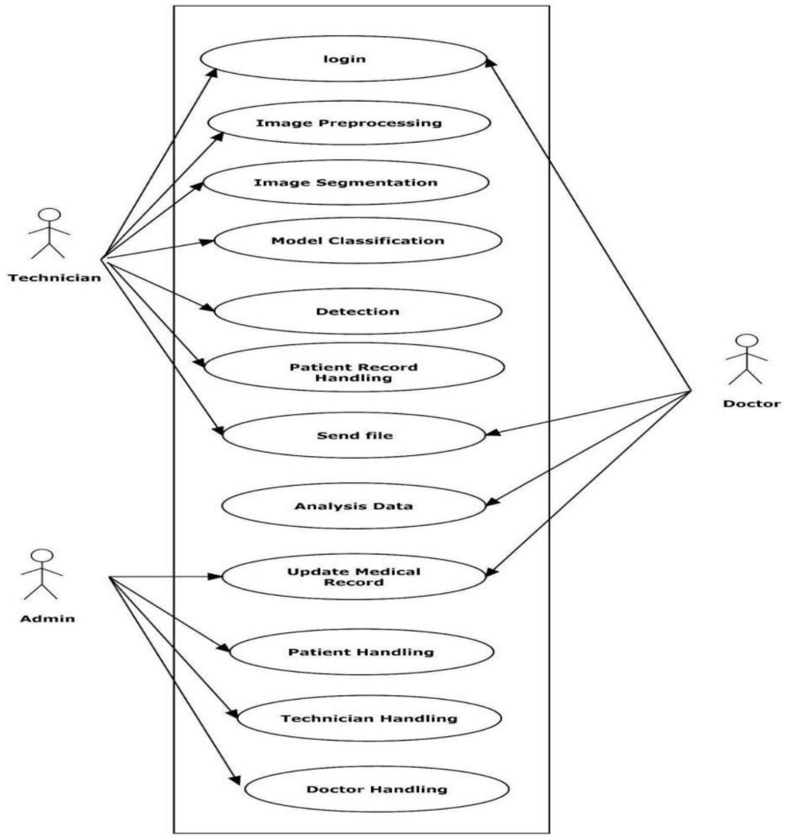


FIGURE 1. Use case diagram of the proposed system

8. RESULT ANALYSIS

To analyze the evaluation of the proposed system statistically, we perform the analysis of variance test, where the proposed system is compared with SVM Model. We propose a fully automatic breast cancer detection system. The proposed system uses U-Net network to extract the breast area from images and propose a deep learning model, which is trained for the classification of abnormal breast tissues using images. The proposed system consists of three main phases, resizing, breast area segmentation and deep learning model for classification. In resizing phase, the images are resized to a smaller size to accelerate computation. In breast area segmentation phase, the breast region is extracted automatically by using U-Net network. In deep learning model for classification phase, we proposed a two-class CNN-based deep learning model, which is trained from scratch and used for the classification of normal and abnormal breast detection.

This part of the study employed state-of-the art deep learning models using the transfer learning approach for the diagnosis of breast cancer. Breast cancer dataset, was used to further train the pre-trained deep CNN networks. For our experiment, each model was standardized with a learning rate of 0.01, a dropout of 0.5, and 38 output classes.

The experiments are carried out on a machine running Microsoft Windows 10 Pro, with an Intel(R) Core (TM) i3-6006U CPU running at 2.00 GHz, 2000 MHz, 2 cores, 4 logical processors, and 120 GB SSD, 1 TB HDD, 8 GB RAM and 26 GB Virtual Memory

The dataset consists of both healthy and diseased images for test and train the ConvNet model. The test and train data can be divided the ratio of 80:20. Here, accuracy as metrics for compilation of a model. To achieve the high accuracy and minimum loss we used the 20 epochs to train the model and to get the high accuracy. This proposed model can achieved the accuracy of 96.3% detecting the diseases in breast.

Table 3. Model accuracy and loss

Method	Validation loss	Accuracy
SVM Model	0.5421	0.8902
CNN Model	0.0929	0.9631

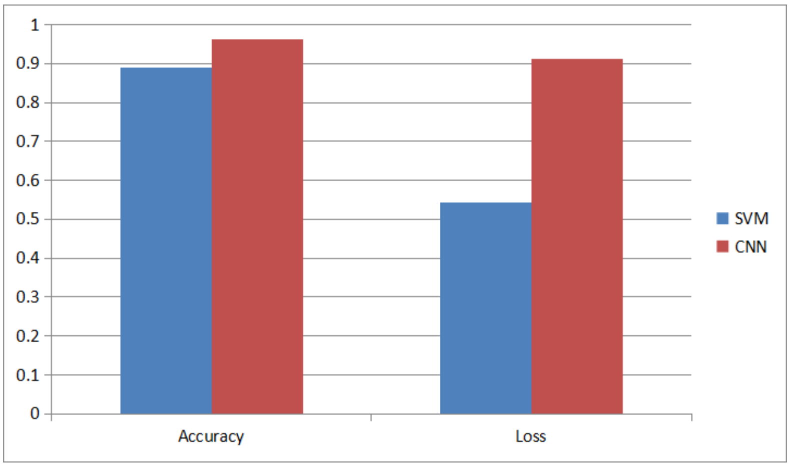


FIGURE 2. Graph showing accuracy and loss of the system

Advantages And Limitations

- Advantage of the proposed system
  - The feature extraction method is automated because of introduction of the CNN.
  - Deeper neural networks allows the model to learn irrespective of the data limitations.
  - Use of residual networks eliminates the vanishing gradient problem.
  - Pre-processing methods are very better.
  - It provides a better result analysis.
  - The result generation is accurate.
  - Better analysis for the treatment and diagnosis.
  - Easy to use and analyse.
  - Wide range of data availability.
- Limitations of the proposed system
  - Because of the shallower neural networks used in the model cannot learn very effective distinctions.
  - Although deep neural networks give very good performance, they need lots of computational power to train networks.
  - Sometimes lots of data cannot be analysed.
  - Takes some time for the training.

Future Scope: A focusing technique based on convex optimization has been used here to increase the penetration in breast cancer imaging scenarios at millimeter waves. Significant enhancements of the field level inside a breast model have been achieved by employing focused apertures, both in theory and in practice. Two CP-RLSA antennas have been designed, manufactured and measured, presenting for the first time measurements of a focused antenna in stratified media. The results are in very good agreement with the theoretical and simulation predictions, validating the overall approach. The optimized antenna presents an improvement of the field level within the breast of more than 15 dB with respect to uniformly radiating aperture and may provide higher sensitivity of imaging systems for cancer diagnostic at millimeter waves.

The methods are simply to visualize and understand. The images generated are so far clear and understanding with no noise. The images that are feeded to the system are up to date images so the comparison and generation of results will be accurate. Thus the medications and treatment will be accurate so the patient can be cured so fast.

9. CONCLUSION

The Backpropagation algorithm looks for the minimum value of the error function in weight space using a technique called the delta rule or gradient descent. we are trying to get the value of weight such that the error becomes minimum. Basically, we need to figure out whether we need to increase or decrease the weight value. Once we know that, we keep on updating the weight value in that direction until error becomes minimum. You might reach a point, where if you further update the weight, the error will increase. At that time, you need to stop, and that is your final weight value. Similarly, we can select the test or query image from any of the category in an image data store. We then resize the selected input image as per pre-trained network model and features are extracted from images by using the pre-trained network and then it’s corresponding category is predicted by classifier and trained features, test features and trained labels. Finally, the classifier predicts the category and the accuracy rate are calculated from confusion matrix by taking the mean value of diagonal elements of confusion matrix.

Finally, in order to extend the application of the millimeter-wave imaging technology to clinical use, it is required that the system be able to detect both melanoma and non-melanoma skin cancers as well as differentiate between benign and cancer tissues. So far, the performance of the UH-MMWI system has been evaluated on non-melanoma specimens. Further work involving melanoma skin cancer as well as benign lesions is needed. The potential to differentiate benign lesions from cancer tissues will result in a decrease in the number of unnecessary biopsies, saving time, effort, and patient discomfort and providing significant cost reductions for both the individual patient and the nation’s healthcare system.

Shading removal: The images that is taken from the dataset contains shade around the region of the cancer this shade for few images is dark and for few is light, removal of the shade in the region of cancer also provides us an clear vision of the cancer which is also helpful in the further enhancements.

Glare Removal: sometime the images are captured from camera the images will contain glare this glare is not visible to the naked eyes, we remove this glare using the Gaussian filters, this minute noise sometimes may affect the accuracy at the end.

Advantage of Proposed System

- The feature extraction method is automated because of introduction of the CNN.
- Deeper neural networks allows the model to learn irrespective of the data limitations.
- Use of residual networks eliminates the vanishing gradient problem
- Pre-processing methods are very better

In our research paper we mainly focus on the technical analysis of the methods implemented.

- Technical Analysis

➢ The technical analysis mainly focus on how the methods implemented can work so the result generated is so far accurate.

➢ It uses the historical data for the comparison and analysis.

➢ It mainly focus on the technical phases that is; how well the algorithm can work so it generates the result accurately and within limited time and cost.

➢ It also consider the cost of implementation of the whole system.

➢ The algorithm should have low error and low time in computation.

- Identified limitations of the related work

➢ As the methods itself show that it’s much complicated in designing and implementation. The data visualization process itself is a complicated process which can generate false results if the image has noise, so the result won't be correct.

- It usually contains two major steps but the further steps are much similar to the above methods ,which makes it further more complicated as it including converting patched images to real images which makes the result inaccurate in diagnosis.
- Since it focus on pre-detection techniques sometimes the post-detection of the cancer cells make them crucial in treatment process as the stage of the cancer has been changed.
- Since, it uses AI it needed to trained with the training datasets. Since the time changes the applied datasets and the detection of the cancer cells in future may be difficult and the result generated is not accurate .So the process is not applicable sometimes.
- The method uses test result so sometimes it may time much more time in generating the result.Also the result can have some inaccuracy in result generation.

In the neural Networks we have used the Back Propagation Algorithm. The Back Propagation is a supervised learning algorithm, for training the multi-layer perceptron's. while designing the neural networks we initialize the weights with some random values as we do not know what exactly the weight can be, so we first give some random weight if the model provides an error with large values. so, we need to need to change the values to somehow minimize the error value. To generalize this, we can just say Calculate the error – How far is your model output from the actual output Minimum Error – Check whether the error is minimized or not. Update the parameters – If the error is huge then, update the parameters (weights and biases). After that again check the error. Repeat the process until the error becomes minimum. Model is ready to make a prediction – Once the error becomes minimum, you can feed some inputs to your model and it will produce the output.

In the neural Networks we have used the Back Propagation Algorithm. The Back Propagation is a supervised learning algorithm, for training the multi-layer perceptron's. while designing the neural networks we initialize the weights with some random values as we do not know what exactly the weight can be, so we first give some random weight if the model provides an error with large values. so, we need to need to change the values to somehow minimize the error value. To generalize this, we can just say Calculate the error – How far is your model output from the actual output Minimum Error – Check whether the error is minimized or not. Update the parameters – If the error is huge then, update the parameters (weights and biases). After that again check the error. Repeat the process until the error becomes minimum. Model is ready to make a prediction – Once the error becomes minimum, you can feed some inputs to your model and it will produce the output.

REFERENCE

[1]. Deep Learning to Improve Breast Cancer Early Detection on Screening Mammography
[2]. [BreastScreening: On the Use of Multi-Modality in Medical Imaging Diagnosis
[3]. Over-diagnosis and breast cancer screening: a case study (theconversation.com)
[4]. Artificial intelligence (AI) for breast cancer screening: BreastScreen population-based cohort study of cancer detection - ScienceDirect
[5]. https://www.researchgate.net/publication/348539397_Millimeter-Waves_Breast_Cancer_Imaging_via_Inverse_Scattering_Techniques
[6]. https://www.researchgate.net/publication/348539397_Millimeter-Waves_Breast_Cancer_Imaging_via_Inverse_Scattering_Techniques
[7]. https://www.mdpi.com/2076-3417/11/1/432(Experimental Validation on Tissue-Mimicking Phantoms of Millimeter-Wave Imaging for Breast Cancer Detection)
[8]. Comparative Analysis of Convolution Neural Network
[9]. Machine Learning Algorithms accuracy in prediction of values
[10]. Marino, N., German, R., Rao, X. et al. Upregulation of lipid metabolism genes in the breastprior to cancer diagnosis. npj Breast Cancer 6, 50 (2020). https://doi.org/10.1038/s41523-020-00191-8
[11]. Lu, D., Song, J., Lu, Y. et al. A shared genetic contribution to breast cancer and schizophrenia. Nat Commun 11, 4637 (2020). https://doi.org/10.1038/s41467-020-18492-8
[12]. Ai, D., Yao, J., Yang, F. et al. TRPS1: a highly sensitive and specific marker for breast carcinoma, especially for triple-negative breast cancer. Mod Pathol (2020). https://doi.org/10.1038/s41379-020-00692- 8
[13]. F. Alsharif and C. Kurnaz, "Wearable Microstrip Patch Ultra Wide Band Antenna for Breast Cancer Detection," 2018 41st International Conference on Telecommunications and Signal Processing (TSP), Athens, 2018, pp. 1-5, doi: 10.1109/TSP.2018.8441335
[14]. T. Kikkawa and T. Sugitani, "Planar UWB antenna array for breast cancer detection," 2013 7th European Conference on Antennas and Propagation (EuCAP), Gothenburg, 2013, pp. 339-343.

[15]. D. A. Woten and M. El-Shenawee, "Broadband Dual Linear Polarized Antenna for Statistical Detection of Breast Cancer," in IEEE Transactions on Antennas and Propagation, vol. 56, no. 11, pp. 3576- 3580, Nov. 2008, doi: 10.1109/TAP.2008.2005545.
[16]. Thippesha D, Pradeep A S, Barsanoor Abhishek, Uma Angadi, Swathi S, & Vanajakshi K N. (2020). Design and Analysis of Wearable Microstrip Patch Antenna Applied for Breast Cancer Detection. International Research Journal of Engineering and Technology (IRJET), 07(08), 2175–2176.
[17]. H. Song, H. Watanabe, X. Xiao and T. Kikkawa, "Influence of Airgaps between Antennas and Breast on Impulse-Radar-Based Breast Cancer Detection," 2019 13th European Conference on Antennas and Propagation (EuCAP), Krakow, Poland, 2019, pp. 1-2.
[18]. I. Iliopoulos et al., "Enhancement of Penetration of Millimeter Waves by Field Focusing Applied to Breast Cancer Detection," in IEEE Transactions on Biomedical Engineering, doi: 10.1109/TBME.2020.3014277.
[19]. Çalışkan, R., Gültekin, S. S., Uzer, D., & Dündar, Ö. (2015). A Microstrip Patch Antenna Design for Breast Cancer Detection. Procedia - Social and Behavioral Sciences, 195, 2905- 2911. doi:10.1016/j.sbspro.2015.06.418
[20]. Top 10 CNN Architectures Every Machine Learning Engineer Should Know | by Anh

