Machine Learning and Deep Learning Techniques for Human Breast Cancer Detection

1st Manish Shah
201601041 DA-IICT
Gandhinagar, Gujarat, India
201601041@daiict.ac.in

2nd Vaibhav Sherekar
201601057 DA-IICT
Gandhinagar, Gujarat, India
201601057@daiict.ac.in

***Abstract*—In our B.Tech. project, we have applied Microwave Imaging techniques for Human Breast Cancer Detection using Machine Learning and Deep Learning methods in Python. Usage of Microwaves for Human Breast Cancer detection, is much more efficient than many primitive and current techniques. We have successfully implemented the research paper on Breast Microwave Imaging[1], presented in the European Conference and Mobile Microwave detection[3] in Python using Spyder, which explains how to efficiently detect tumour within human breasts.**

***Index Terms*—Microwaves, Breast cancer Detection, Machine Learning (ML), Support Vector Machine (SVM) , Logistic Regression, Breast Microwave Imaging (BMI)**

I. Introduction

Breast cancer in women has always been a major concern for the biology world. Breast cancer impacts nearly 2.1 million women every year. Around 15% of all cancer deaths happen due to breast cancer in women. Breast cancer is one of common cancer arising in most of the women. If detected early, chances of survival increase more than 90%. As a result, regular checkups and screenings for breast cancer are very necessary after a certain age for all women.

The techniques which are used today for breast cancer detection and screening is X-ray mammography, ultrasound, biopsy (removing a sample of breast cell for testing in the laboratory) and Magnetic Resonance Imaging (MRI). All these methods have certain drawbacks. Some of the difficulties include high cost, the need for breast compression, using ionising radiation. All these can further increase the risk of developing tumour. To overcome these difficulties a new method of Microwave Imaging (BMI) emerged. It has shown the potential of being better breast cancer detection method. BMI makes use of the scattering of incident electromagnetic pulses within the microwave range to produce the images. It makes use of inverse scattering method and back-scattered microwave data to produce results. The back-scattered data is fed into Machine Learning (ML) classifiers to detect presence and absence of tumour. This method is more accurate compared to previous systems and the set-up for this system can be easily made.

As part of our B.Tech. project, we first researched about BMI and the use of Machine Learning and Deep Learning in BMI implementation, the use of various classifiers of ML, the importance of Microwaves in tumour detection, use of neural networks in the detection of malignant and benign breast tissues. Learning all these method and techniques, finally we try to implement [1], which makes use of Logistic Regression method (which is one of the many algorithms in ML for classification) for Breast Cancer Detection using the supervised learning method. Accuracy of this algorithm to correctly detect the tumour was (85 ± 4)%, which shows the usefulness of ML in BMI cancer detection.

II. Approach and Methodology

A. *Research work*

To understand about Breast Cancer Detection using Microwaves with the help of Machine Learning, the paper on deep neural-networks[2] was very helpful. This research has been done to examine the practicality of BMI systems. It gives information about nonlinear methods and techniques used in ML. The paper also educates about the knowledge and working of neural networks, forward scattering, backward scattering and usefulness of microwaves in breast tumour detection. The paper employed certain equations to implement backward scattering method and detect tumour which is explained below in the simplest form, $A * X = B$.

where,
A = system matrix
X = dielectric constant of the materials
B = scattered waves data or matrix

The system matrix is known to us, the dielectric values are mentioned in the dataset file. From these values, we conclude B by convolution of A and X. Now, from A and B, we verify the value of X using supervised learning and see the accuracy of the ML algorithm classifier.

The thesis work by Jorge A. Sacristan [3], helps in acknowledging a new domain of mobile microwaves that are used for Breast Cancer Detection. It detects the tumor by analysing the back-scattered signal from the breast. Two different datasets were used. The first data-set was composed of signals that were generated using a frequency-domain simulation of microwaves, scattering from breast models that were made up of skin (outer layer), fatty tissues (inner layers) and tumours. This did not contain fibro-glandular tissues. The second dataset was based on models that included fibro-glandular patches of different sizes along with the components that were used in first data-set to produce breasts with densities that range from 0% to 25% in terms of fibro-glandular content grid. In this experiment 5 frequencies were used namely - 2.3 GHz, 3.35 GHz, 4.4 GHz, 5.45 GHz and 6.5 GHz, chosen to cover bandwidth supported by the antenna. This system can be employed anywhere easily.

The steps involved in this process include:

- Placing of patient's breast in-between the transmitting and receiving sensors. Here there is only one transmitting antenna and an array of receiving antennas.
- The transmitting antenna sends a microwave pulse towards the breast while the scattered signals (when reflected from the breast) are captured by the sensors and are logged into the on-board computer.
- The servo-motor rotates the chamber that holds the transmitting antenna and sensors. Steps 2 and 3 are repeated until the scattered signals have been acquired at five different rotation angles.
- An array containing the values of all scattered signals from all different angles is given as an input to a classifier that was previously trained. The classifier indicates whether if a tumour is present.

In order to test the feasibility of the proposed experimental system, Richmond's frequency-domain simulation was performed and the results were fed into a Support Vector Machine (SVM), a K-Nearest neighbours (KNN) classifier and six other classifiers for evaluation.

Forward scattering is when one predicts the scattered electromagnetic field given the geometry, find numerous engineering applications in computer-aided design. In inverse scattering, one reconstructs the physical geometry of a scatter particle/object from the measured scattered field. Hence, it finds applications in image and profile reconstructions. When using a linear method, multiple scattering within a scatter particle/object is ignored. By using a nonlinear inverse scattering method, such multiple scattering effect is accounted for. Image and profile reconstruction using such nonlinear inverse algorithm can remove artefacts that linear methods would not remove.

The paper[4] discusses convolutional neural networks. The paper gives a solution to solving inverse scattering problem using artificial neural networks as it is not effective to train the networks in order to cover different classes of devices. The paper deals with the use of neural networks in order to solve inverse electromagnetic problems to which they conclude that it is an effective strategy if the desired geometry is known. Later we can use training sets around this geometry, SNN(Spiking Neural Network) and some training sets to get the job done.

In [5], Artificial Neural Networks (ANN) has been used in combination with microscopy to visualize the 3-dimensional structure of a cell (Tumour cell). In this paper, the tomography technique is used which can reconstruct 3-dimensional images from a given sample of 2-dimensional images. Multiple scattering which is shown in Fig. 1, is one of the most challenging problems in optics, if solved, we can be able to see through fog, dark or dull (greyish) water or even human tissue to detect cancer or any anomalies.

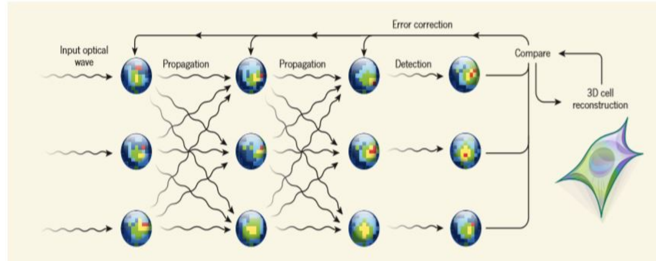

Fig. 1. Use of Artificial Intelligence algorithm to explain how the stage of optical light changes as it passes through 3D biological sample[5].

The paper[6], describes Deep Learning(DL) used as a framework to construct a 3-dimensional image. As in [5], the back-projection cannot be implemented due to its memory requirements (as a fully connected layer). Tomographic CB-CT is one of the most famous challenges in Deep Learning where they have to reconstruct images from incomplete data. This method is more reliable than the handcrafted network structure in order to decrease maximum error bounds.

The discussion about computer vision and image analysis is mentioned in [7]. It describes how image reconstruction has experienced three phases of development, they are:

- analytical methods that uses mathematical model of an imaging system (idealised model).
- iterative reconstruction.
- data-driven and learning-based methods (recent).

Microwave Imaging for early detection discusses the experimental testing of microwave imaging (MWI) system[8], which is low cost and precise for the early detection of breast cancer. The system has two components - a transmitter and a DC receiver. This system is designed to eliminate the use of costly vector network analysers.

As shown in [9] that how important is the shape and size of the breast tumour in order to get accurate results. Here, different types of breast tumours are classified into different shapes and sizes using machine learning algorithms with a system which is UWB (Ultra Wide Band) microwave imaging system.

B. Methodology

Method involved in deep neural networks[2] includes, the investigation domain denoted by D_{inv}, into which the material of interest is held, is successively illuminated by TM-polarized incident waves E(n) , n = 1,2,...,N (with n being the index of the nth illumination and N being the total number of transmitters). Both transmitters and receivers remain in the observation domain and is denoted by F and exterior to D_{inv}. For each method, the M receivers uniformly distributed over F, collect the electric fields scattered from the probed scene. For the nth method(step) and the m_{th} (m = 1,2,...,M) receiver, the scattered electrical field $E_{sca}^{(n)}$ at the location of r_m is governed by a pair of coupled equations,

$$E_{sca}^{(n)}(r_m) = k_0^2 \int_{D_{inv}} G(r_m, r')\chi(r')E^{(n)}(r')dr' \quad (1)$$

and

$$E^{(n)}(r) - E_{inc}^{(n)}(r) = K_0^2 \int_{D_{inv}} G(r, r')\chi(r')E^{(n)}(r')dr' \quad (2)$$

where r =(x, y) and r' =(x', y') denote the field and source points respectively[2], and $E^{(n)}$ represents the total electric field resulted from the interaction of probed scene with incident field $E_{inc}^{(n)}$. G(r,r') = (i/4)$H_0^{(1)}$(k_0|r - r'|) denotes a 2-D Green's function in free space, where $H_0^{(1)}$ is the first kind zeroth-order Hankel function[2].

For computational imaging, the investigation domain D_{inv} is uniformly divided into pixels such that total electric field, contrast currents and contrast functions are assumed uniform[2].

$$E_{sca}^{(n)} = G_d E^{(n)} \chi \quad (3)$$

and

$$E^{(n)} - E_{inc}^{(n)} = G_s E^{(n)} \chi \quad (4)$$

To solve (3) and (4), iterative methods are applied.

Mobile breast cancer detection method[3], includes use of 8 different classifiers namely:

- Gaussian Naive Bayes
- K- Nearest Neighbours
- Decision tree
- Random forest
- AdaBoost classifier
- Support Vector Machine (SVM) with linear kernel
- Support Vector Machine with Radial Basis Function (RBF) kernel
- Multi Layer Perceptron (MLP)

The method includes the use of a transmitting antenna and several microwave sensors. In absence of any material inside the chamber, the microwaves sensors could detect the electromagnetic fields produced by the transmitting antenna and scattered by the PVC walls of the chamber. However, as external material was introduced in the chamber, the electromagnetic field was perturbed, and the values logged by the sensors were affected. The limit to which electromagnetic field scattering occurs is caused by a difference between relative permittivity ϵ_r of the introduced medium and its surrounding medium[3]. The contrast in permittivity exhibited by healthy and malignant tissues are relevant for tumour detection purposes, as it produces electromagnetic scattering patterns that were analysed by various ML techniques mentioned above.

$$\epsilon_r = \epsilon_\infty + \frac{\epsilon_s - \epsilon_\infty}{1 + (j\omega\tau)^{1-\alpha}} + \frac{\sigma_s}{j\omega\epsilon_0} \quad (5)$$

Fig. 2, presents the permittivity values and their variation with frequencies between 1 GHz and 11 GHz. Electromagnetic scattering experiments were simulated taking the cross-section

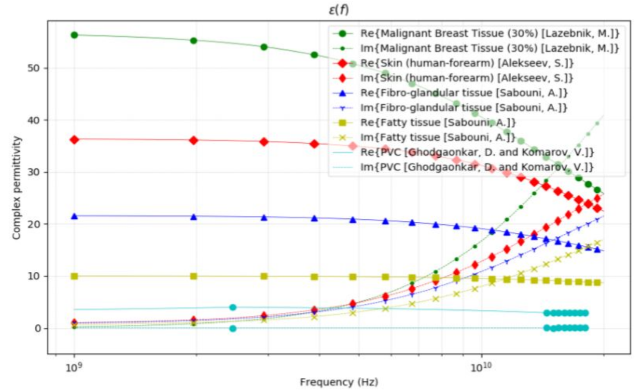

Fig. 2. Value of complex permittivity as a measure of frequency for the different tissues used to build breast models over five datasets. Solid lines indicate real part of the complex permittivity, while the dotted lines indicate imaginary part[3].

of a breast model into account, as well as the PVC walls that make up the chamber.

In breast model classification[9],uses breast tumour models that mimic the real breast using dielectric properties to create a database which consists of 13 malignant tumour models and 13 benign tumour models. Their sizes was between 13mm - 40mm. Later these models were then placed in two types of 3-Dimensional models:

- Homogeneous 3-D breast model
- 3-D breast model with some fibrograndular tissues.

Models were then placed in a 1 GHz to 6 GHz frequency range of a microwave imaging prototype system. The algorithms used to classify the breast tumour models were:

- k-nearest neighbours (kNN)
- Naive Byes (NB)
- Principal Component Analysis (PCA)
- Decision Trees (DT)

Regression analysis is a method of analysis, which examines the relation between a dependent variable; also called target and independent variable; also called predictor. In linear regression, data is developed using a straight line. While logistic regression includes probability of event represented as a linear function of combination of predictor variables. In case of logistic regression, output is the probability of occurrence of an event. In this project[1], we have implemented logistic regression for this reason.

III. Implementation

The implementation part of Mobile microwave detection[3] involves, the comparison of 11 classifiers on a 2-class synthetic dataset in Sypder Software. The output results is shown in the Fig. 3.

The BMI explained in the manuscript[1], introduces the second generation of 3D printed phantoms developed in the laboratory. The results obtained with the first generation, implies the development of the second generation of phantoms.

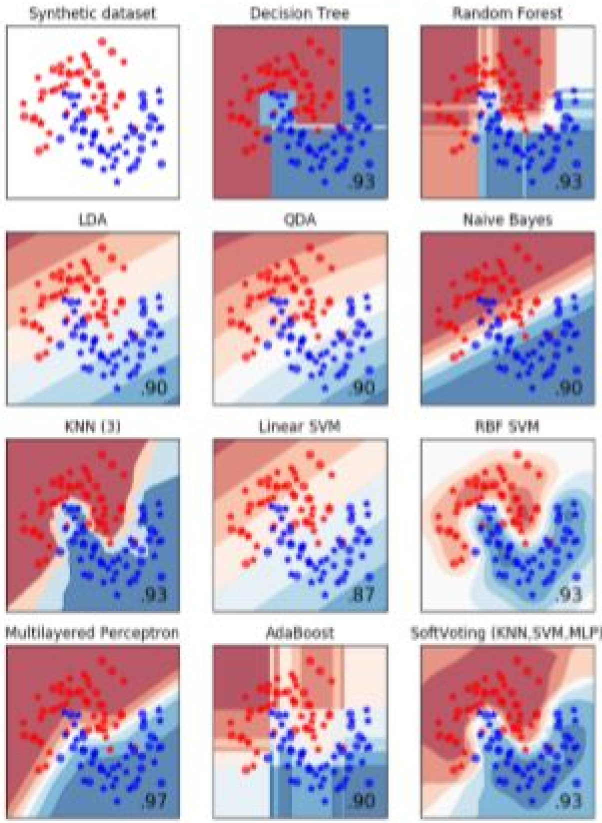

Fig. 3. Comparison result of 11 classifiers on a 2-class synthetic dataset. The upper left tile shows two classes of data (red and blue) represented by two features (x,y). Stars represent training phase and circles represent testing sets. The darkness represent confidence of the classifier. The score obtained by each classifier is represented in the lower right corner of each tile[3].

The second generation was designed to increase phantom diversity, increasing the number of unique phantoms from 13 in the first generation to 66 in the second[1]. Both generations of 3D-printable breast phantoms were developed using the MRI data provided by the University of Wisconsin-Madison. The MRI models were converted into binary models corresponding to fibroglandular and adipose components.

The first generation of phantoms contains three adipose shells and five fibroglandular shells. The second generation consisted of nine adipose shells and nine fibroglandular shells[1]. A photograph of the 3D-printed shells of the second generation of phantoms is displayed in Fig. 4(a)-4(b).

This project makes use of the frequency-domain S-parameters, the time-domain S-parameters obtained via the inverse discrete Fourier transform, and the time-domain S-parameters obtained via the inverse chirp z-transform, evaluated at 1024 time points between 0 ns and 6 ns response times performed in laboratory.

To demonstrate an initial use of UM-BMID (University of Manitoba Breast Microwave Imaging Dataset), a logistic regression classifier for tumor detection was implemented. An implementation of a logistic regression classifier was created in Python.

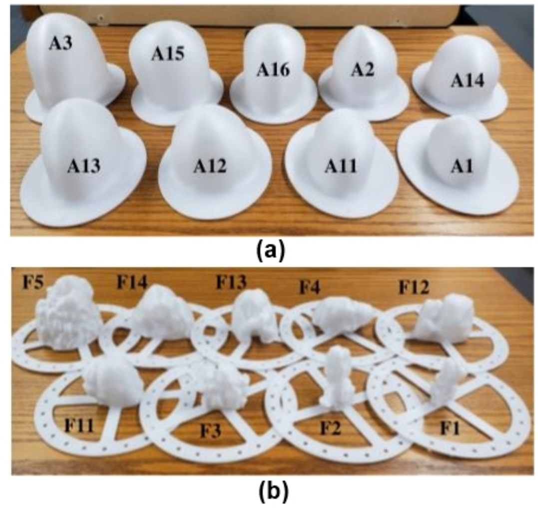

Fig. 4. Second generation of MRI derived breast shells; (a) Adipose and (b) Fibrograndular [1]

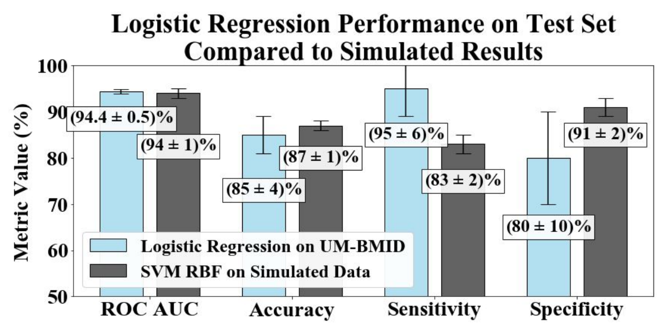

Fig. 5. Results of the logistic regression classifier on the test set (blue bars). Reported values are the averages across 10 runs, reported uncertainties are the standard deviations across 10 runs. Results by SVM-RBF on simulated BI-RADS Class II dataset (grey bars) is also displayed for comparison

For the implementation of BMI, the dataset was split randomly into training and testing sets, with 199 samples included in the training set and 50 samples included in the testing set. To reduce over-fitting, no samples in the test set had the same tumor size, tumor position, and breast phantom ID similar to training set[1]. With these conditions satisfied, the samples in the test set were randomly picked from the dataset. The logistic regression classifier was trained on the training set using gradient descent with a learning rate of 1.

IV. Nascent Theorems

After training was completed, the optimal decision threshold was selected that maximized classification accuracy on the training set. It was used to predict the labels for samples in the test set. The sensitivity,specificity,and diagnostic accuracy at this threshold were then computed. The area under the curve of the receiver operator characteristic curve (ROC AUC)was also determined. The ROC AUC, diagnostic accuracy, sensitivity, and specificity of the trained logistic classifier evaluated on the test set are displayed in Fig. 5.

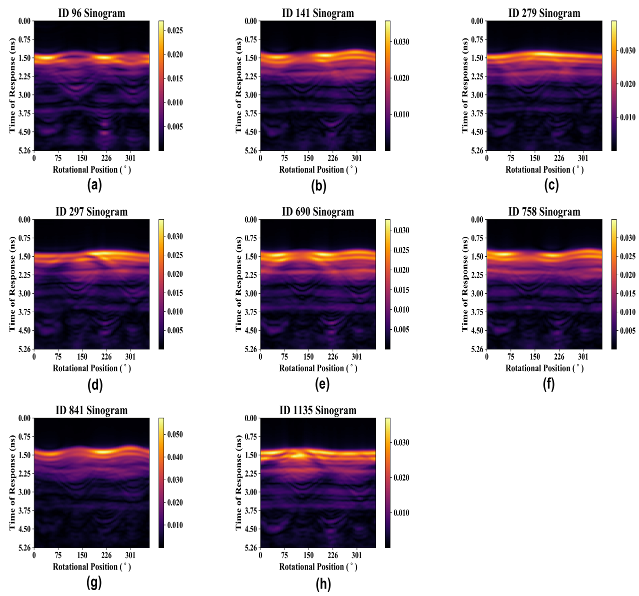

Fig. 6. Sinograms representing Time Domain representation of measured S-parameters from a phantom scan

The ROC AUC and diagnostic accuracy results obtained with the simulated results, which shows the usefulness of machine learning based classification in tumor detection with experimental data. The promising results indicate that tumor-detection via machine learning in BMI is an eventually decent approach. More advanced machine learning methods, including more sophisticated optimization routines, deep learning approaches, or the use of support vector machines, may improve classification performance.

V. Future Work

Future scope for mobile microwave breast cancer detection[3] includes,

- As the prototype is only a 2-Dimensional design which contains only one transmitting and twelve receiving sensors, future prototypes can include moving sensors that would cover the whole breast from the outer layer to the nipple. This will help because sometimes tumour is located far away from the sensors and might go undetected.
- Other option in updating the prototype would be to have several layers of sensors that will form a 3-Dimensional structure.
- In future research, one could study the relation between the position of the sensors and the classification performance in order to find the optimal sensor position.

Future work in Convolutional Neural Network (CNN) for Inverse Problem[10] solving includes,

- As CNN's are powerful and flexible there is a huge scope for it in the future. CNN could be used in Biomedical Imaging where images are formed when applied with inverse problem technique, which would be much simpler than the complicated techniques like MRI and CT.
- In future, Deep learning techniques[12] will be used along with Artificial Intelligence to improve algorithms through reinforcement learning.
- In making 3-D phantoms of the breast[11], the liquid that is filled in the breast model will need further investigation and it is likely that a more optimal fluid will be used to get accurate results.
- As Microwave Imaging has taken over X-ray, Microwave Imaging has some limitations like: to develop a microwave sensor with high sensitivity and limited resolution of an image which would reduce the cost of the system and improve the image resolution.
- Finally, in the future, while making 3-D Phantom models of breasts, different Telomere Maintenance Mechanism(TMM) should be also considered in improving the accuracy to locate the tumour.

VI. Conclusion

This project work explains the use and effectiveness of Microwaves in Breast Cancer Detection using Machine Learning and Deep Learning techniques. The method of BMI and its improvised performance over certain methods of the same kind is notable. The project helps to learn about various ML classifiers useful for breast tumour detection, determine which suits them best and analyse results. In addition, learning to use CNN, DL and ML methods for breast cancer detection is worthwhile. Various types of ML methods-supervised, unsupervised and reinforcement learning, and getting most out of it, is understood from this project. The dataset used in the project inhibits the scope for the implementation of techniques and the effects of breast size and density, tumor size and position on the methods in BMI. The performance and accuracy by logistic regression classifier in cancer tumour detection are promising. The project can be used as a reference for other future works.

Acknowledgement

We would like to thanks Prof. Deepak Ghodgaonkar for his guidance and support throughout the project. We would also like to thanks Dr. Hardik Patel for his assistance and support.

Contribution

During our BTP, we have done following works of implementation and contributed in the following manner:

- We have successfully implemented Machine Learning classifiers comparison on 2-class synthetic dataset for Mobile Microwave Breast Cancer Detection[3] and have represented the output in Fig. 3 and have worked diligently to implement the complete thesis work.
- We have also implemented the work of Breast Microwave Imaging[1], using the dataset available from Wisconsin-Madison and UM-BMID. We have implemented Logistic Regression Machine Learning classifier for this purpose and have shown my comparison results, accuracy to the previously implemented work versions and output

in sinograms in Fig. 5 and Fig. 6 respectively. We have implemented all work in python using Spyder software.

- Apart from this, we have collected the information needed for this project from the journal papers, research papers, online materials, IEEE papers and conference papers, papers from the BTP mentor as well as from all the available and possible sources.
- To gain knowledge and have practical experience of Machine Learning, we have completed two online courses: one by Andrew NG on Coursera which provides exposure to theoretical information and practical experience in ML, and second by Edureka on YouTube, which familiarize with tools for ML like Anaconda, Spyder, Jupyter Notebook and acknowledge about Artificial Intelligence (AI), ML and DL.
- In addition to all these, we have also implemented several other works and projects based on Machine Learning for better understanding of various ML algorithms which are helpful in Breast Tumour Detection.

REFERENCES

[1] T. Reimer, J. Krenkevich, and S. Pistorius, "An open-access experimental dataset for breast microwave imaging,", 2020 European Conference on Antennas and Propagation (EuCAP 2020), accepted.

[2] Lianlin Li, Long Gang Wang, Fernando L. Teixeira, Che Liu, Arye Nehorai and Tie Jun Cui, "DeepNIS: Deep Neural Network for Nonlinear Electromagnetic Inverse Scattering", IEEE Transactions on Antennas and Wireless Propagation Letters, vol. 67, No. 3, March 2019.

[3] Jorge A. Sacristán, "Preliminary Study of a Mobile Microwave Breast Cancer Detection Device Using Machine Learning", A thesis submitted to The Faculty of Graduate Studies of The University of Manitoba, Dec 2017.

[4] S. Ratnajeevan H. Hoole, "Artificial Neural Networks in the Solution of Inverse Electromagnetic Field Problems", IEEE Transaction on Magnetics, vol. 29, No. 2, March 1993.

[5] Laura Waller and Lei Tian, "Computational imaging: Machine learning for 3D microscopy", Article in Nature, vol 523, pp. 416-417, July 2015.

[6] Tobias Wurfl, Mathis Hoffmann, Vincent Christlein, Katharina Breininger, Yixing Huang, Mathias Unberath and Andreas K. Maier, "Deep Learning Computed Tomography: Learning Projection-Domain Weights from Image Domain in Limited Angle Problems", IEEE Transaction on Medical Imaging, 2018.

[7] Ge Wang, Jong Chu Ye, Klaus Mueller and Jeffrey A. Fessler, "Image Reconstruction Is a New Frontier of Machine Learning", IEEE Transaction on Medical Imaging, vol. 37, No. 6, June 2018.

[8] J. A. Tobon Vasquez, F. Vipiana, M. R. Casu, M. Vacca and A. Pulimeno, "Microwave Imaging for Early Breast Cancer Detection: Experimental Testing of a Low-Cost Portable System", Microwave Imaging for Combined Early Diagnostics of Breast Cancer (MICENEA) FIRB project, IEEE, March 2016

[9] Raquel C. Conceicaol, Hugo Medeiros, Daniela M. Godinho1, Martin O'Halloran, Diego Rodriguez-Herrera, Daniel Flores-Tapia, and Stephen Pistorius, "Classification of Breast Tumour Models with a Prototype Microwave Imaging System", Accepted Article, American Association of Physicists in Medicine, Feb. 2020.

[10] Michael T. McCann, Kyong Hwan Jin and Michael Unser, "A Review of Convolutional Neural Networks for Inverse Problems in Imaging", IEEE, Oct 2017.

[11] Matthew J. Burfeindt, Timothy J. Colgan, R. Owen Mays, JacobD. Shea, Nader Behdad, Barry D. Van Veen, and Susan C. Hagness, "MRI-Derived 3-D-Printed Breast Phantom for Microwave Breast Imaging Validation", IEEE Transactions on Antennas and Wireless Propagation Letters, vol. 11, Dec. 2012.

[12] Danna Voth, "Using AI to Detect Breast Cancer", In The News, Published by the IEEE Computer Society, Jan. 2005.

[13] Branislav Gerazov and R.C. Conceicao, "Deep learning for tumour classification in homogeneous breast tissue in medical microwave imaging", Research Gate Conference Paper, July 2017.