Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 10 Issue 03, March-2021

Prediction of the Malignant Tumour Size in Breast Cancer with the Aid of Machine Learning

Pramodh B R
Associate Data Annotator
Tika Data services
Bengaluru, India

Thippesha D
Associate Data Annotator
Tika Data services
Bengaluru, India

***Abstract*— Breast cancer is one of the fatal diseases which affects the women regardless of class, from faulty gene to the carcinogen there multiple reasons for it. Detecting in the early stage and treating is the most effective way to cure it. In this article the machine learning is used to predict the malignant tumour size employing WEKA and Python on the electric field data obtained from a wearable antenna, the obtained result is compared with each other. The machine learning algorithms are tested on different sets of data to determine accuracy, error, and performance.**

Keywords*— Breast Cancer, Malignant Tumour, WEKA, Python, Machine Learning, Antenna, Random Tree, Random Forest, XG Boost, Support Vector Regression, Decision Tree, Multi-Layer Perception.

I. INTRODUCTION

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 [1-3]. Breast cancer 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.

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. Researches showed that it is possible to diagnose breast cancer with the help of an electromagnetic device [4-10]. The device measures the dielectric difference in 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 harder to 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 [10-14].

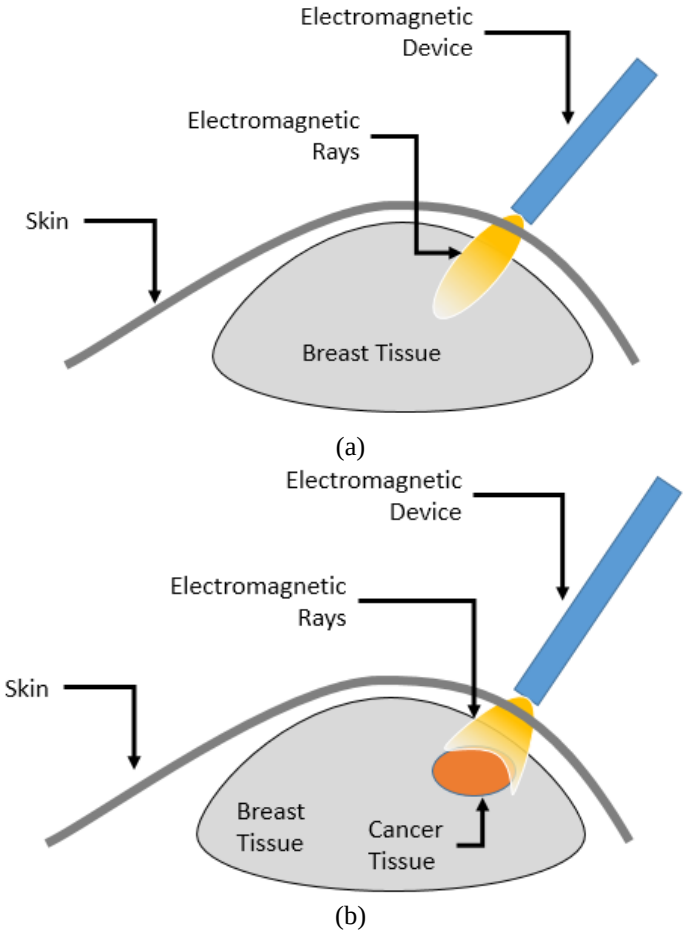


Fig. 1. The electromagnetic device diagnosing breast (a) without cancer tumour, (b) with cancer.

II. METHODOLOGY

The methodology consists of two phases as explained below.

A. Data acquisition and processing

The electric field data is obtained from the Ansys HFSS v15 simulator, the Ansys HFSS is selected in particular because of its efficiency in the calculation of dielectric property. A 3D breast model was created and an electromagnetic wave of 2.4 GHz is illuminated on that model, the rate of absorption and refraction will depend on the dielectric property as cancer has different dielectric property than normal breast tissue it absorbs part of the EM wave and refract the rest of the EM wave creating its signature pattern, and this field pattern can be measured and recorded. The field

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 10 Issue 03, March-2021

measurements consist of 180 data points for a single measurement, each data point denotes the intensity of the electric field in that direction. The overall data consists of 90 datasets for the tumour size varies from 0 to 30mm.

B. Training and Testing

The acquired data used trained and tested the machine learning model using two software tools one is the Python with scikit and the second one is the WEKA, in this process at least five algorithms used to train the model. In the Python algorithm such as Random Forest, XG Boost, Support Vector Regression, Decision tree, and Multi-Layer perception are employed. In the WEKA algorithms such as Random Tree, Support Vector Regression, K-means, Decision Tree and Multi-Layer Perception are employed. The algorithms are selected based on the ease of implementation or execution in that respective software tool. The results obtained from these models were plotted and compared with each other.

The Python and WEKA process the data a bit different than each other, mainly the Python express accuracy in terms of R2 score but WEKA express it in terms of the correlation coefficient, the R2 score expresses how accurate the predictions are but the correlation coefficient expresses how much the input and output are related in other words how much it is predictable in addition error related parameter such as Mean Squared Error (MSE), Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) & Relative Absolute Error.

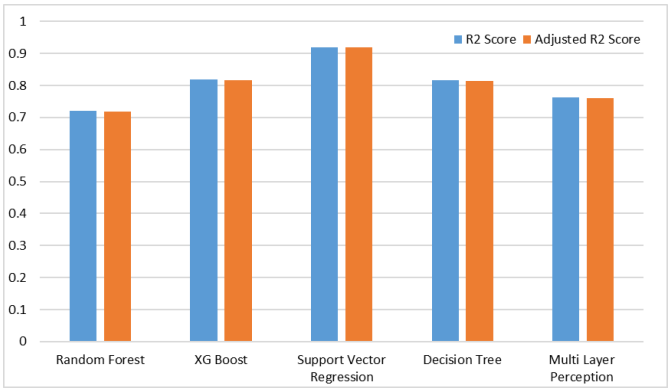
III. RESULTS AND DISCUSSIONS

A. Preadictions by Python

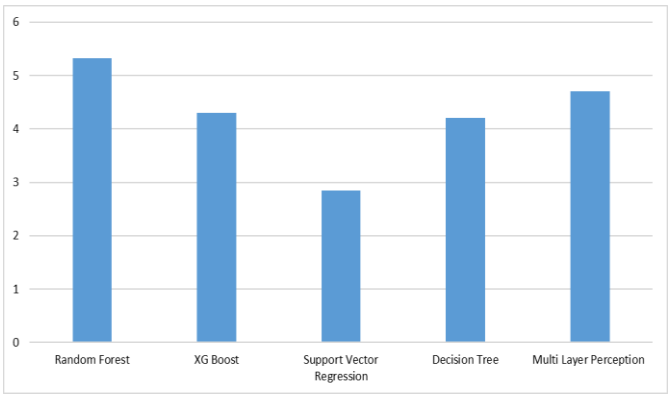
The results obtained from the Python are tabulated in TABLE I. The Random Forest algorithm is given an accuracy of 72.08% with a root mean squared error of 5.32 and the adjusted R2 score is 71.78% which proves that the model is neither over-fitted nor under-fitted but from table 1 it's observed that it has more error compared with the other algorithms. The XG Boost gives an accuracy of 81.78% with a root mean square error of 4.3 and it has an adjusted R2 score of 81.59%, observing the R2 score and adjusted R2 score confirms that the in training and in testing it gives the same accuracy but it's not a suitable candidate for training the model because it still has more error rating compared with other algorithms. The decision tree gives a model with 81.61% accuracy with a root mean square error of 4.21 and an adjusted R2 score of 81.42%, if observed it produced nearly identical results compared to the XG boost hence it's not the suitable algorithms for training the model. The multi-layer perception algorithm produces a model with 76.37% accuracy with a root mean square error of 4.70 and it has an R2 score of 76.12%. By these results, it can be concluded that its performance is similar to Random forest hence it can't be used for training. The Support Vector Regression has a very good accuracy of 92% with a root mean square error of only 2.85 and it has an adjusted R2 score of 91.91%, as it has the highest accuracy and lowest error rate compared to any other algorithms in the table it can be concluded that it's the most suitable algorithm to train the model. These parameters are plotted in the Fig. 2 for ease of understanding.

TABLE I. PYTHON RESULTS

Parameters / Algorithms	MSE	MAE	RMSE	R2 Score	Adjusted R2 Score
Random Forest	28.3664	4.1787	5.3260	0.7208	0.7178
XG Boost	18.5069	3.4537	4.3019	0.8178	0.8159
Support Vector Regression	8.1245	2.0475	2.8503	0.9200	0.9191
Decision tree	17.7327	3.1896	4.2110	0.8161	0.8142
Multi-Layer perception	22.1685	3.7445	4.7083	0.7637	0.7612



(a)



(b)

Fig. 2. (a) Acuuracy of preadictions, (b)Error of preadiction.

B. Preadictions by WEKA

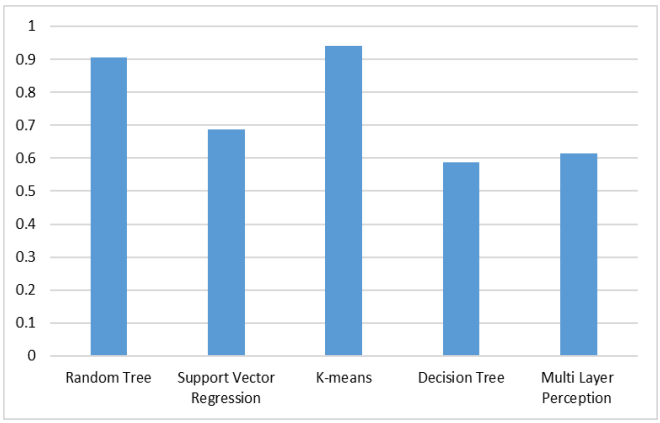
The results obtained by the WEKA are tabulated in TABLE II. . In this tool, the Random Tree correlation coefficient of 0.90 with root mean squared error 4.18. The K-means has a 0.94 correlation coefficient with a root mean square error of 3.29. The decision tree gives a correlation coefficient of 0.58 with a root mean square value of 8.18. The multi-layer perception gives a correlation coefficient of 0.61 with a root mean squared error of 8.84. And in last the support vector regression gives a correlation coefficient of 0.68 with a root mean square error of 8.07. By analyzing the results it's concluded that the K-means algorithm performed better in this tool. For ease of understand all the results are plotted in Fig. 3.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

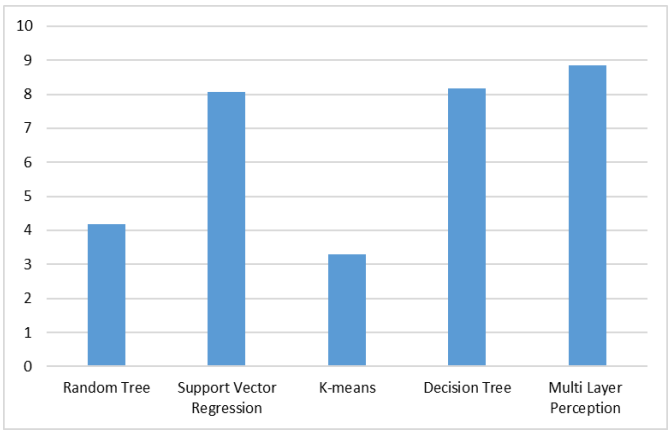
International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 10 Issue 03, March-2021

TABLE II. WEKA RESULTS

Parameters / Algorithms	Correlation Coefficient	MAE	RAE	RSE
Random Tree	0.9047	2.7931	34.5889	4.1895
K-means	0.9404	2.2069	27.3295	3.2905
Decision Tree	0.5869	6.1724	76.4372	8.183
Multi-Layer Perception	0.6131	6.207	76.8657	8.8442
Support Vector Regression	0.6866	7.1764	88.87	8.0735



(a)



(b)

Fig. 3. *Acuuracy of preadictions, (b)Error of preadiction.*

IV. CONCLUSION

After analyzing the results from both the tools it's confirmed that a malignant tumour size can be predicted using a 2.4 GHz spectrum electric field data with the help of machine learning algorithms. This can be served as a base model for developing future medical diagnosis devices which can be cost-effective and portable. In the future, the same method of analysis can be applied to predict the soft foreign objects with the body which are invisible to the normal diagnosis technique.

REFERENCES

[1] Marino, N., German, R., Rao, X. et al. Upregulation of lipid metabolism genes in the breast prior to cancer diagnosis. npj Breast Cancer 6, 50 (2020). https://doi.org/10.1038/s41523-020-00191-8

[2] 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

[3] 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

[4] 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

[5] 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.

[6] 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.

[7] 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.

[8] H. Song, H. Watanabe, X. Xiao and T. Kikkawa, "Influence of Air-gaps 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.

[9] 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.

[10] 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

[11] N. Amral, C. S. Ozveren and D. King, "Short term load forecasting using Multiple Linear Regression," 2007 42nd International Universities Power Engineering Conference, Brighton, 2007, pp. 1192-1198, doi: 10.1109/UPEC.2007.4469121.

[12] H. Shakouri, G. R. Nadimi and F. Ghaderi, "Fuzzy linear regression models with absolute errors and optimum uncertainty," 2007 IEEE International Conference on Industrial Engineering and Engineering Management, Singapore, 2007, pp. 917-921, doi: 10.1109/IEEM.2007.4419325.

[13] X. Chen, "Recursive local polynomial regression estimation and its applications," Proceedings of the 31st Chinese Control Conference, Hefei, 2012, pp. 2043-2048.

[14] A. V. Omelchenko and O. V. Fedorov, "Polynomial regression coefficients estimation in finite differences space," 2015 25th International Conference Radioelektronika (RADIOELEKTRONIKA), Pardubice, 2015, pp. 257-260, doi: 10.1109/RADIOELEK.2015.7129024.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)