

A comparative Survey of Breast Cancer Identification with Machine Learning Algorithms

¹ Tadesse Mega, ² Dr.S.K Manju Bargavi

¹ Student, ² Professors

¹ MCA Department, Jain (Deemed-to-be-University), Bangalore, India,

² MCA Department, Jain (Deemed-to-be-University), Bangalore, India

Submitted: 01-03-2022

Revised: 10-03-2022

Accepted: 12-03-2022

ABSTRACT: Around the world, breast disease is the main kind of cancer growth in ladies, representing 25% of all cancer cases. In 2018, it brought about 2.5 million cancer cases and 628,000 deaths. As WHO reported in 2018, breast cancer is the main disease in Ethiopia women. Similarly, according to Ethiopia health institution report, It turns into the most dangerous disease growth rapidly and represents 33% of malignant growth cases in ladies. The right diagnosis helps to find valuable treatment. In Ethiopia, Black Lion Hospital is the only main place for the diagnosis of cancer disease. The Hospital is treating only 1% of the absolute patients and the pathologists analyze biopsy slides physically or manually. This method is tedious, might be inclined to blunder, and time-consuming method. The ML method of examination is more accurate and recovers analysis time. This paper plans to introduce an examination of the great well-known algorithms to forecast cancer disease. Specifically random classification algorithms, decision classification tree algorithms, and artificial neural network Algorithm. The dataset is taken from the Kaggle, to build our model. The outcomes acquired are competitive and the precise model can be utilized for diagnosis.

KEYWORDS: - Artificial neural network, benign, decision tree, malignant, Random Forest

I. INTRODUCTION

The breast is an associate degree organ of the human frame, that is assessed as an associate degree secretor, designating its practicality to produce breast milk. The breast is specifically created of fat, animal tissue, and many little structures, called lobules. [1] The cell is that the tiniest organic unit with numerous systems and able to unbiased continuation.[2] cancer may be outlined

as a nation of illness that terminates the cells to answer to regular stimuli. associate degree irregular cell that spreads and breaks out of management can offer rise to a tumor. A tumor that has dilated on the far side of the layer of tissue wherever it advanced, and is spreading into peripheral healthful tissue is alleged to be persistent. [3] carcinoma is represented as a cancerous tumor that arises from breast tissue, often from the animal tissue cells at intervals the wall of milk ducts that offer the ducts with milk of the breast.[4,5] the first peril determinants for carcinoma square measure being feminine, antique age, and genetic components. alternative relevant factors embody overdue 1st parturition, null parity, and biological time [6, 7]. some researchers additionally show proof of alcohol ingesting being a hazard issue [8], whereas breastfeeding and physical interest have been established to lower the uncertainty of breast most cancers [9, 10]. carcinoma prognosis is that the technique of locating the existence of most cancers mobile within the patient's breast and multidisciplinary try. at some stage within the breast most cancers diagnostic the doctors selects and use distinctive kinds of sorting out techniques to return across most cancers and to visualize if cancer has increased to completely different components of the body out of doors the breast and also the liquid body substance nodes underneath the arm [11]. The preponderance of occurrences of irregularities at intervals at the breast is known by the manner of screening mistreatment imaging technologies comprehensive of diagnostic procedure, resonance imaging, or ultrasound and diagnostic assay.[11,12,13] The breast most cancers-detecting result's reported in any of the 5 general FNA biological science classes is an insufficient pattern (not comfortable animal tissue cells for analysis), Benign (now not cancer),

unusual, atypical/uncertain however probable benign, Suspicious, and doubtless malignant (cancer) and Malignant (most cancers) for the distinctive patient. most cancers may be handled in many alternative ways, counting on the kind of foremost cancer, its region, grade, and its metastasis level (whether or not it unfolds or not). the most common varieties of treatments for breast most cancers are as follows: surgery, secretion remedy, radiation therapy, therapy biological treatment choices. The guide carcinoma detection methodology is accomplished by employing a health practitioner/pathologist while not the wish of a laptop computer or computer machine (which is specially designed for slide microscopic digital photograph analysis) the utilization of a light magnifier. carcinoma detection Associate in Nursing prediction approach may be the utilization of machine gaining information of Algorithms specifically designed for detection the carcinoma from the given at the Wisconsin Diagnostic carcinoma data-set that is derived from a digitized photograph of a resonance imaging. during a processed carcinoma discovery device, no wish for Associate in Nursing old health professional or a crew of medical doctors for breast cancer-detecting result verification.

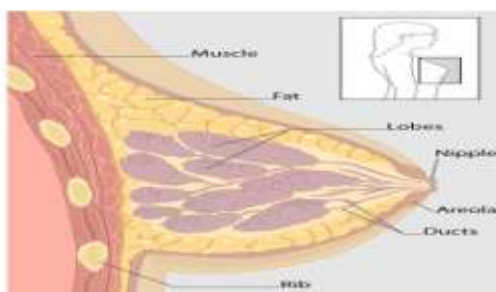


Figure 1: The structure of breast

II. 2. LITERATURE SURVEY

2.1 Related Research

This study was built from the existing studies. In this section, the related previous researches and their gap, about ML (machine-learning) and algorithms will discuss. Many research works have been conducted on breast cancer classification. Creating a machine learning model with high accuracy is the most critical in health sector to make a significant change in a patient survival rate.

There have been numerous proposed strategies for classifying breast cancer patterns, up to now different researchers have attempted to build a model using different machine learning techniques with different classification accuracy. Most of

them build a model which has an accuracy value less than 96% . Our objective on this Research paper is to make a high accuracy model (more than 96%) with a low error rate by using deep learning technique on the same dataset.

1. Breast cancer cell Detection with decision Classification tree and adaptive-boosting model, Author name Tsehau Admasu and two others, Published on 2020. This project used decision Classification tree model and adaptive-boost model to predict breast cancer and the maximum accuracy using adaptive boosting model in this project is 92.53%, it has a 8% error rate.[18]

2. Classification of the Breast related Cancer Data Using ML different Algorithms, Author name, Akbugday, Published by TIP -2019. This project used KNN and SVM, NB and WAKE tools to predict breast cancer and maximum accuracy using SVM model in this project is 96.8% and this model has a 3.2% error rate[19].

3. Breast related Cancer perdition by ML, Author name Gemeches, Published by International Journal of Scientific Development and Research, 2021. This project used logistic Classification model, decision tree model AND RF Classification model and used WAKE tools for predicting breast cell cancer and the highest accuracy in this project is 96% using RF model and this model has a 4% error rate[20].

4.A Probabilistic Detection Model for Breast Cancer, Author name Shweta Kharya two others, Published on April 2014 by International Journal of Computer Applications .This project proposed a Naive Bayes Classifiers of machine learning algorithms to predict breast cancer and maximum accuracy using Naïve Bayesian model in this project is 93% and this model has a 7% error rate

5. Classification of malignant and benign tissue with logistic regression,, Author name Mohammad Abdu Hasid, Published in: 2019 by Informatics in Medicine. this project used logistic regression to predict breast cancer and the result shows it's accuracy is 96% and this model has a 4% error rate[21].

6. Cancer cell Detection Using ML technique, Author name Shubha Sharm and others, published by 2018 on International Computational Conference. This project used RForest, KNearestNeighbor, and Naïve algorithm to predict breast infected by cancer and the result shows RF=94.7%, NB=94.4 % and KNN has 95% accuracy, The max accuracy of model in this project 95% and this model has a 5% error rate

7. Breast Cancer Classification using SVM Classifier, Author name E. Karthikeyan and one other, Published on 2019 by International Journal of Technology and Engineering .The project

proposed SVM to detect Breast Cancer and the proposed algorithm has achieved an accuracy of 89.11% and it has 11% error rate[22].

2.2 Machine Learning Techniques

ML is concerned with the design of algorithms and strategies that enable machines to learn and predict based on previous information automatically. Machine learning is the process of extracting information from data using mathematical and statistical methods [15]. ML is a branch of computer data science concerned with the research and development of algorithms that learn from and predict data. Without even being explicitly trained how and where to search for, machine learning helps computers to discover previously undiscovered insights.

Machine learning are divided into three types.

Supervised learning's are developed using inputs and accurate outputs, with the algorithm's goal being to develop a general norm that links inputs and outputs.

Unsupervised learning;

The method is only provided input data without any matching output data, and its purpose is to examine the data and identify structure within it.

Semi-supervised learning;

It offers the functionality of both learning algorithms (supervised & unsupervised). It learn from both labeled data and unlabeled data in a dataset

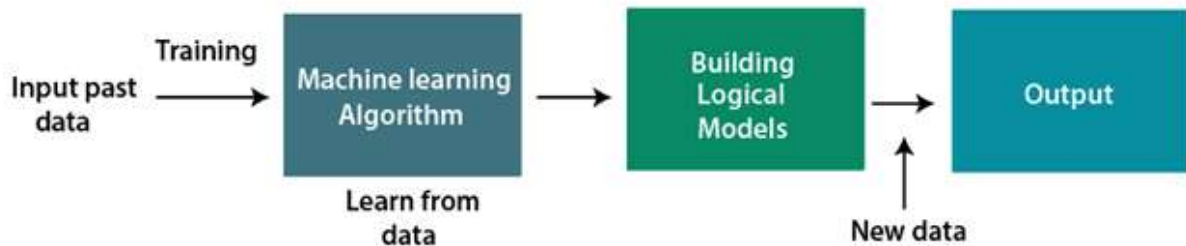


Figure 1. ML Pipeline

2.2.1 Decision classification algorithm

It is a supervised classification algorithm which is used to perform classification issues, but it is to solve regression patterns. It is just a leaf structured classifier, with inner nodes shows dataset features, branches shows classification rules, and a child nodes shows the result. There are two nodes in a Decision classification algorithm: Decision tree node and Child(Leaf) tree Nodes. Child nodes are the result (out put) but it has no branching, whereas Deciding nodes are being used to do decisions and

also have numerous branches.

The work of DT

The procedure for determining the class of a particular dataset in a Decision classification starts at the root tree node. The algorithm jumps to the second tree node after comparing the results of a root tree with the result or values of real dataset features.

The algorithm then compares the two values once more. It will keep going until it reaches the leaf tree node. The procedure can be illustrated in the diagram below:

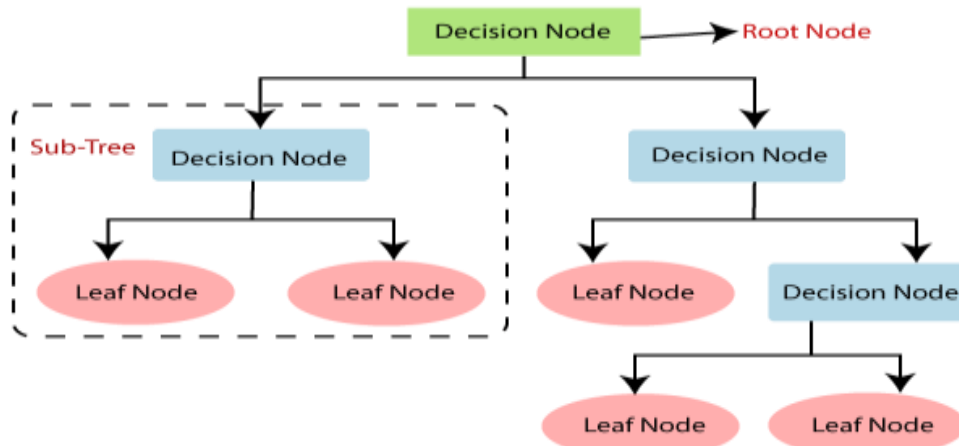


Figure 3. Visualization of the decision tree classifier algorithm

2.2.2. Random forest is a type of ML algorithm and it has many classification trees. Every tree has option for sorting a set of data. After that, the algorithm adopts the categorization that received the most votes from each individual tree. The algorithms are a form of supervised learning algorithm that uses the trained dataset to generate a decision classification tree that forecast the results of the test sample dataset. When the decision classification tree is given the training sets (which consisting of goal and feature), it Generates a list of norms. These criteria are

therefore applied to test data to forecast the class of new data. It generates a forest of numerous decision models, each with its own set of rules.

Working with RF:-

The random forest has the following steps:

- 1: random items are chosen at random from the total data collection
- 2: For each data sample, one decision trees are built.
- 3: Every decision tree produces a result or output
- 4: the final outcome is generated based on Majority rule Voting

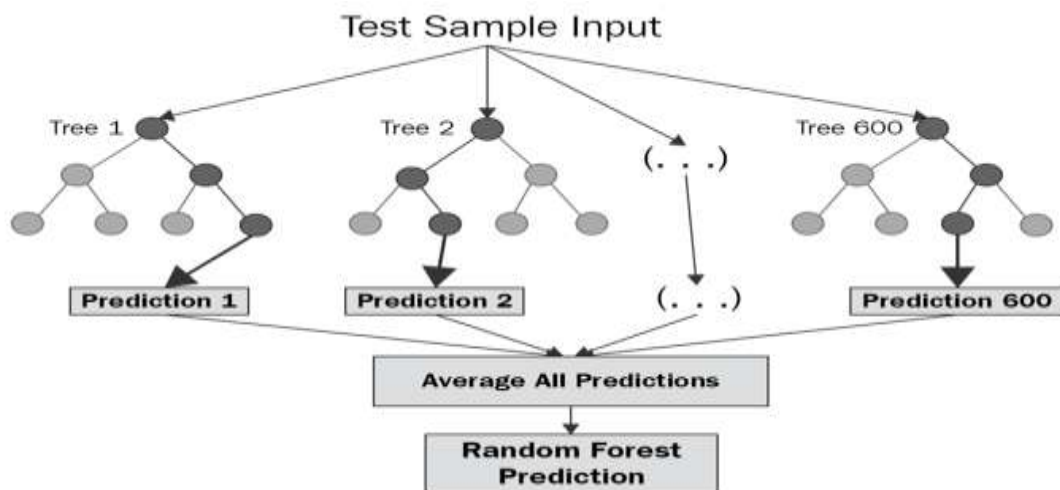


Figure 4 Visualization of the Random forest classifier

2.1.3. Artificial Neural Networks

They are determined from by genetic neurons. Many of the brain's activities are still unknown, but we do know that biological neurons allow the brain to handle large quantities of data in complicated manner

The way ANN operate is analogous to how neurons Performs in our bodies system. To know how it works, first we have to understand it's structured. Neural network have three essential layers –

Layers of Input

This layer is the 1st layer of an ANN model, and it collects input signals in many formats

Layers of Hidden

They are in the midsection of a ANN .

There

could be a single hidden layer or several hidden layers. It evaluates input signal mathematically and determines the relationship between the parameters /input & output described in [16]

Layer of Output

The output produced through numerous computations done by the central layer is obtained in this layer.

The operation of a neural network is influenced by a number of hyper parameters. Those parameters are frequently concerned with in the outcome of ANNs. Biases and weights are only a few of them. . Every node in the ANN has a certain amount of weight. [16].

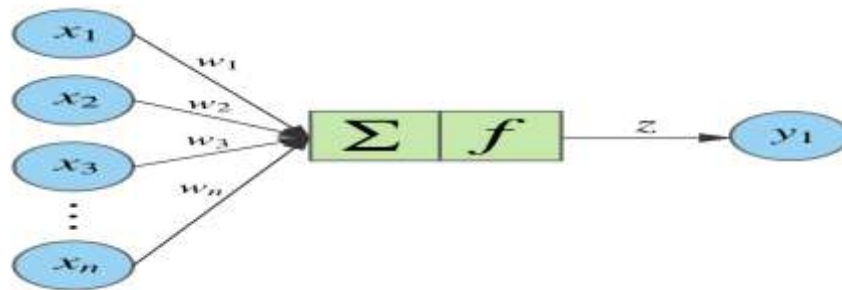


Figure 5. working with artificial neural networks.

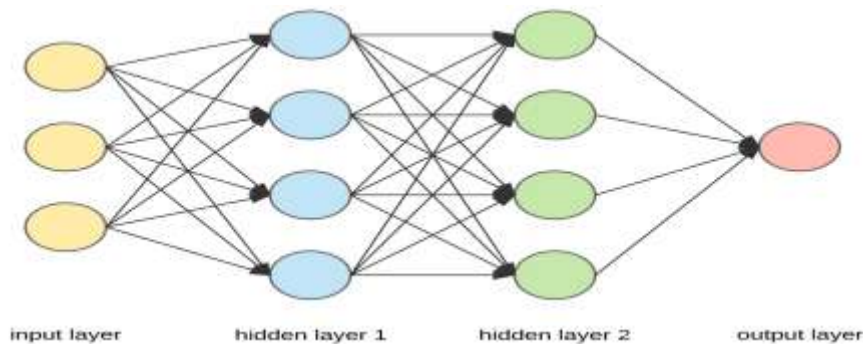


Figure 6. Visualization of the artificial neural network

III. CONCLUSION

This survey article looked at the strategies and algorithms offered for identifying breast diseased cells and, in certain circumstances, assessing their stage so that the cancer victim can receive correct therapy and enhance his quality of life. Although digitized mammography is commonly used in the early stage of cancer diagnosis, because of its harmful effects on a human. Additional other methods like as biopsy, infrared have been advocated. Mammography techniques and biopsy were shown to be the most accurate methods. Machine learning techniques are frequently applied in medicine center and have proven to be an effective screening tool that assists clinicians in processing data. In this project we use decision tree, random forest and ANN algorithm and compare the result of the models, The choice of appropriate algorithms with a best dataset will result in the increased accuracy of prediction systems and reduce error. These technologies can assist in determining the best treatment options for a cancer patient.

REFERENCES

- [1]. S. D. Tzikopoulos, M. E. Mavroforakis, H. V. Georgiou, N. Dimitropoulos, S. eodoridis, A fully automated scheme for mammographic segmentation, and classification based on reast density and asymmetry, computer methods and programs in biomedicine, 102 (2011) 47-63
- [2]. M. Murat Dundar, Sunil Badve, Gokhan Bilgin, Vikas Raykar, Rohit Jain, Olcay Sertel, and Metin N. Gurcan, "Computerized Classification of Intraductal Breast Lesions Using Histopathological Images", IEEE Transactions on Biomedical Engineering, Vol.58, No. 7, July 2011, pp. 1977-84.
- [3]. Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., Walter, P., Molecular Biology of the Cell, Garland science, Taylor & Francis Group, LLC, 2008.
- [4]. World Health organization "Cancer" retrieved from <http://www.emro.who.int/health-topics/cancer/index.html>, last accessed on 28 May 2017.
- [5]. National Cancer Institute, "Breast Cancer", retrieved from <http://www.cancer.gov/cancertopics/types/breast>, last accessed on 29 June 2014.
- [6]. Verywell, "Mammary Epithelial Cells: Function and Abnormalities" retrieved from <https://www.verywell.com/> last accessed on 29 May 2017.
- [7]. Boyle, P., Levin, B. (Eds.), World Cancer Report 2008. Ann Oncol; 2008; 19(9): 1519-1521, "Need for global action for cancer control".
- [8]. Boyle, P., Levin, B. (Eds.), World Cancer Report 2008. Ann Oncol; 2008; 19(9): 1519-1521, "Need for global action for cancer control". Collaborative Group on Hormonal Factors in Breast Cancer, Alcohol, tobacco and breast cancer—collaborative reanalysis of individual data from 53 epidemiological studies, including 58,515 women with breast cancer and 95,067 women without the disease, British Journal of cancer 87 (11) (2002) 1234–1245.
- [9]. Thune, I., Brenn, T., Lund, E., Gaard, M., Physical activity and the risk of breast cancer, New England Journal of Medicine 336(18) (1997) 1269–1275.
- [10]. Collaborative Group on Hormonal Factors in Breast Cancer, Breast cancer and breastfeeding: collaborative reanalysis of individual data from 47 epidemiological studies in 30 countries, including 50302 women with breast cancer and 96973 women without the disease, Lancet 360 (9328) (2002) 187–195
- [11]. Cancer.Net, "Breast cancer diagnosis", retrieved from www.cancer.net, last accessed on 19 May 2017.
- [12]. Cochrane Nordic, "Screening for breast cancer with mammography", retrieved from <http://nordic.cochrane.org/screening-breast-cancer-mammography> last accessed on 15 July 2017.
- [13]. National Breast Cancer Centre Incorporating the Ovarian Cancer Program "Breast fine needle aspiration cytology and core biopsy: a guide for practice", First Edition Prepared by the National Breast Cancer Centre Funded by the Department of Health and Ageing. National Breast Cancer Centre 2004, ISBN Print: 174127036 7 Online: 174127 042 1, CIP: 618.190758.
- [14]. NHS Cancer Screening Programs, "Guidelines for Non-Operative Diagnostic Procedures and Reporting in Breast Cancer Screening, Stream line Offset, Hoddesdon, Herts, Publication No 50, ISBN 1 871997 44 5, June 2001.
- [15]. Muhammad Jamil Moughal, "Which Machine Learning algorithm to use? Muhammad Jamil Moughal – Medium," 2018.
- [16]. "jvatpoint.com/machine-learning-decision-tree-classification-algorithm" Feb 13th, 2019.
- [17]. Introduction to Artificial Neural Networks at

- <https://data-flair.training/blogs/artificial-neural-networks-for-machine-learning/>
- [18]. Breast_cancer_prediction_model_with_decision_and_adaptive_boosting <https://www.researchgate.net/publication/349890120>
- [19]. Classification_of_Breast_Cancer_Data_Using_Machine_Learning_Algorithms <https://ieeexplore.ieee.org/document/8895222>
- [20]. /8895222
- [21]. Breast_Cancer_Detection_and_Prediction_using_Machine_Learning <https://ijsdr.org/viewpaperforall.php?paper=IJSDR2105051>
- [22]. Classification_of_malignant_and_benign_with_logistic_regression <https://www.sciencedirect.com/science/article/pii/S2352914818301497>
- [23]. Breast_Cancer_Classification_using_SVM_Classifier <https://www.ijrte.org/wpcontent/uploads/papers/v8i4/D7227118419.pdf>