

Analysis of Machine Learning Techniques for Breast Cancer Prediction

Mrs N. Vanitha¹, R. Srimathi² and J Haritha³

¹Assistant Professor, Department of Information Technology, N.G.P Arts and Science College Coimbatore, INDIA

²Student, Department of Information Technology, N.G.P Arts and Science College Coimbatore, INDIA

³Student, Department of Information Technology, N.G.P Arts and Science College Coimbatore, INDIA

¹Corresponding Authors: vanitha969@gmail.com

ABSTRACT

The most frequently happening cancer among Indian women is breast cancer, which is the second most exposed cancer in the world. Here is a chance of fifty percent for fatality in a case as one of two women diagnosed with breast cancer die in the cases of Indian women. With the rapid population growth, the risk of death incurred by breast cancer is rising exponentially. [2] Breast cancer is the second most severe cancer among all of the cancers already unveiled. A machine learning technique discovers illness which helps clinical staffs in sickness analysis and offers dependable, powerful, and quick reaction just as diminishes the danger of death. In this paper, we look at five administered AI methods named Support vector machine (SVM), K-closest neighbours, irregular woodlands, fake/ Artificial neural organizations (ANNs). The performance of the study is measured with respect to accuracy, sensitivity, specificity, precision, negative predictive value. Furthermore, these strategies were evaluated on exactness review region under bend and beneficiary working trademark bend. At last in this paper we analysed some of different papers to find how they are predicted and what are all the techniques they were used and finally we study the complete research of machine learning techniques for breast cancer.

Keywords-- Breast Cancer, Prediction, Machine Learning

I. INTRODUCTION

The most two mainstream AI methods have been utilized for grouping of Wisconsin Breast Malignancy (Unique) dataset and the order execution of these procedures. Machine learning is a method of data analysis that automates analytical model building. It is a branch of artificial intelligence based on the idea that systems can learn from data, identify patterns and make decisions with minimal human intervention. [2]. Machine learning algorithms build a model based on sample data, known as "training data", in order to make predictions or decisions without being explicitly programmed to do so. Machine learning algorithms are used in a wide variety of applications, such as computer vision, where it is difficult or unfeasible to develop conventional algorithms to perform the needed tasks

II. TYPES OF MACHINE LEARNING

There are two types of machine learning, supervised learning and unsupervised learning. Listed in the below figure 1

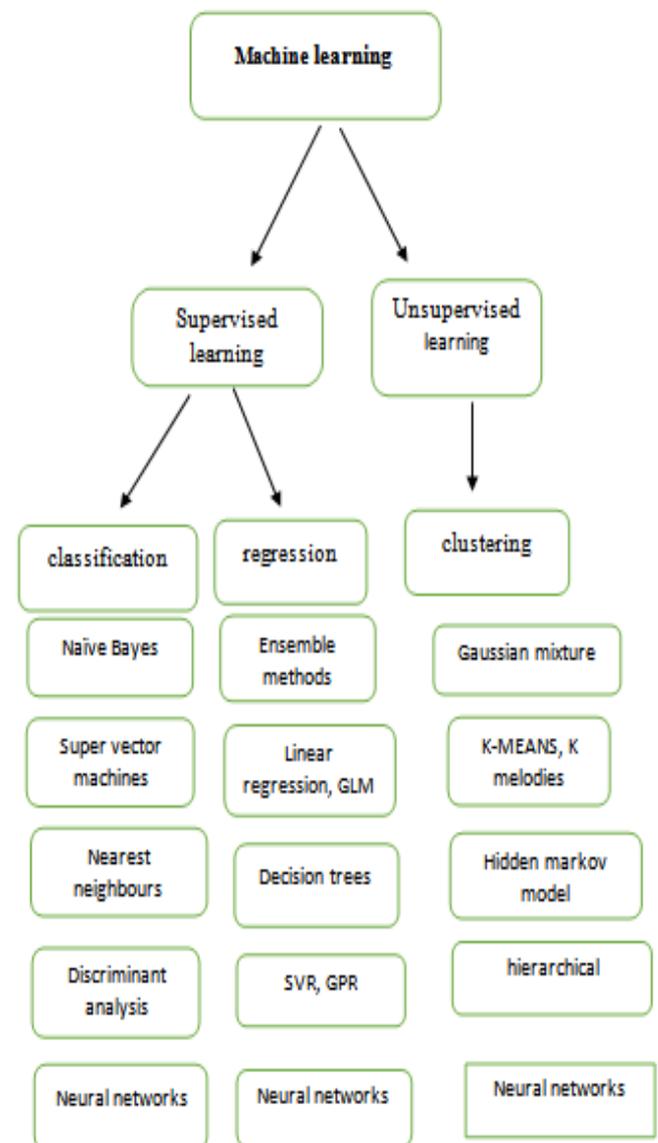


Figure 1: Types of machine learning

1) Supervised Learning: Regulated learning is the AI undertaking of learning a capacity that maps a contribution to a yield dependent on model info yield sets. It deduces a capacity from named preparing information comprising of a bunch of preparing model.

2) Unsupervised Learning: Learning In this technique, the machine is forced to train from an unlabelled dataset and then differentiating it based on some characters and allowing the algorithm to act on that information without external guidance.

3) Reinforcement Learning: The learning process continues from the environment in an iterative fashion. All possible system states are eventually learned by the system over a prolonged period.

III. MACHINE LEARNING ALGORITHMS

1) Support Vector Machine (SVM in): Support vector machine (SVM) may be a discriminative classifier characterized by an isolating hyper plane. In two-dimensional space this hyper plane may be a line separating a plane in two sections where in each class lay in either side. SVM gives 83.33% accuracy with linear kernel.

2) Multilayer Perceptron: Multilayer perceptron (MLP) has several layers, two terminating layers are called input-output layer and intermediate layers are called hidden layers. divided into training and testing instances within the ratio of 80:20. It means out of 699 instances; 560 instances.

3) K-Nearest Neighbour (KNN): KNN is valuable for an enormous dataset don't utilize numerical examination. In the most noticeably terrible situation, KNN needs more memory to check all informational indexes. Also, its normal exactness in all circle is 89.29%.

4) Random Forest: It creates a group of decision trees from randomly selected subset of training set. It then combines the results from different decision trees to predict the target class of the test instance. It works well

because a single decision tree may be susceptible to noise, but the collective outputs of multiple decision trees reduce the effect of noise giving better results. It helps to control over-fitting which in turn leads to better predictive accuracy.

5) Artificial Neural Network: Artificial Neural Network (ANN) is one the most advanced and powerful machine learning algorithm that can be used for many purposes such as classification, regression, voice recognition, targeted marketing, etc. A single unit of an ANN is called a perceptron, it was first invented in 1957 by Frank Rosenblat, as he wanted to invent something that could learn and adjust itself to meet the required needs. The way in which ANNs work is a bit complex but understandable, firstly the input nodes are taken, and the number of input nodes is equal to the number of independent variables, then these inputs are weighted through synapses and passed on to the next neuron [4].

IV. SIGNIFICANCE OF MACHINE LEARNING

- Decision tree algorithms, lazy learning methods and resembling techniques are applied to standard datasets to analyse and compare the data classification algorithms on nominal data.
- Random Forest and K-Nearest-neighbour classifiers achieved high accuracy for data classification
- High-value predictions that can guide better decisions and smart actions in real-time without human intervention.
- Through Virtual Assistant solutions, machine learning automates tasks that would otherwise need to be performed by a live agent – such as changing a password or checking an account balance.

There are different types of applications. Listed below in the figure2

V. APPLICATIONS OF MACHINE LEARNING

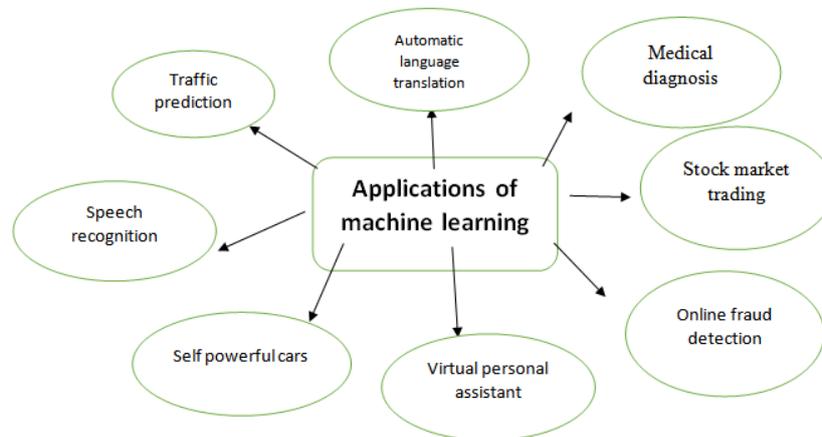


Figure 2: Applications of machine learning

VI. BREAST CANCER

Breast cancer is one of the leading causes of women mortality world-wide. It is common among the younger age groups (in their thirties and forties) and mostly occurs in women aged from 30 to 69. A regular diagnosis helps us to accumulate the variations and its impact on a patient. Almost 70 percent of the deaths from cancer disease happen in countries that have low and middle income. The most common cancer type among women are breast, lung and colorectal, which totally symbolize half of the all cancer cases learning (ML) methods studying the data and extracting key characteristics of relationships and information from dataset.

Types of Breast Cancer

Daze duplicate- Blind carbon copy (BCC) means to decide the reasonable therapy, which can be forceful or less forceful, contingent upon the class of the malignancy. To make a decent prognostic, bosom malignant growth characterization needs nine attributes which are:

1. Determine the layered constructions (Clump Thickness);
2. Assess the example size and its consistency (Uniformity of Cell Size);
3. Malignancy cells spread everywhere on the organ and ordinary cells are associated with one another (Marginal Adhesion);
4. Proportion of the consistency, augmented epithelial cells are an indication of threat (Single Epithelial Cell Size);
5. In kind hearted tumour's cores isn't encircled by cytoplasm (Bare Nuclei);
6. Depicts the core surface, in kind hearted cells it has a uniform shape. The chromatin will in general be coarser in tumours (Bland Chromatin);
7. In typical cells, the nucleolus is generally undetectable and little. In disease cells, there are more than one nucleolus and it turns out to be substantially more unmistakable, (Normal Nucleoli);
8. Gauge of the quantity of mitosis that has occurred. The bigger the worth, the more prominent is the opportunity of threat (Mitoses)

Significance of Study on Breast Cancer Prediction

During the beginning phases of the sickness, the indications are not introduced well and thus conclusion is deferred.

- It is suggested by the NBCF (National Breast Cancer Foundation) that ladies beyond forty a year old enough must to get a mammogram

once per year. A mammogram is a X-beam of the bosom

- Breast disease can be relieved with flow clinical medicines and new spearheading procedures.
- Breast malignant growth identification utilizing Relevance Vector Machine, acquired an exactness of 97% utilizing Wisconsin unique dataset.
- Breast disease expectation and guess are significant test to clinical local area.
- Recurrence of disease is the greatest apprehensions for malignancy patient, and this can influence their personal satisfaction.
- The point of this exploration is to foresee bosom malignancy from disease highlights with high exactness.

Background Study

- In 2019, Naveen, Dr R K Sharma, Dr Anil Ramachandran published a journal in iee on the topic of Effective breast cancer prediction using ensemble machine learning model, Decision tree and KNN gives 100% accuracy with ensemble technique. K nearest neighbour gives maximum accuracy as 100%.
- In 2019, Ebru aydindag, Bayrak, Pinar kirchi Ramachandran published a journal in iee on the topic of Comparison of machine learning methods for breast cancer Diagnosis and SVM algorithm in machine learning techniques has showed best performance in the accuracy of 96.9955% for diagnosis and prediction
- In 2019, Dhanya irenic rose Perl, Sai Sindhu, Madhumathi, Siva Kumar. Published a journal in iee on the topic of AComparative study of breast cancer prediction using machine learning and feature selection and Random forest classifier gives highest accuracy with features selection for both the datasets, f test wrapper method constantly improves the accuracy of the classifier for WDBC dotnet
- In 2018, Shubham Sharma, Archit Aggarwal, Tanupriya Choudhry published a journal in iee on the topic of Breast cancer classification using machine learning and Result shows that KNN gives highest accuracy 97.51% with lowest error rate the NB classifier 96.19%
- In 2018, MeriamAmarna, ikram Gagauz published a journal in iee on the topic of Breast cancer detection using machine learning algorithm and the result obtained is very competitive and can be used for detection and treatment.

Table 1: Background study for breast cancer using machine learning techniques

| S. no | Author | Journal | Title of the paper | Year-method | Output |
|-------|--|-------------|--|---|--|
| 1 | Naveen, Dr R K Sharma, Dr Anil Ramachandran | Ieee Xplore | Effective breast cancer prediction using ensemble machine learning model | 2019 / decision tree, KNN feature Scaling, Cross validation various Ensemble machine learning techniques with bagging model | Ensemble model improves the system performance with un-biasing. Decision tree and KNN gives 100% accuracy with ensemble technique. K nearest neighbour gives maximum accuracy as 100% |
| 2 | Ebru aydindag, Bayrak, Pinar kirchi | Ieee Xplore | Comparison of machine learning methods for breast cancer Diagnosis | 2019/ Support vector method, K nearest neighbour, Native Bayes and Decision tree, ANN (artificial neural network and SVM popular machine learning technique applied in Wisconsin breast cancer technique. | SVM algorithm in machine learning techniques has showed best performance in the accuracy of 969955% for diagnosis and prediction |
| 3 | Dhanya irenic rose Perl, Sai Sindhu, Madhumathi, Siva Kumar. | Ieee Xplore | AComparative study of breast cancer prediction using machine learning and feature selection. | 2019/ feature selection techniques by existing machine learning algorithm | Random forest classifier gives highest accuracy with features selection for both the datasets, f test wrapper method constantly improves the accuracy of the classifier for WDBC dotnet. |
| 4 | Amarna, ikram Gagauz Meriam | Ieee Xplore | Breast cancer classification using machine learning | 2018/ Native Bayes, k nearest neighbour | Result shows that KNN gives highest accuracy 97.51% with lowest error rate the NB classifier 96.19%, if data is large the KNN time for running will increase |
| 5 | ShubhamSharma, Archit Aggarwal, Tanupriya Choudhry. | Ieee Xplore | Breast cancer detection using machine learning algorithm | 2018/ Random forest, KNN (K Nearest neighbour) and native Bayes | The result obtained is very competitive and can be used for detection and treatment.it has been observed that each of the algorithm had an accuracy of more than 94% to determine benign tumour/malignant tumour. KNN is most effective in detection of breast cancer and it had accuracy, precision and FIS cone over the other algorithm |

Table 2: Analysis of machine learning techniques for breast cancer prediction

| Author/Technique | Decision tree algorithm | Support vector machine | Multilayer perception | K nearest neighbor | Random forest algorithm | Artificial neural network | Feature selection algorithm | Naïve Bayes |
|--|-------------------------|------------------------|-----------------------|--------------------|-------------------------|---------------------------|-----------------------------|-------------|
| Naveen, Dr R K Sharma, Dr Anil Ramachandran | | | | | | - | - | - |
| Ebru aydindag, Bayrak, Pinar kirchi | - | | - | - | - | | - | - |
| Dhanya irenic rose Perl, Sai Sindhu, Madhumathi, Siva Kumar. | - | - | - | - | - | - | | - |
| Amarna, ikram Gagauz Meriam | - | - | - | | - | - | - | |
| ShubhamSharma, Archit Aggarwal, Tanupriya Choudhry. | - | - | - | | | - | - | yes |

This below table indicates the analysis of machine learning techniques for breast cancer prediction

Research Gap Identified

In those papers they used few features only for the diagnosis of cancer. Algorithms available in the study produces minimum accuracy only. In those papers they were not properly used some methods such that decision tree algorithm, support vector machine, K nearest method.

VII. CONCLUSION AND FUTURE WORKS

Breast cancer is one of the leading causes of death in women. So, early detection is very important. Thus, we studied many machine learning algorithms to detect the breast cancer. In future we will try on all features of machine learning techniques and to achieve best accuracy. Our investigation likewise energizes that nobody procedure can be named being the ideal AI strategy. Hence there is a solid requirement for better understanding into the legitimacy and consensus of many the talked about methods.

REFERENCES

[1] Naveen & Dr. K Sharma. (2019). Effective breast cancer prediction using ensemble machine learning

model. In: *International Conference on Recent Trends on Electronics, Information, Communication & Technology*.

[2] Ebru Aydindag & Bayrak, Pinar Kirchi. (2019). *Comparison of machine learning methods for breast cancer diagnosis*. Book ISBN: 978-1 7281-1013.

[3] Siva Kumar. (2019) A comparative study of breast cancer prediction using machine learning and feature selection. In: *Conference on Intelligent Computing and Control System, Amrita Vishwa Vidyapeetham, Amritapuri, India*.

[4] Amarna, Ikram & Gagauz Meriam. (2018). *Breast cancer classification using machine learning LRDSI laboratory*. University of Blida 1, Blida, Algeria 1.

[5] Tanishk Thomas & Nitesh, Pradhan. (2020). Comparative analysis to predict breast cancer using machine learning algorithm. In: *Conference on Inventive Computational Technology, Manipal University Jaipur*.

[6] Gupta, P. & P. S. (2018). Analysis of machine learning techniques for breast cancer prediction. *International Journal of Engineering and Computer Science*, 7(05), 23891-5.

[7] S. Mythili & A. V. S. Kumar. (2015). CTCHABC-hybrid online sequential fuzzy extreme kernel learning method for detection of breast cancer with hierarchical artificial bee. In: *IEEE International Advance Computing Conference (IACC), Bangalore*, pp. 343-348.