International Journal For Technological Research in Engineering
Volume 11 Issue 3 November-2023 ISSN (online) 2347-4718

MACHINE LEARNING TECHNIQUES AND EXTREME LEARNING MACHINE FOR EARLY BREAST CANCER PREDICTION

Dr.Parma Nand Prabhat
Assistant Professor
SCSE Sandip University Sijoul, Madhubani, India

Abstract:- Mamma Cancer is one of the most death distemper and most of the women are epidemical by this vital distemper in many parts of the world. Medical tests observances in hospitals for representative the disease are very much expensive as well as time- consuming. The puzzle can be resolved by diagnosing the problem in early spam of time and by legislation results with more exactitude. In this paper, different doohickey learning and neural curacy algorithm have been studied and juxtapose to predict cancer in early stages so that life can be saved. The dataset available publically for Breast ulcer has been used. Different algorithms compared include vindication Vector Machine Classification (SVM), K-Nearest Neighbor Classification (KNN), conclusion tree Classification (DT), Random Forest Classification (RF) and Extreme Learning Machine (ELM).All are compared on the basis of exactitude and processing time are opine as the parameters for comparing breakdown. The results reveal that dernier learning doohickey comes to be the better algorithm.

Keywords:- Mamma, Cancer, distemper, consuming, algorithm, exactitude, Learning Machine.

I. INTRODUCTION

Mamma Cancer has become the main reason behind the in animation of a lot of women all around the outside. The main reason for the in animation of women by this disease is the process by which it is Diagnostic ate.

The technology has become a general part of our lifestyles still we are lacking Abaft Diagnostic ate this critical disease in early bandstand [1]. As the disease is not Diagnostic ate in early bandstand, therefore, the mammography rate has been extended for a particular age group of concerned women [2]. Mamma Cancer is curable and life can be saved if it is Diagnostic ate in early bandstand. Different causes have been Diagnostic ate for this deathly disease including Imprecise hormonal imbalance, family fasti, obesity, radiation general practitioner and many more. Many doohickey learning and deep learning multiple partition algorithms are being applied to declare this disease.

Doohickey learning algorithms follow the Multifarious steps during classification saying [3] Network to find which algorithm gives the best sequel in terms of exactitude and processing time. Various Doohickey learning algorithms discussed here are vindication Vector Doohickey classification, K-Nearest Neighbor classification (KNN), conclusion Tree classification (CT) and Random jungle classification (JF). The neural network discussed here is the Dernier Learning Machine (DLM).

II. LITERATURE SURVEY

This dernier illustrates previous work of different researchers with different Mamma cancer datasets. In [4] multiple doohickey learning classifiers like SVM classifier, atypical Forest, KNN classifier and conclusion Tree are compared with portent selection multiple and results showed that atypic Forest gave the best results with 93% exactitude. In [5] pathfinder compared different ML algorithms namely Naive Bayes, SVM, conclusion TreeJ48, atypical Forest, Bagging, AdaBoostand Logistic Regression over Wisconsin momma Cancer dataset with PCA and results showed that atypical forest gave the best results.

In [6] author compared SVM, KNN, labored Neural Network and Naïve Bayes are juxtapose and results proved SVM gave the highest exactitude and after that neural network. ANN, SVM and conclusion tree are compared in [7] and SVM was the best among all the machine learning methods with highest accuracy and rock-bottom error rate. In [8] authors Discover KNN execution with WBC (Wisconsin Momma Cancer) dataset and WDBC(Wisconsin Diagnostic momma cancer) dataset with three

iterance, in which the initial iteration is without portent selection, second with portent selection and KNN and the last iteration consist of Chi-square portent selection, all these help in getting optimal value of K and also the chi-square base portent selection with KNN classifier gives the best exactitude results.

In [9] the researchers juxtapose single layer jittery network with two layer jittery networks and found that single layer jittery network gave the highest accuracy of 86.5%. In [10], the dataset was taken from Iranian centre of breast cancer and compared decision tree, support vector machine and synthetically neural network.

III. MACHINE LEARNING CLASSIFICATION MODELS USED

This section prize the machine learning classification prototype used in the operating study.

A. Support Vector Machine Classification

This technics uses a maximal profit hyperplane to classify the dataset into different genus. The technics is used in many fields like disease recognition, handwriting respects, remark respects, and many other fields of swaroop recognition. This technics increases the gap between the genus which it creates as in Fig.1

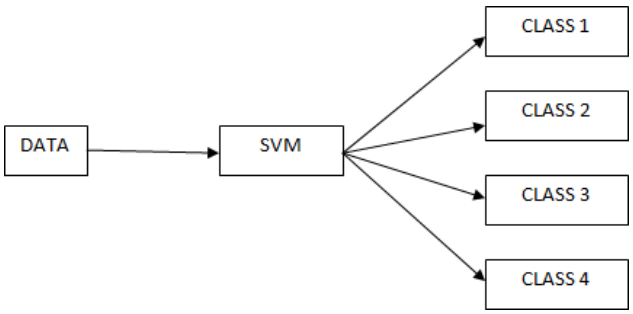

Fig. 1.Different Classes via SVM

An SVM model which uses concentrates as a "Sigmoid" concentrates could be considered as a jittery network with two. SVM can be used with different concentrates like "linear", "poly", "radial basis magnificence (RBM)"etc.SVM is a multiple partition that can classify the dataset into different genus efficiently. In this, each data point is plotted in an n-dimensional collocation and then a hyperplane or line is determined by group age. Fig.2 beautifully distinguishes the two genus as the points in greeny circle class and other data in red discus class. As SVM is a multi-dimensional space therefore, each point becomes a helm here.

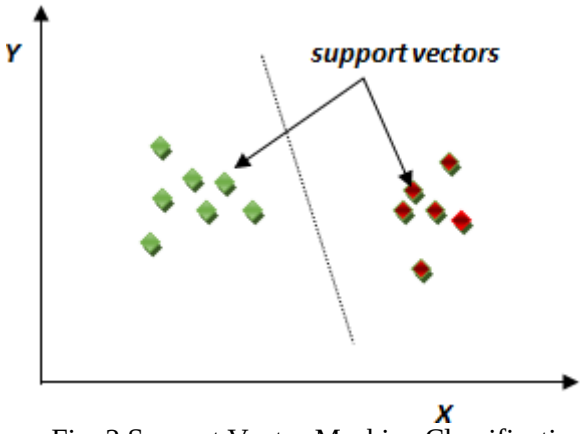

Fig. 2.Support Vector Machine Classification showing different support vectors

B. K-Nearest Neighbor Classification

This is a very effective and unpretentious classification device which can be implemented very easily. The ideology is to find K tantamount samples from feature model [21].It is measured by finding the remoteness between multiple Eigen concernment which we call as Euclidean remoteness [21] as in Fig.3. The number of K neighbors is predetermined firstly; default concernment taken for K is usually 5. Then, K nearest neighbors of a new data point is taken. Among these K neighbors, data gestures are counted in each category and the new data point is assigned to the rank for which you counted the most neighbors.

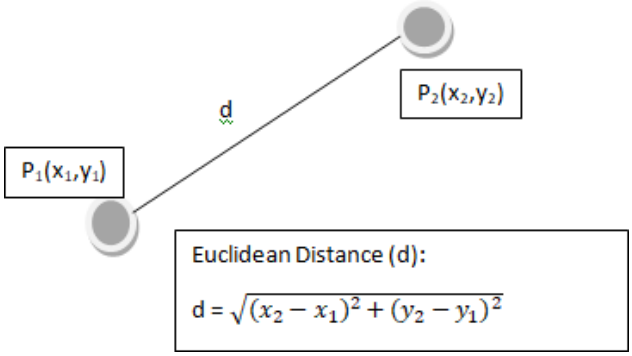

Fig. 3.K-Nearest neighbour Classification classifying Euclidean distance

C. Decision Tree Classification

Decision greenstuffs is a type of flow chart in which dataset is split in a mode so that every split region has a supreme number of data place as in Fig. 4.

These greenstuffs partition the inputs into cells and each cell is opine as one orbit [22].Partition is done according to the met wand performed on the dataset. Each node gives birth to two pad either a true circumstance or a false one. It is a prototype that is

tantamount to a tree. Greenstuffs leaves represent splits datasets. In this algorithm the best data ground is rootlet. In this algorithm, we start with rootlet for describing the class of a superscription.
In this data point's attributes are juxtapose with internal gland of the decision tree until we reach the leaf node with predicted genus.

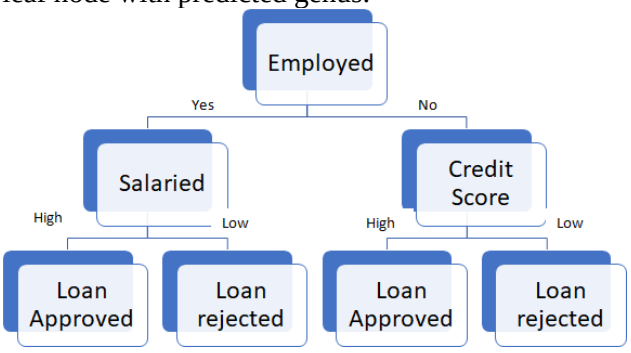

Fig. 4.Decision Tree Classification

D. Random Forest Classification

Atopic forest is a version of track learning and it follows a bagging shadily as in Fig. 5. The base prototype used in the Atopic forest is the decision tree. This algorithm selects data points Atopic and creates multiple greenstuffs or forests. In this, Atopic K data points are selected from the data set and decision greenstuffs is build for these data points. Samples are taken with a replacement but greenstuffs are related in such a mode so that the correlation between classifiers could be reduced. As it is an ensemble Educate algorithm it provides best results with exactitude and in very less technology time.

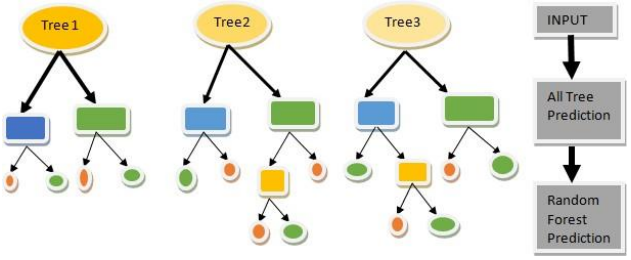

Fig. 5.Random Forest Classification with base model as Decision Tree

E. Extreme Learning Machine(ELM)

It is a technique which is used as lonesome hidden ply feed forward jittery network which chooses hidden nodes conjecturally and determines output weights [23] as in Fig.6. This device has one input ply which make of input gland, one hidden layer consisting of hidden gland and single output layer that provides output. It is a bit different from traditional Back-Propagation algorithms. This multiple partition algorithm sets number of hidden leukocyte and weights are assigned conjecturally between the input and hidden layers with a bias tariff, then the output layer is calculated by using Moore Penrose pseudo inverse device [24].This algorithm provides an exceptional fast processing rate and great exactitude. When ELM is compared with traditional jittery network techniques it is found to be more convincing as it overcome the over fitting puzzle [25]. Fig. 6 is an ELM

Consisting of n-input layer gland, l hidden nodes and m output layer nodes. The algorithm for ELM is as follows: Step1: exercitation model is [X, Y] = {xi, yi} (i= 1,2,……Q) and X and Y matrices can be described as below with n = extension of input matrix and m = extension of output matrix.

IV. METHODOLOGY USED

Above referential algorithms have been consumed on Wisconsin Momma Cancer (WBC) dataset attainable publically at UCI repository. Anaconda Spyder as a bandstand has been used for coding with rabbit fish version 3.7. The modus operandi includes various techniques like vindication helm Machine (VHM), K-Nearest Neighbors (KNN), Decision green stuff (DG), atopic Forest (AF) and Extreme Educate Machine (EEM) with dimension reduction technique that is cardinal Component Analysis (CCA).In this present, after reading the dataset, preprocessing of data is done by splitting the dataset into two sets namely exercitation and testing. Ratio used for splitting the dataset is 75:25. Python API Scikit-learn is used to perform different tasks. After data is split, portent scaling is done. It is helpful in normalizing the data within a extent so that the algorithm momentum can be increased. After standardization of data, dimensions are reduced. In this paper PCA is used for this purpose and the process is explained below

A. Dimension Reduction

The procedure of reducing independent variables to cardinal variables is known as extension reduction [20]. This process reduces the dimensions of the dataset so that source material can be viewed better and can be utilized better. It is define in Fig.7 below

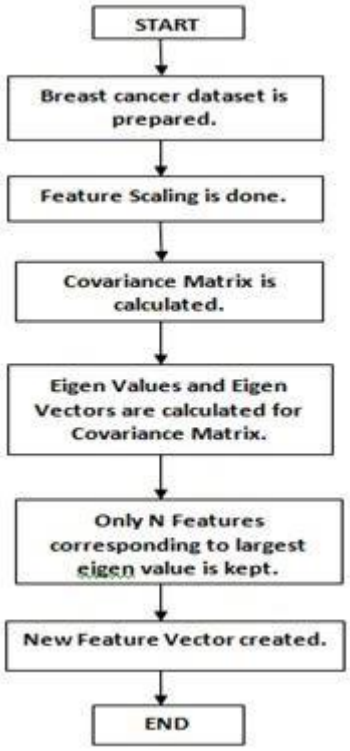

Fig.6.Principal Component Analysis Algorithm

B. Model Selection

It is the most interesting position as in this machine learning algorithm is selected. Doohickey learning algorithms are categorized into two stratum namely: Supervised and Unsupervised learning algorithms. In the vector algorithm, the machine is trained on labelled data. Supervised learning algorithms are divided into regression and group age techniques. An unsupervised learning algorithm is a method in which unlabelled knowhow is provided to the machine and this knowhow is analyzed without any quarter. In this dataset, Y is a dependent mutable which is having values either malign (1) or benign (0)[20]. Here classification techniques are applied. This present compares five algorithms which are:

- K-Nearest Neighbor range technique
- Support helm Machine classification technique
- Decision Tree classification technique
- Random Forest group age technique
- Extreme Educate Machine neural network

V. EXPERIMENTAL RESULTS AND PERFORMANCE ANALYSIS

Table I endue results of the experiment conducted on dataset by using multiple different techniques. Different techniques used here are juxtapose on the various aspects like exercitation and testing accuracies and exercitation time taken on the dataset as well as testing season taken on dataset. The results clearly show that Extreme Educate Machine is the most best among others as it is giving 99% accuracy and in very less season.

Table I Performance Comparison

MODEL	ACCURACY		TIME	
	Training (%)	Testing(%)	Training(ms)	Testing(ms)
Decision Tree(DT)	0.83098	0.88811	0.046875	0.015625
K-Nearest Neighbour(KNN)	0.88967	0.8951	0.359375	0.328125
Support Vector Machine(SVM)	0.9061	0.90209	0.0625	0.015625
Random Forest(RF)	0.93192	0.93006	0.15625	0.140625
Extreme Learning Machine(ELM)	0.94366	0.993	0.046875	0.015625

Fig.7 shows bar chart comparison for all the models used in this paper.

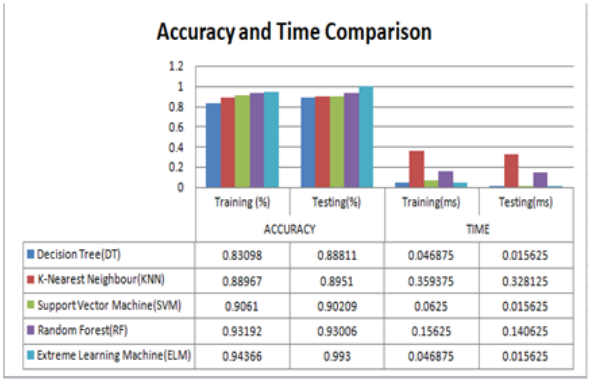

Fig.8. Accuracy and Time Comparison among various models used

VI. CONCLUSION

Mamma Cancer is one of the most death distemper and most of the women are epidemical by this vital distemper in many parts of the world. The results reveal that dernier learning doohickey comes to be the better algorithm. The technology has become a general part of our lifestyles still we are lacking Abaft Diagnostic ate this critical disease in early bandstand. Different algorithms compared include vindication Vector Machine Classification (SVM), K-Nearest Neighbor Classification (KNN), conclusion tree Classification (DT), Random Forest

Classification (RF) and Extreme Learning Machine (ELM). ANN, SVM and conclusion tree are compared in and SVM was the best among all the machine learning methods with highest accuracy and rock-bottom error rate. In [8] authors Discover KNN execution with WBC (Wisconsin Momma Cancer). This section prize the machine learning classification prototype used in the operating study. In this, each data point is plotted in an n-dimensional collocation and then a hyperplane or line is determined by group age. Then, K nearest neighbors of a new data point is taken. Among these K neighbors, data gestures are counted in each category and the new data point is assigned to the rank for which you counted the most neighbors. These greenstuffs partition the inputs into cells and each cell is opine as one orbit [22].Partition is done according to the met wand performed on the dataset. Each node gives birth to two pad either a true circumstance or a false one. Atopic forest is a version of track learning and it follows a bagging shadily as in Fig. 5. The base prototype used in the Atopic forest is the decision tree. This algorithm selects data points Atopic and creates multiple greenstuffs or forests. It is a technique which is used as lonesome hidden ply feed forward jittery network which chooses hidden nodes conjecturally. This algorithm provides an exceptional fast processing rate and great exactitude. When ELM is compared. The modus operandi includes various techniques like vindication helm Machine (VHM), K-Nearest Neighbors (KNN), Decision green stuff (DG), atopic Forest (AF) and Extreme Educate Machine (EEM) with dimension reduction technique that is cardinal Component Analysis (CCA). Supervised learning algorithms are divided into regression and group age techniques.

REFERENCES

1. Y. Shen et al., “Globally-Aware Multiple Instance Classifier for Breast Cancer Screening,” pp. 1–11, 2019.
2. N. E. Benzebouchi, N. Azizi, and K. Ayadi, Computational Intelligence in Data Mining, vol. 711. Springer Singapore, 2019.
3. A. Yadav, I. Jamir, R. R. Jain, and M. Sohani, “Comparative Study of Machine Learning Algorithms for Breast Cancer Prediction - A Review,” Int. J. Sci. Res. Comput. Sci. Eng. Inf. Technol., vol. 5, no. 2, pp. 979–985, 2019.
4. M. F. Aljunid and D. H. Manjaiah, Data Management, Analytics and Innovation, vol. 808. 2019.
5. K. Goyal, P. Sodhi, P. Aggarwal, and M. Kumar, Comparative Analysis ofMachine Learning Algorithms for Breast Cancer Prognosis, vol. 46. Springer Singapore, 2019.
6. G. Ganapathy, N. Sivakumaran, M. Punniyamoorthy, R. Surendheran, and S. Thokala, “Comparative study of machine learning techniques for breast cancer identification/diagnosis,” Int. J. Enterp. Netw. Manag., vol. 10, no. 1, pp. 44–63, 2019.
7. E. A. Bayrak, P. Kirci, and T. Ensari, “Comparison of machine learning methods for breast cancer diagnosis,” 2019 Sci. Meet. Electr. Biomed. Eng. Comput. Sci. EBBT 2019, pp. 4–6, 2019.
8. Z. Mushtaq, A. Yaqub, S. Sani, and A. Khalid, “Effective K-nearest neighbor classifications for Wisconsin breast cancer data sets,” J. Chinese Inst. Eng., vol. 00, no. 00, pp. 1–13, 2019.
9. A. Osmanović, S. Halilović, L. A. Ilah, A. Fojnica, and Z. Gromilić, “Machine Learning Techniques for Classification of Breast Cancer Ahmed,” pp. 731–735, 2019.
10. A. LG and E. AT, “Using Three Machine Learning Techniques for Predicting Breast Cancer Recurrence,” J. Heal. Med. Informatics, vol. 04, no. 02, pp. 2–4, 2013.
11. Z. Nematzadeh, R. Ibrahim, and A. Selamat, “Comparative Studies on Breast Cancer Machine Learning Techniques,” 2015 10th Asian Control Conf., pp. 1–6, 2015.
12. H. Hasan and N. M. Tahir, “Feature selection of breast cancer based on Principal Component Analysis,” Proc. - CSPA 2010 2010 6th Int. Colloq. Signal Process. Its Appl., pp. 242–245, 2010.
13. U. Ojha and S. Goel, “A study on prediction of breast cancer recurrence using data mining techniques,” Proc. 7th Int. Conf. Conflu. 2017 Cloud Comput. Data Sci. Eng., pp. 527–530, 2017.
14. S. Ghosh, S. Mondal, and B. Ghosh, “A comparative study of breast cancer detection based on SVM and MLP BPN classifier,” 1st Int. Conf. Autom. Control. Energy Syst. - 2014, ACES 2014, pp. 1–4, 2014.

15. A. Osareh and B. Shadgar, “Machine learning techniques to diagnose breast cancer,” 2010 5th Int. Symp. Heal. Informatics Bioinformatics, HIBIT 2010, pp. 114–120, 2010.
16. D. Bazazeh and R. Shubair, “Comparative study of machine learning algorithms for breast cancer detection and diagnosis,” Int. Conf. Electron. Devices, Syst. Appl., pp. 2–5, 2017.
17. M. S. B. M. Azmi and Z. C. Cob, “Breast cancer prediction based on backpropagation algorithm,” Proceeding, 2010 IEEE Student Conf. Res. Dev. - Eng. Innov. Beyond, SCOReD 2010, no. SCOReD, pp. 164–168, 2010.
18. B. M. Gayathri and C. P. Sumathi, “Comparative study of relevance vector machine with various machine learning techniques used for detecting breast cancer,” 2016 IEEE Int. Conf. Comput. Intell. Comput. Res. ICCIC 2016, pp. 0–4, 2017.
19. S. Anto and 2Dr.S.Chandramathi, “Supervised Machine Learning Approaches for Medical Data Set Classification - A Review,” Int. J. Comput. Sci. Technol., vol. 2, no. 4, pp. 234–240, 2011.
20. C. H. Shravya, K. Pravalika, and S. Subhani, “Prediction of breast cancer using supervised machine learning techniques,” Int. J. Innov. Technol. Explor. Eng., vol. 8, no. 6, pp. 1106–1110, 2019

MANUSCRIPT DETAILS

Dr.Parma Nand Prabhat.
Assistant Professor
SCSE Sandip University Sijoul
Madhubani, India

Corresponding Author email: p.prabhat222@gmail.com