Prognostic And Diagnostic Prediction Models Of Invasive Cancer In Women And Its Detection Accuracy Via Developing Machine Learning Techniques

Rajesh Kumar Maurya[1], Naveen Kumar[2], Vikas Srivastava[3], Sanjay Kumar[4] and Vimal Kumar[5]

[1]Department of Computer Application, ABES Engineering College, Ghaziabad, U.P- INDIA
[2]School of Computing, DIT University, Dehradun, UK, INDIA
[3]Department of Computer Science and Engineering, MIET Meerut, U.P- INDIA
[4]Department of Computer Science and Engineering, Tula's Institute Dehradun, UK, INDIA
[5]Department of Computer Science and Engineering, Galgotia's University, Greater Noida, U.P, INDIA
[1]rajesh2in@gmail.com , [2]naveen.it23@gmail.com, [3]vikassrivasta@gmail.com, [4]santyagi01@gmail.com, [5]vimalmnnit16@gmail.com

Abstract

Cancer is a common disease, and it takes several human lives every year. It is generally classified on the basis of the tissue from where it originates, such as carcinoma or lymphoma. However, a single tumour may possess several types of heterogeneity. Cancer management is a tedious task due to its heterogeneity. Cancer causes >50% of the total deaths in developing countries due to poor diagnosis. Breast Cancer (BC) is the most heterogeneous cancer and massive molecular data is available on it; however, classification of the information is imperative for its management. Machine learning is an evolving tool that can be used to classify heterogeneous breast cancer datasets.

Objective: Proper classification of tumour diversity and diagnosis of BC using ML approaches can improve the chances of survival; effective prognosis can help clinicians recommend the right treatment.

Methods: Machine learning (ML) classifiers used for classification in the study are Classification and Regression Trees Naive Bayes Classifier, Artificial Neural Networks, Support–Vector Machines and Logistic Regression.

Results: ML is the most efficient tool for classifying the diversity of breast cancer datasets existing with heterogeneity in risk factors. The experimental findings show that the logistic regression model gives the highest accuracy (96.60%), with a lower error rate than other models.

Conclusion: ML appears to be a powerful and practical tool for the categorisation of breast cancer. Massive data concerning breast cancer is available, and crucial features can be extracted through ML. This research paper presents a review of machine learning tools used in the extraction of vital elements of tumour categorisation proposing a prognosis.

Keywords: Cancer, machine learning tools, accuracy, recall, recurrence, survival, prognosis.

1. INTRODUCTION

Each cell in the body divides to support the growth and repair process. The division of a cell follows a process called the cell cycle. Cell division is a highly regulated process, as tight regulation of the cycle helps maintain an appropriate number of cells in the body, and failure of cell division control results in the formation of a tumour. DNA refers to the genetic material in all eukaryotic cells, and the genetic defect is the cause of uncontrolled cell proliferation that produces cancer. Genes are the regions on the DNA where information for cell structures and functions is coded. The flow of genetic information starts from the DNA to the RNA and finally to proteins. Particular gene products act as guardians and take care of the expression of other genes, such as genes controlling cell proliferation.

Among cancers, breast cancer (BC) is considered the most heterogeneous one. BC is a major cause of death in women worldwide. The mutation is the cause of the defect in genetic material and the production of the altered gene product. Mutation results in either gain or loss of the function of genes. The modified gene product fails to control cell division, leading to the formation of tumours or benign tumours. Later on, the constituent cells of tumour mass gain the ability to invade the neighbouring tissue and occupy distant places in the body. The tumour mass is called a malignant tumour—the malignant tumour mass is composed of a different type of cells. Breast cancer is an epithelial cell-derived tumour known as carcinoma. The epithelial cells comprise the outer and inner boundaries of the organs. A healthy immune system fights cancer cells, but when these cells escape immune surveillance, tumours are formed. Since the genetic composition of each cell is different, the susceptibility to cancer varies across a population, causing stress [1–4]. Moreover, BC is highly dependent on the female hormones oestrogen and prolactin.

The molecular data extracted by researchers using the microarray (MA) technique, protein array (PA) and sequencing of DNA and RNA provide a vast platform for understanding the heterogeneity of the BC dataset to limit the manual classification limit. Evolving computational techniques are now fulfilling this need. Computational methods are novel tools for analysing a complex set of data and might be a helpful option for clinicians [5].

1.1 Importance of Machine Learning (ML) Tools for BC symptom Analysis

BC is a life-threatening disease, especially in women. In women, BC is not only a disease but also a social issue. The heterogeneity of BC makes its diagnosis and management more difficult than that of any other cancer. Error-free diagnosis is the key to cancer management. ML methods are evolving tools that meet the demand of clinicians by providing early error-free diagnosis, and the predictions of a BC ML analysis could improve the chances of survival for BC patients and help maintain the pace of the treatment [6]. ML algorithms are a set of instructions that analyse and forecast the type of cancer based on records. The goal of ML algorithms in cancer is to evolve into a diagnostic evaluation algorithm that can provide a timeline of the cancer patient, which can accurately predict the disease and its treatment. The main issue with ML algorithms pertains to the diversity of data and types of cases reported. The effectiveness of ML tools can be determined by comparing the performance of two more algorithms [7, 8].

2. ML ALGORITHM FOR CANCER PREDICTION

Healthcare is critical for developing as well as developed nations. The diagnosis of a disease navigates the path of treatment. Even though massive molecular and clinical data are available, we have a limited prediction and prognosis mechanism. ML classifies, arranges and extracts available data sets and provides a pattern for clinical progress. The advantages of ML tools in healthcare include the identification of diseases, treatment innovation, medical imaging diagnosis, personalised medicine, ML-based behavioural change, health records, research, clinical trials, data collection and improved radiotherapy to support medical practitioners [9]. Several BC forecast models have been designed and proved useful to discover the characteristics of a disease as well as to diagnose and predict treatment using ML and statistical techniques. ML involves data mining algorithms that can be utilised by a healthcare expert to develop new kinds of algorithms. The algorithms can provide BC patients' dataset to machines that can analyse

clinical outcomes more precisely. The following practical ML tools, based on filtering and classification techniques, are widely used to analyse massive datasets of cancer patients:

- Naive Bayes (NB)
- Logistic regression (LR)
- Support-vector machines (SVM)
- Artificial neural networks (ANN)
- Classification and Regression Trees (CART)

These tools present the right treatment option, which can help minimise deaths due to cancer [10, 11]. Proper investigation and the timely detection of cancer via ML tools can support the doctors in protecting the patient. ML applications are very helpful in categorising multifarious records correctly. Consequently, it offers hope to cancer patients. Prediction tools create a dataset that would be very useful to surgeons in monitoring the performance of a drug against BC and precautionary measures and remedies depending on a patient's response to medication. Useful investigation tools have been developed to identify hidden relationships in BC records [12–14]. The essential steps in the classification of BC dataset are as follows:

a. Collection of BC dataset
b. Purification of a redundant dataset
c. Selection of suitable ML techniques
d. ML technique for BC dataset.
e. Assessment of the efficiency of ML classification techniques with regard to BC
f. Prediction of cancer (susceptibility, recurrence and susceptibility)

2.1 Naive Bayes (NB)

NB is the oldest statistical classification method based on probabilistic algorithms and so-called Bayesian classifiers for the cataloguing of real-life problems such as BC [15].

$$P\left(\frac{L}{N}\right) = P\left(\frac{N}{L}\right)P(L)/P(N) \qquad (1)$$

Equation (1) represents the NB classifier where **L** belongs to a certain class, **N** is the observed data points, **P (L/N)** is the posterior probability of target given attributes, **P (N/L)** is likelihood, **P (L)** is the class prior probability and **P(N)** is prior probability of predictor. In 2012, NB classification was used to study the WBCD dataset, obtaining an efficiency of 95.99% [17]. In 2015, it was used to study the WBCD dataset, obtaining an efficiency of 99.11% (specificity), 98.25% (sensitivity) and 98.54% (accuracy) [18]. In 2019, the BC classification method was used to study the WBCD dataset, obtained 95.7% (accuracy), 97.2% (precision) and 97.1% (recall) [19]. An NB approach used for the classification of a malignant tumour (MT) and benign tumour is shown in Fig. (1), where P1, P2, P3, P4…PN represents the probability of the occurrence of cancerous features.
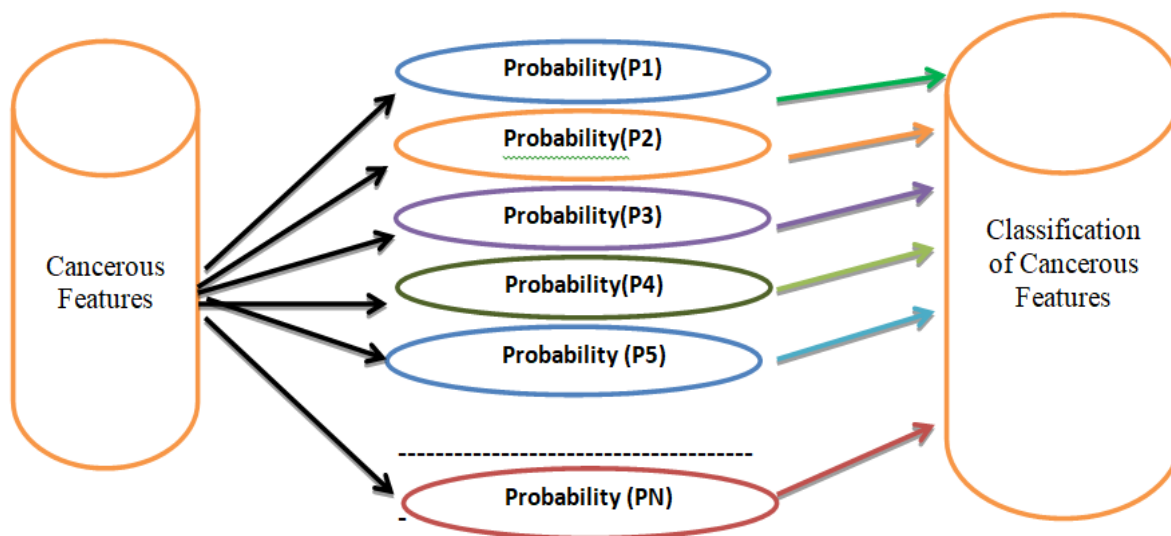
Fig.(1). Working of NB Classifier on WDBC dataset

2.2 Logistic regression

The LR method is like other statistical regression techniques commonly employed to identify the relationship among the variables. In 1958, David Cox, a statistician, designed a supervised ML classification technique for forecasting. LR is one of the most commonly used algorithms that simplifies the idea of a predictive analysis to conditions where the resultant variable is certain. This algorithm usually emphasises a class of binary classification (values between 0 and 1) of problems and produces a logistic curve.

$$P1\,(Y=1): \text{Forecast Chances}\,(\,0 < P1 < 1\,) \tag{2}$$

$$0 - \text{Negative Class(for examle Benign Tumor)},$$

$$1 - \text{Positive Class(for example Malignant Tumor)}$$

Here, the outcome labels are projected in terms of benign/malignant BC with 59.1% accuracy on WPBC and 95.24% accuracy on WDBC using this method [20, 21]. For studying and forecasting in-patient deaths of cancer patients, an LR-based examination with cluster amendments was implemented [22]. A sample approach, as shown in Fig. (2), was used for the forecasting of a malignant or benign tumour using the LR technique. Equation (2) represents forecast the chances. Equation (3) represents the linear model, and Equation (4) represents the logistic model. In Equation

(3), α_0 is constant and α_1 is a slope. In Equation (4), the symbol β represents the LR coefficient. In Fig. (2), the emblem x represents tumour size (input) and y (0 or 1) represents the type of tumour – malignant or benign (output).

$$P1\bigl(Y = 1\bigr) = \alpha_0 + \alpha_1 x \tag{3}$$

$$P1\bigl(Y = 1\bigr) = \frac{e^{\beta_0} + e^{\beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}} \tag{4}$$
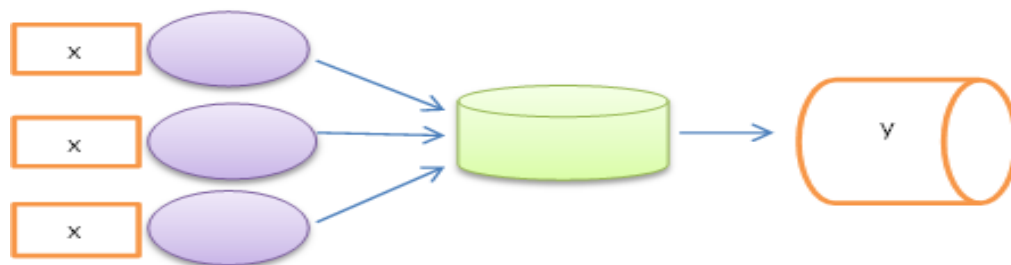
Fig.(2). A semantic approach shown for tumour (benign and malignant) classification using the LR classifier

2.3 Support-vector Machine

SVM classifiers are useful when a researcher has no idea about the dataset. This method, which uses a supervised class of a classifier, is one of the most effective in finding the best linear separator between two features. The classifier tries to make a hyper-plane or decision boundary in such a way that the separation between the two sections – Class A and Class B – is as wide as possible. The idea is to find a hyper-plane that enables the finest separation of a dataset into two sections. The most significant distance and classes fall between adjacent data points of margin. The objects belong either to Class A or B [23]. One approach shown in Fig. (3) is used for forecasting malignant and benign tumours.
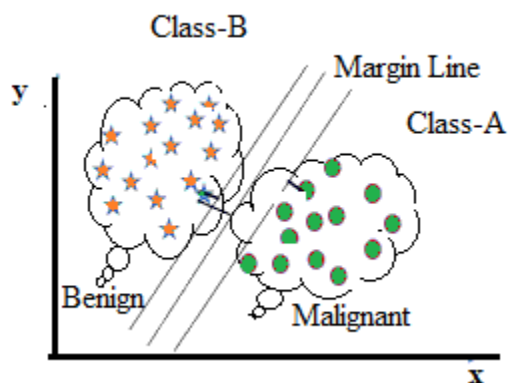
Fig.(3). A semantic approach for classifying benign and malignant tumours using the SVM technique

The performance of the SVM classifier on WBCD datasets was evaluated: 96.2927% accuracy (9-attribute sample), 96.5610% accuracy (4-attribute sample) and 96.5610 % accuracy (13-attribute sample) were obtained [24]. When this method was used to study the cancer dataset, 96.6 % accuracy was achieved [25]. The SVM method has been applied to study the WBCD dataset, obtaining 97.2% accuracy. When Least Square SVM (LS-SVM) method of WBCD dataset was used, 98.53 % accuracy was obtained [26]. When the SVM method was applied on WBCD dataset, 97.7 % accuracy was obtained [27]. Using the LSVM method, QSVM and FG-SVM methods on WBCD dataset, 93.3%, 92.7% and 91.3% accuracy, respectively, were obtained [28]. This technique was applied to study the WBCD dataset with 97.2% accuracy while using SVM and 98.53 % accuracy while using the LS-SVM method [26]. Using this technique to study the WBCD records, 97.7 % accuracy [27] were obtained, with 93.3% efficiency using LSVM, 92.7% using QSVM and 91.3% using FG-SVM [28].

2.4 Classification and Regression Tree

Leo Breiman developed the CART method. He was an eminent professor and statistician at the University of California, Berkeley. This method involves decision tree (DT) procedures used for feature selection problems and classification. CART is a compelling approach with significant potential and clinical utility. In this method, records (cancer patients) can be divided into several groups based on their characteristics. A tree-like arrangement is used in DTs, where the leaves of the tree represent the classification criteria and the sprig symbolises causes [29]. CART is used to predict the disease outbreak in medical organisations for finding the cure of illnesses [30, 31]. The approach, shown in Fig. (4), is used for the forecasting of malignant and benign tumours.
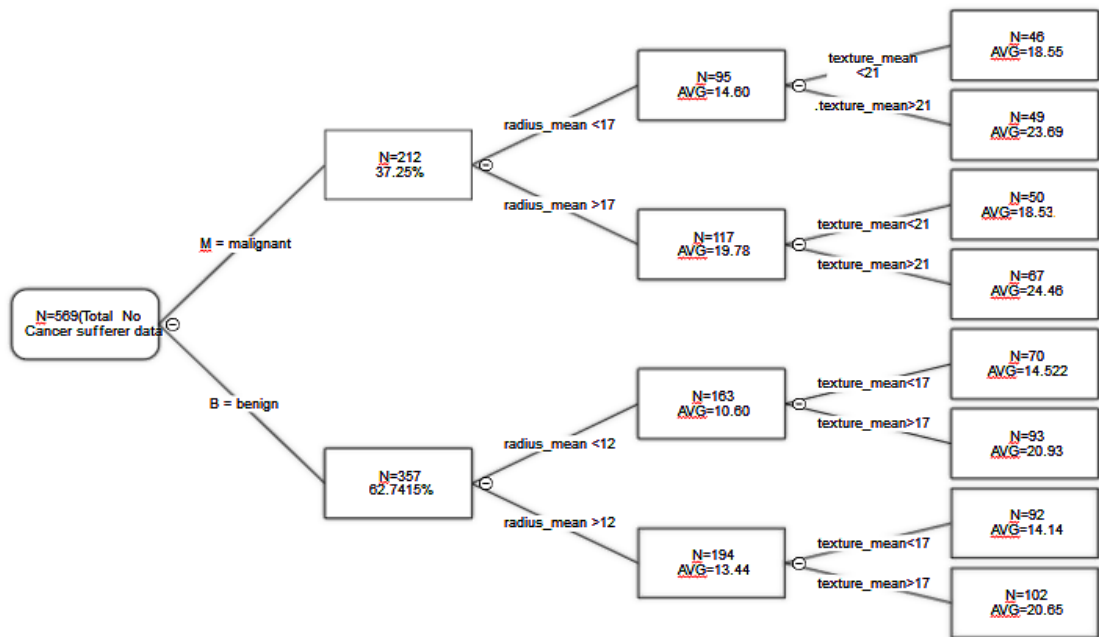

Fig.(4). An example of a decision tree for BC

2.5 Artificial Neural Network

In 1960, a mathematical model was developed by Professors Bernard Widrow and Ted Hoff. The ANN model mimics the structure and function of the human brain for processing information. The method was helpful in the classification and prediction of chronic diseases like BC. ANN can easily show straightforward relationships of BC. This algorithm was designed and implemented in a particular application, such as pattern cataloguing from end to end using an ML procedure. The linking arrangement designed within different layers is called the edges of the ANN model. The arrangement of neurons form three layers:

       a)   Input layer
       b)   Output layer
   c)   Intermediate layer

The neurons are interconnected with edges, and each side is associated with their vertex, which is called 'wait'. ANNs are sets of connected neurons [32, 33]. Using the probabilistic neural network (PNN) technique to analyse the BC dataset [34, 35], 96% accuracy was obtained. An ANN approach is shown in Fig. (5), used for the forecasting of malignant and benign tumours.
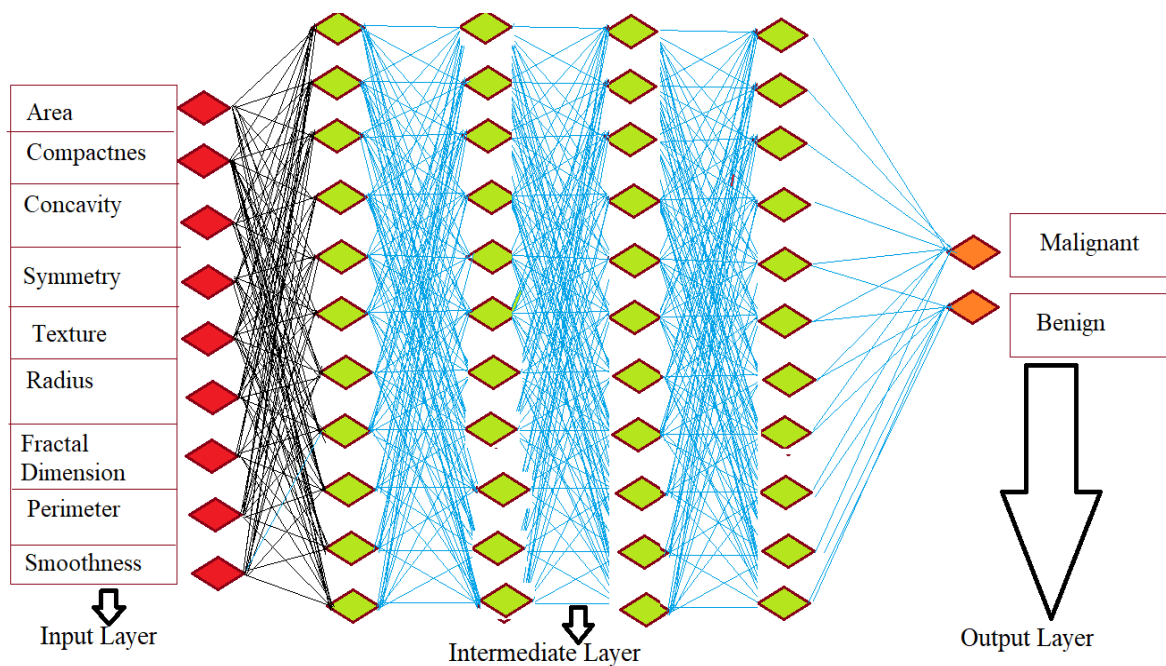
Fig.(5). A semantic model for classifying in what way ANN is trained to predict using different layers for results

The ANN model was used on the labelled dataset, and the train test split was used to check the precision of the classification model. The data was split into an 80:20 ratio for train and test. A neural network design was developed with three hidden layers with 64, 32 and 16 nodes. The activation function sigmoid was used in hidden layers and soft-max was used in the output layer. The input layer had 29 nodes equivalent to the number of features. The record ID was not considered as an input feature as it did not contain any feature value. Upon testing it with 1000 epochs, the accuracy was around 96.49%. The AUC-ROC curve, which is a measure of model accuracy for ANN, is depicted in Fig. (6)[66-67].
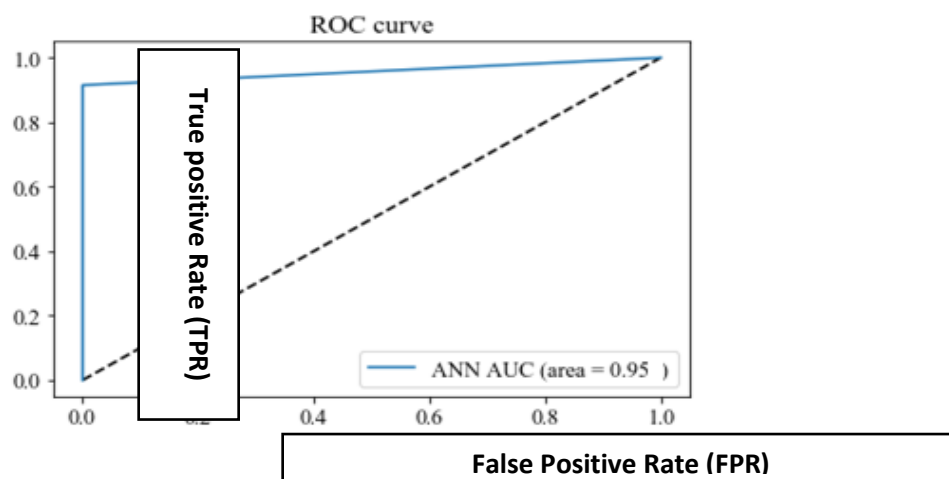
Fig.(6). AUC-ROC curve for model peformance of ANN

3. BC PROGNOSIS

The prognosis and forecasting of BC growth concentrates on three significant areas: risk assessment, forecast of BC relapse and a forecast of the BC endurance rate. The area includes a projection of the possibility of increasing the cancer symptoms before the BC patient is diagnosed [36]. Chemotherapy for early-stage symptoms BC (for instance, changes in bowel, urination difficulties, habits, testicular changes, weight loss, breast lumps, skin, mouth sores, persistent cough, nausea, stomach pain, etc.) is often managed after breast surgery. The end of the prognosis procedure of BC is entirely dependent on the quality of the remedial treatment. The prognostic predictions here are supplementary, as opposed to those from unpretentious problem-solving inspection. The study of the prognostic efficiency of ML-based tools would provide visual valuations to those of established clinical predictors The study of the prognostic efficiency of ML-based tools is compared with that of the established clinical predictors and visual valuations [37–39]. Among the various ML classifiers, SVM has been found to provide higher accuracy [40]. The gene mutations of cancerous cases profiles has been effectively used with ML approaches to discover clinically clear separation of BC records [41].

3.1 BC Susceptibility

BC is the most significant public malignancy in the world. Reproductive genes for cancer are hereditary factors, which cause a genetic disturbance in the breast, and mutations in these genes raises the associated possibility of developing BC [42]. Breast is a highly dynamic tissue that is affected by lactation period, age and sex. It is partial by lactation period, age and sex, and this makes it difficult to follow the syndrome after the next medical treatment. It also produces a substantial amount of heterogeneity in BC statistics [43, 44]. This poses a considerable challenge for the diagnosis of the disease. ML offers dynamic techniques for classifying genetic variability. Genetic analysis and evaluation in an extensive series of BC patients can address the actual possibilities of cancer [45, 46].

3.2 BC Recurrence

The migration of cancer cells to a distant location in the body is known as metastasis. In-transit metastatic or locally recurrent BC can come back after a month or a year after the treatment and surgery in the same place in the body, and this is known as repeated BC. The doctor typically removes the lumps, which are identified by the pathologist based on the laboratory tests. If the signs of the physical examination are abnormal, the pathological consequences can be confident or uncertain in terms of the metastases. Typical signs include nipple discomfort, secretion, growing shape, inverted nipple, new shape or size of breast nipples, weight loss, shortness of breath and bone pain [47, 48]. The ML-based methods of feature selection and classification can be used to develop modules for the prediction of relapse of cancer [49, 50]. With the combined efforts of cancer surgeons and scientists' notes on various recent techniques for BC, the key future trends of computational biology seem to have obtained positive results with admirable correctness [51].

3.3 BC Survival

The goal of clinicians and researchers is to increase the healthy life expectancy of BC patients. With the considerable advancement in the number of cancer surgeries available for breast cancer and the number of experienced surgical and breast radiation teams, there is a possibility of diagnosis of the disease, and with them, the life-sustaining ratio is minimal. Microarray technology is one of the tools used to estimate the survival time of lymphoma patients [52, 53]. The determination of BC's sensitivity using effective medical findings is essential in improving care and providing necessary details of the disease to doctors and patients [3, 54]. Classification models to detect strategies and genetic characteristics, which allow treatment responses and the overall existence of patients and are an excellent way to differentiate between surviving categories [55, 56], are a result of computational development and design. In recent

years there has been a great promise of equipment and related appearances from various ML classifiers, signifying that such techniques are capable of representing high-quality learning and non-sequential characteristics [57].

4. EXPERIMENTAL BC DATASET

The BC database used in this manuscript has been taken from the learning store, i.e.,UCI Machine Learning Repository, which was developed by Dr. William H. Wolberg. A number of researchers and biologists have shown the efficiency of the ML Classifier on WBCD in terms of various analytical factors. The analysed WDBC has a total of 464 records of BC cases with 62% (342 – benign) and 38% (212 – malignant) tumours [58]. Factors of the BC dataset include perimeter, radius, texture, concavity, area, concave points, diagnosis (malignant/benign), smoothness, compactness and symmetry, as shown in Table 1.

Table **1.** Attributes of WDBC.

S.N	Range	Parameters
A1	Patients ID	ID number
A2	A-Malignant, B-Benign	Diagnosis
A3	6–28	Radius
A4	9–40	Texture
A5	43–188	Perimeter
A6	143–2500	Area
A7	0.05–.10	Smoothness
A8	0.19–.35	Compactness
A9	0–.43	Concavity
A10	0–.20	Concave points
A11	0.1–0.3	Symmetry
A12	.04–.09	Fractal dimension

5. EXPERIMENTAL ENVIRONMENT AND RESULTS

The following steps are involved in the assessment of a dataset with ML tools: a) construction of an ML model followed by the recovery of data from the BC dataset and b) performing of a test for equating the performances of different ML tools such as NB, LR, CART, ANN and SVM on BC dataset in the python programming environment[59-60].

5.1 Programming Environment

Python is the most frequently used programming language, and it was developed in the 1980s by Guido van Rossum[65]. Cancer patients' entire experimental records have been evaluated with the help of Python in an ML programming background that has an excellent library of ML tools for classification. Python libraries and its applications offer a unique development environment for programming and the needs of future and current work in bioinformatics estimate [63-64].

5.2 Results

Table 2 demonstrates the performances of five classification models on the patient's entire dataset. Here, the dataset investigation done using five classification models and their performance have been tabulated. For implementing the models, the data has been summarised on the BC dataset. For evaluating and training the five ML classifiers, we

divided the training dataset into different sections by performing a 10-fold cross-validation scheme. We then examined the performance of the five models in terms of standard deviation (SD), meantime and accuracy, as shown in Table 2 and Fig. (7) [61-62].

Table **2.** Shown efficiency of five classifiers on WDBC .

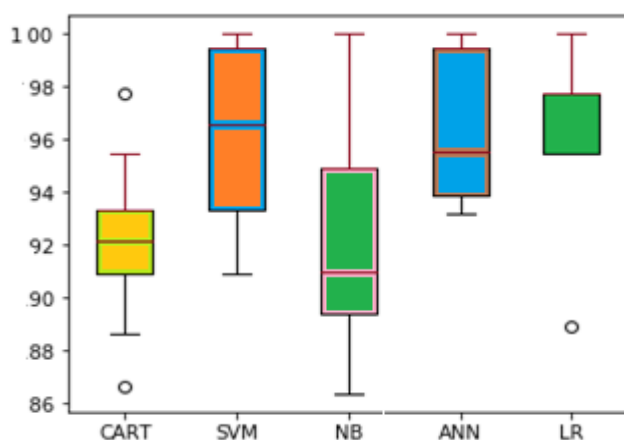| Assessment measures	CART	NB	SVM	ANN	LR
Run time	03.15	00.20	00.24	07.77	01.02
SD	00.03	00.038	00.03	00.027	00.03
Accuracy (%)	92.11	92.10	96.16	95.70	96.60

Fig. (7). Performance analysis of 5-ML Models

6. CONCLUSION

This manuscript briefly discussed the utility of unsupervised and supervised ML tools to investigate the number of BC species and their impediments. An early opinion on BC is imperative in the first stage for proper handling of the disease. Amended cataloguing and forecast accuracy of ML techniques and their usefulness in the BC forecast, the prognosis can cut the loss-of-life ratio all over the world. The ML-based expanded analysis of the study emphasises the development and design of correctness and the highly efficient risk prediction of models using ML tools. Even though a number of right ML procedures have obtained a very high degree of accuracy on WBCD, there is a massive requirement of value-added quality tools to redeem the life of BC patients. Usability and application of the right ML tools, in-depth understanding of the classification approach for cancerous pattern study and detection of multidimensional genetic records integration are excellent ways to achieve thoughtful incorrect proficiency diagnosis, prognosis and predictions of BC [63-64]

7. CURRENT AND FUTURE DEVELOPMENTS

In cancer, several types of genomic disorders occur to produce various kinds of phenotypes. Therefore, it is difficult to categorise a particular cancer patient for therapeutics and prognosis. In this context, ML is a developing tool that provides an opportunity to classify a complex set of available data for better prediction. Although these kinds of ML tools are in use for biomedical research, prognosis and therapeutic purposes, there is a need to further refine these tools for improved classification of data for future use.

8. Availability of data and materials: The BC database used in this research has been taken from the learning store, i.e., the UCI Machine Learning Repository, which was developed by Dr. William H. Wolberg [58].

References

1. J. Baselga, and L. Norton, "Focus on breast cancer", Cancer Cell, vol. 1, no. 4, pp. 319-322, 2002.
2. S.J. Greco, "Breast cancer risk in a rapidly aging population: advances and approaches to study the aging tissue microenvironment. Breast Cancer: Targets and Therapy, 11, 111. Bombonati, A., & Sgroi, D. C. (2011). The molecular pathology of breast cancer progression", J. Pathol., vol. 223, no. 2, pp. 308-318, 2019.
3. C.M. Lynch, B. Abdollahi, J.D. Fuqua, A.R. de Carlo, J.A. Bartholomai, R.N. Balgemann, V.H. van Berkel, and H.B. Frieboes, "Prediction of lung cancer patient survival via supervised machine learning classification techniques", Int. J. Med. Inform., vol. 108, pp. 1-8, 2017.
4. F. Chiriac, A. Baban, and D.L. Dumitrascu, "Psychological stress and breast cancer incidence: a systematic review", Cluju1 Med., vol. 91, no. 1, pp. 18-26, 2018.
5. M.W. Huang, C.W. Chen, W.C. Lin, S.W. Ke, and C.F. Tsai, "SVM and SVM ensembles in breast cancer prediction", PLoS One, vol. 12, no. 1, 2017.e0161501
6. D. Bazazeh, and R. Shubair, "Comparative study of machine learning algorithms for breast cancer detection and diagnosis",
7. P. Suryachandra, and P.V.S. Reddy, "Comparison of machine learning algorithms for breast cancer", 2016 International Conference on Inventive Computation Technologies (ICICT), vol. Vol. 3, , 2016pp. 1-6
8. A.M. Priya, K.E. Kannammal, S.P. Kavya, R. Iswarya, and M. Nivetha, "Identify Breast Cancer Using Machine Learning Algorithm. Journal Homepage, 2019", http://www. ijesm. co
9. F. Maes, D. Robben, D. Vandermeulen, and P. Suetens, The Role of Medical Image Computing and Machine Learning in Healthcare.Artificial Intelligence in Medical Imaging.. Springer: Cham, 2019, pp. 9-23.
10. J.T. Wang, M.J. Zaki, H.T. Toivonen, and D. Shasha, Introduction to data mining in bioinformatics.Data Mining in Bioinformatics.. Springer: London, 2005, pp. 3-8.
11. H. Wang, L. Huang, R. Jing, Y. Yang, K. Liu, M. Li, and Z. Wen, "Identifying oncogenes as features for clinical cancer prognosis by Bayesian nonparametric variable selection algorithm", Chemom. Intell. Lab. Syst., vol. 146, pp. 464-471, 2015.
12. R. Jing, Y. Liang, Y. Ran, S. Feng, Y. Wei, and L. He, "Ensemble Methods with Voting Protocols Exhibit Superior Performance for Predicting Cancer Clinical Endpoints and Providing More Complete Coverage of Disease-Related Genes", Int. J. Genomics, vol. 2018, 2018.8124950
13. N. Borisov, V. Tkachev, M. Suntsova, O. Kovalchuk, A. Zhavoronkov, I. Muchnik, and A. Buzdin, "A method of gene expression data transfer from cell lines to cancer patients for machine-learning prediction of drug efficiency", Cell Cycle, vol. 17, no. 4, pp. 486-491, 2018.
14. O. Hamzeh, A. Alkhateeb, and L. Rueda, "Predicting Tumor Locations in Prostate Cancer Tissue Using Gene Expression", International Conference on Bioinformatics and Biomedical Engineering, , 2018pp. 343-351
15. S. Kharya, S. Agrawal, and S. Soni, "Naive Bayes classifiers: A probabilistic detection model for breast cancer. International Journal of Computer Applications, 2014, 92(10), 0975-8887",
16. A. Wood, V. Shpilrain, K. Najarian, and D. Kahrobaei, "Private naive bayes classification of personal biomedical data: Application in cancer data analysis", Comput. Biol. Med., vol. 105, pp. 144-150, 2019.
17. G.I. Salama, M. Abdelhalim, and M.A.E. Zeid, "Breast cancer diagnosis on three different datasets using multi-classifiers", Breast Cancer, vol. 32, no. 569, p. 2, 2012. [WDBC].
18. M. Karabatak, "A new classifier for breast cancer detection based on Naïve Bayesian", Measurement, vol. 72, pp. 32-36, 2015.
19. A. Yadav, I. Jamir, R.R. Jain, and M. Sohani, "Comparative Study of Machine Learning Algorithms for Breast Cancer Prediction-", RE:view, 2019.
20. P. Ferroni, F.M. Zanzotto, S. Riondino, N. Scarpato, F. Guadagni, and M. Roselli, "Breast Cancer Prognosis Using a Machine Learning Approach", Cancers (Basel), vol. 11, no. 3, p. 328, 2019.
21. A. Sharma, S. Kulshrestha, and S. Daniel, "Machine Learning approaches for breast cancer diagnosis and prognosis",
22. J. Upshaw, R. Ruthazer, K. Miller, S. Parsons, J. Erban, and A. O'Neill, "A TOOL FOR PERSONALIZED DECISION MAKING: A CLINICAL PREDICTION MODEL FOR ANTHRACYCLINE CARDIOTOXICITY IN EARLY STAGE BREAST CANCER", J. Am. Coll. Cardiol., vol. 71, no. 11, p. A696, 2018.
23. Y. Jiang, J. Xie, Z. Han, W. Liu, S. Xi, L. Huang, and J. Yu, Immunomarker Support Vector Machine Classifier for Prediction of Gastric Cancer Survival and Adjuvant Chemotherapeutic Benefit., 2018.
24. K. Polat, and S. Güne¸s, "Breast cancer diagnosis using least square support vector machine", Digit. Signal Process., vol. 17, pp. 694-701, 2007.
25. K.P. Bennett, and J.A. Blue, "A support vector machine approach to decision trees", Proceedings of the IEEE International Joint Conference on Neural Networks, vol. Volume 3, , 1998pp. 2396-2401
26. J.A.K. Suykens, and J. Vandewalle, "Least squares support vector machine classifiers", Neural Process. Lett., vol. 9, pp. 293-300, 1999.

27. V. Chaurasia, S. Pal, and B.B. Tiwari, "Prediction of benign and malignant breast cancer using data mining techniques", J. Algorithm. Comput. Technol., vol. 12, no. 2, pp. 119-126, 2018.

28. R. Karthiga, and K. Narasimhan, "Automated Diagnosis of Breast Cancer Using Wavelet Based Entropy Features", 2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA), , 2018pp. 274-279

29. R.J. Lewis, "An Introduction to Classification and Regression Tree (CART) Analysis", Annual Meeting of the Society of Academic Emergency Medicine in, 2000

30. H. Leopord, W.K. Cheruiyot, and S. Kimani, "A survey and analysis on classification and regression data mining techniques for diseases outbreak prediction in datasets", Int. J. Eng. Sci., vol. 5, no. 9, pp. 1-11, 2016.

31. S. Vural, X. Wang, and C. Guda, "Classification of breast cancer patients using somatic mutation profiles and machine learning approaches", B.M.C. Syst. Biol., vol. 10, no. 3, suppl. Suppl. 3, p. 62, 2016.

32. S. Agrawal, and J. Agrawal, "Neural network techniques for cancer prediction: A survey", Procedia Comput. Sci., vol. 60, pp. 769-774, 2015.

33. H. Shaish, S. Mutasa, J. Makkar, P. Chang, L. Schwartz, and F. Ahmed, "Prediction of Lymph Node Maximum Standardized Uptake Value in Patients With Cancer Using a 3D Convolutional Neural Network: A Proof-of-Concept Study", A.J.R. Am. J. Roentgenol., vol. 212, no. 2, pp. 238-244, 2019.

34. S. Swathi, B.G. Anjan, R.K. Sendhil, and S.N. Bhukya, "Performance of ART1 Network in the Detection of Breast Cancer", 2nd International Conference on Computer Design and Engineering (ICCDE 2012), 2015 Singapore

35. K. Ajay, R. Sushil, and A. Tiwari, "Cancer Survival Analysis Using Machine Learning. 2019, Available at SSRN 3354469",

36. I. Mihaylov, M. Nisheva, and D. Vassilev, "Application of Machine Learning Models for Survival Prognosis in Breast Cancer Studies", Information, vol. 10, no. 3, p. 93, 2019.

37. S. Mani, Y. Chen, X. Li, L. Arlinghaus, A.B. Chakravarthy, V. Abramson, S.R. Bhave, M.A. Levy, H. Xu, and T.E. Yankeelov, "Machine learning for predicting the response of breast cancer to neoadjuvant chemotherapy", J. Am. Med. Inform. Assoc., vol. 20, no. 4, pp. 688-695, 2013.

38. K. Kourou, T.P. Exarchos, K.P. Exarchos, M.V. Karamouzis, and D.I. Fotiadis, "Machine learning applications in cancer prognosis and prediction", Comput. Struct. Biotechnol. J., vol. 13, pp. 8-17, 2014.

39. S. Aruna, and S.P. Rajagopalan, "A novel SVM based CSSFFS feature selection algorithm for detecting breast cancer", Int. J. Comput. Appl., vol. 31, no. 8, 2011.

40. P.H. Abreu, M.S. Santos, M.H. Abreu, B. Andrade, and D.C. Silva, "Predicting breast cancer recurrence using Machine Learning techniques: a systematic review", A.C.M. Comput. Surv., vol. 49, no. 3, p. 52, 2016. [CSUR].

41. L. Falzone, S. Salomone, and M. Libra, "Evolution of cancer pharmacological treatments at the turn of the third millennium", Front. Pharmacol., vol. 9, p. 1300, 2018.

42. A. Mehrgou, and M. Akouchekian, "The importance of BRCA1 and BRCA2 genes mutations in breast cancer development", Med. J. Islam. Repub. Iran, vol. 30, p. 369, 2016.

43. W. Kim, K.S. Kim, J.E. Lee, D.Y. Noh, S.W. Kim, Y.S. Jung, M.Y. Park, and R.W. Park, "Development of novel breast cancer recurrence prediction model using support vector machine", J. Breast Cancer, vol. 15, no. 2, pp. 230-238, 2012.

44. B.J. Kim, and S.H. Kim, "Prediction of inherited genomic susceptibility to 20 common cancer types by a supervised machine-learning method", Proc. Natl. Acad. Sci. U.S.A., vol. 115, no. 6, pp. 1322-1327, 2018.

45. J. Xu, R. Jing, Y. Liu, Y. Dong, Z. Wen, and M. Li, "A new strategy for exploring the hierarchical structure of cancers by adaptively partitioning functional modules from gene expression network", Sci. Rep., vol. 6, p. 28720, 2016.

46. P. Apostolou, and F. Fostira, "Hereditary breast cancer: the era of new susceptibility genes", BioMed Res. Int., vol. 2013, 2013.747318

47. W. Yue, Z. Wang, H. Chen, A. Payne, and X. Liu, "Machine Learning with Applications in Breast Cancer Diagnosis and Prognosis", Designs, vol. 2, no. 2, p. 13, 2018.

48. F. Bertucci, and D. Birnbaum, "Reasons for breast cancer heterogeneity", J. Biol., vol. 7, no. 2, p. 6, 2008.

49. B.A. McKinney, D.M. Reif, M.D. Ritchie, and J.H. Moore, "Machine learning for detecting gene-gene interactions: a review", Appl. Bioinformatics, vol. 5, no. 2, pp. 77-88, 2006.

50. M.D. Ritchie, J.R. Davis, H. Aschard, A. Battle, D. Conti, M. Du, E. Eskin, M.D. Fallin, L. Hsu, P. Kraft, J.H. Moore, B.L. Pierce, S.A. Bien, D.C. Thomas, P. Wei, and S.B. Montgomery, "Incorporation of biological knowledge into the study of gene-environment interactions", Am. J. Epidemiol., vol. 186, no. 7, pp. 771-777, 2017.

51. Y.C. Chen, W.C. Ke, and H.W. Chiu, "Risk classification of cancer survival using ANN with gene expression data from multiple laboratories", Comput. Biol. Med., vol. 48, pp. 1-7, 2014.

52. A. Moslemi, H. Mahjub, M. Saidijam, J. Poorolajal, and A.R. Soltanian, "Bayesian survival analysis of high-dimensional microarray data for mantle cell lymphoma patients", Asian Pac. J. Cancer Prev., vol. 17, no. 1, pp. 95-100, 2016.

53. A. Bashiri, M. Ghazisaeedi, R. Safdari, L. Shahmoradi, and H. Ehtesham, "Improving the Prediction of Survival in Cancer Patients by Using Machine Learning Techniques: Experience of Gene Expression Data: A Narrative Review", Iran. J. Public Health, vol. 46, no. 2, pp. 165-172, 2017.

54. B.M. Kennedy, M. Rehman, W.D. Johnson, M.B. Magee, R. Leonard, and P.T. Katzmarzyk, "Healthcare providers versus patients' understanding of health beliefs and values", Patient Exp. J., vol. 4, no. 3, pp. 29-37, 2017.

55. A.A. Tabl, A. Alkhateeb, L. Rueda, W. ElMaraghy, and A. Ngom, "Identification of the Treatment Survivability Gene Biomarkers of Breast Cancer Patients via a Tree-Based Approach", International Conference on Bioinformatics and Biomedical Engineering, , 2018pp. 166-176

56. D. Paul, R. Su, M. Romain, V. Sébastien, V. Pierre, and G. Isabelle, "Feature selection for outcome prediction in oesophageal cancer using genetic algorithm and random forest classifier", Comput. Med. Imaging Graph., vol. 60, pp. 42-49, 2017.

57. M.T. Lash, Y. Sun, X. Zhou, C.F. Lynch, and W.N. Street, "Learning rich geographical representations: Predicting colorectal cancer survival in the state of Iowa. 2017, arXiv preprint arXiv:1708.04714".

58. U.C.I. Machine Learning Repository, Wisconsin Breast Cancer data set.http://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wiscosin+(Original)

59. Narayan, Vipul, and Daniel AK. "CHOP: Maximum Coverage Optimization and Resolve Hole Healing Problem using Sleep and Wake-up Technique for WSN." ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal 11.2 (2022): 159-178.

60. Narayan, Vipul, and A. K. Daniel. "Design Consideration and Issues in Wireless Sensor Network Deployment."." Invertis Journal of Science & Technology (2020): 101.

61. Pramanik, Sabyasachi, et al. "A Novel Approach Using Steganography and Cryptography in Business Intelligence." Integration Challenges for Analytics, Business Intelligence, and Data Mining. IGI Global, 2021. 192-217.

62. Choudhary, Shubham, et al. "Fuzzy Approach-Based Stable Energy-Efficient AODV Routing Protocol in Mobile Ad hoc Networks." Software Defined Networking for Ad Hoc Networks. Springer, Cham, 2022. 125-139.

63. Narayan, Vipul, and A. K. Daniel. "IOT based sensor monitoring system for smart complex and shopping malls." International Conference on Mobile Networks and Management. Springer, Cham, 2021.

64. Narayan, Vipul, and A. K. Daniel. "CHHP: coverage optimization and hole healing protocol using sleep and wake-up concept for wireless sensor network." International Journal of System Assurance Engineering and Management 13.1 (2022): 546-556.

65. Narayan, Vipul, and A. K. Daniel. "Energy Efficient Protocol for Lifetime Prediction of Wireless Sensor Network using Multivariate Polynomial Regression Model." (2022).

66. Srivastava, Swapnita, and P. K. Singh. "HCIP: Hybrid Short Long History Table-based Cache Instruction Prefetcher." International Journal of Next-Generation Computing 13.3 (2022).

67. Srivastava, Swapnita, and P. K. Singh. "Proof of Optimality based on Greedy Algorithm for Offline Cache Replacement Algorithm." International Journal of Next-Generation Computing 13.3 (2022).