Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 44 No. 3 (2023)

Breast Cancer Recurrence Prediction Using Data Mining

Charanpreet Kaur[1]*
Dr. Rosy Madaan[2]
[1] *Research Scholar, Department of Faculty of Engineering & Technology, MRIIRS, Faridabad, Haryana, India*
[2] *Associate Professor, Department of Faculty of Engineering & Technology, MRIIRS, Faridabad, Haryana, India*

***Abstract*:** Today, the Healthcare sector is generating a vast amount of meticulously detailed information on individuals and their medical issues. Medical data may be investigated for hidden patterns or connections using data mining techniques. The spread and return of the illness are the main causes of breast cancer fatalities. One of the most important areas of study in the field of medicine is the early diagnosis of breast cancer. Healthcare businesses can benefit from data mining by forecasting patient illnesses and behaviour trends. This is performed by analyzing data from many angles and by using data mining to find hidden patterns and associations between seemingly unrelated bits of information. Months after the original diagnosis and breast cancer therapy, a metastatic recurrence of the disease may happen. Only a few number of techniques have been researched in the state of the art due to the challenges associated with early breast cancer recurrence prediction during its medication. The aim of the research is to review the work done by different researchers in this field and to compare the various traditional data mining classifiers (Logistic Regression, Naïve Bayes, K-Nearest Neighbours, Decision Tree, Support Vector Machines, and Random Forest) applied to Breast Cancer Data Set from UCI Machine Learning Repository. Out of these, the best results were given by Random Forest Algorithm with 98.4% accuracy.

Keywords: *Data mining, Breast cancer, Recurrence, Healthcare sector, Prediction*

1. Introduction

According to data from January 2021, there already exist more than 4.66 billion digital natives globally. This represents a total of 59.5 percent of all people on the planet. In just over five years, there has been a rise of almost 83% in online users. Google handles more than 40,000 queries per second, for a daily total of 3.5 billion. This translates to over 2 trillion queries performed a year globally. The data may be analyzed using specialized software programs and strategies. Finding hidden information and recognizing related patterns, correlations, and market trends is an incredibly challenging task. In today's data analysis, statistical modeling, data mining, text analytics, and data optimization are frequently used techniques. The analysis of market patterns, the creation of new business prospects, and the exploration of yet-unknown information are all facilitated by data analytics. With the use of data mining, healthcare firms have improved patient survival rates, early illness diagnosis, customer relationship management, etc. Medical personnel can better understand medical help and treat patients by using data mining.

The most precious asset in everyone's life is their health. Diseases hit the human body suddenly and without prior notice. Genetics or family history may occasionally be the root cause of an illness, over which no one has any influence. Different health issues, such as Heart Issues, Diabetes, Depression, Cancer, Blood Pressure, etc., are widespread nowadays and frequently result in mortality. More than 10 million people will die from cancer worldwide in 2020, making it one of the deadliest diseases in recent years. 1.09 million instances of stomach cancer, 2.26 million occurrences of breast cancer, 2.21 million cases of liver cancer, and 16 1.20 million patients of skin cancer are reported by the World Health Organization (WHO). There have already been too many breast cancer deaths—about 685,000 people globally. Early breast cancer therapy can increase patients' chances of surviving and even save their lives (Latif, M.Z., Shaukat, et al., 2020). If there are any symptoms or pain in the breast, one has to see a doctor for a clinical evaluation. The doctor checks the breast to look for lumps or other symptoms, such as changes in the texture, form, or size of the nipple.

The following are the three main types of breast cancer: -

1. Benign breast cancer: This form of cancer does not pose a threat to life and does not metastasize to other body areas. These cells frequently do not spread outside of their original location. This malignancy is also known as ductal carcinoma in situ since it began within milk ducts (DCIS).
2. Malignant Breast Cancer - If malignant breast cancer is not properly treated, it might be fatal. Other bodily organs may also become inappropriately colonized by these cancer cells. This specific kind of ductal carcinoma affects 80% of individuals and is typically curable by surgery.
3. Other forms of Breast Cancer - are often rare and include milk-producing lobules, papillary breast cancer, tubular breast cancer, medullary breast cancer, and inflammatory breast cancer.

The study's goal is to ensure that women with breast cancer are properly assessed. Breast cancer risk factors for women are growing exponentially. Regardless of the controversy surrounding risk analysis, classification into benign and malignant subtypes remains a crucial diagnostic technique for breast cancer. Breast cancer recurrence can cause a wide range of symptoms, making studies into the disease extremely useful in the treatment of cancer (Walczak, S. and Velanovich, V., 2018).

2. Related Work

Merouane, E. and Said, A., (2022) employed deep learning and the Cross Industry Standard Process for Data Mining (CRISP-DM), which allowed for the provision of appropriate care and decreased the return of malignant cells. The study utilized 449 Moroccan records in total. The performance metrics of Naive Bayes (NB), K-Nearest Neighbours (KNN), Support Vector Machine (SVM), and Logistic Regression were compared. SVM has a 90% accuracy rate, making it the most accurate model.

Ahmad, S., Ullah, T., et al., (2022) introduced the idea of a hybrid AlexNet-gated recurrent unit (AlexNet-GRU) (LN). The suggested AlexNet-GRU model outperformed the convolutional neural network and CNN, long short-term memory model, by 3% on accuracy and other efficiency metrics (CNN-LSTM). Further layers can be added to the recommended model to increase its accuracy even more.

Okagbue, H.I., Adamu, P.I., et al., (2021) utilized a number of independent variables, including as age, sex, length of illness, region of the tumor, and manner of diagnosis, to determine whether patients with breast cancer would survive. The collection had 175 entries for women and 25 records for men. This study employed 10 distinct models, out of all these, AdaBoost delivered the most accurate results.

Kaur, I., Doja, M.N., et al., (2021) performed an investigation on ovarian cancer patient death rates. The therapy administered to the patients was tracked using sequence mining techniques, which also helped to define time intervals for the treatment data. The patients were categorized as having survived or having passed away using the Ensemble approach. The research used a dataset of 140 individuals with ovarian cancer. For the proposed model, the Ensemble technique yielded a precision of 76.4%.

Magboo, V.P.C. and Magboo, M.S.A., (2021) deployed four classifiers, including Logistic Regression, Naive Bayes, K-Nearest Neighbours, and Support Vector Machines, to identify patients who experienced an early cancer relapse following therapy. Based on the characteristics of the patient, a treatment plan for cancer relapse can be suggested early. Using a low-dimensional, four-category high-dimensional data set, breast cancer recurrence was categorized.

Alwohaibi, M., Alzaqebah, M., et al., (2021) used the Brain Storming Optimization technique to investigate two datasets related to breast cancer (BSO). The statistical feature extraction approaches (SFM), which choose features based on relevance levels and associations, were applied at the first level. Three important measures were taken. Next is the second-level step, when each method is assessed using several classifiers. The third stage involves choosing the ideal feature combination. Applied techniques included logistic regression, SVM, and linear discrimination analysis.

Nicolò, C., Périer, C., et al., (2020) proposed early-stage prediction of metastatic recurrence in patients with breast cancer. The model was put into practice utilizing a dataset of 642 patients with 21 clinicopathologic factors. It also used Random Survival Forest, Logistic Regression, Gradient Boosting, SVM, RF, and Cox methods. The results showed that the mechanistic model could reproduce the data best.

Kabiraj, S., Akter, L., et al., (2020) examined the causes of the relapse of the disease. The key factors for the prediction of breast cancer recurrence were determined to be age, sex, acquired genetic deformations, and skin type. The Naive Bayes classifier was used as the foundation for the Bagging approach, which the scientists used

to predict whether or not the patient will experience a relapse of cancer. During the study procedure, an accuracy of 73.8182% was noted.

Zain, Z.M., Alshenaifi, M., et al., (2020) forecasted the recurrence of breast cancer using KNN, REPTree, and NB models. The principle component analysis feature extraction technique and its absence from these models were used in their construction (PCA). KNN performed better than others in terms of accuracy without implementing PCA. In the future, the performance and effectiveness of the suggested models can be improved by using techniques for feature extraction than PCA.

Mosayebi, A., Mojaradi, B., Bonyadi Naeini, A. and Khodadad Hosseini, S.H., (2020) analyzed the machine learning classifier methods for predicting breast cancer recurrence. The authors employed KPCA-SVM, LVQ neural network, Multilayer Perceptron, Bayesian Neural Network, Random Forest, Artificial Neural Network, C5.0, and Random Forest neural networks. With an accuracy rate of 81%, the C5.0 algorithm was the best at predicting breast cancer relapse in the first five years.

Goyal, K., Aggarwal, P., and Kumar, M., (2020) explored the diagnosis and prognosis of the disease. For individuals dealing with breast cancer problems once more, they offered guidance on proper medical care and effective data processing. To calculate the likelihood of a cancer recurrence, the scientists have used a variety of algorithms, including Generalized Regression Neural Networks (GRNN), SVM, DT, and Feed Forward Back Propagation Neural Networks (FFBPN). With an accuracy percentage of 85.18%, FFBPN outperformed other techniques.

Akinnuwesi, B.A., Macaulay, B.O. and Aribisala, B.S., (2020) claimed that although the precise cause of breast cancer (BC) is still unknown to experts, it is among the most prevalent malignancies in women around the globe over. To achieve this, clinical competency must be employed in combination with medical imaging and computational techniques. To quickly identify BCa in research projects, early analytical models based on support vector machines (SVM) and Principal Component Analysis (PCA) were created. PCA was used to extract structures in the first step of pre-processing; however, the collection of characteristics was further condensed in the second stage, which was carried out following the first. Using multi-predicted data that has been processed, SVM was utilized to predict breast cancer risk and diagnosis.

Roberto Cesar, M.O., German, et al., (2020) built a model to forecast the likelihood that patients would have a recurrence of breast cancer following surgery for the condition, using J48 and random forest, Naive Bayes and Naive Bayes Simple, SMO Poli-kernel, and SMORBF-Kernel classifiers. The K-Means cluster method has been implemented after the best classification algorithm was found using the data mining WEKA tool.

Lou, S.J., Hou, M.F., et al., (2020) investigated several deep learning algorithms to predict the recurrence of breast cancer after a breast cancer operation after 10 years. The authors accessed six illness recurrences and identify the potential risk factors for this fatal condition. Preoperative quality of life, clinical information, demographics, and treatment quality were all factors taken into account. The ANN model demonstrated the greatest accuracy. The unique aspect of their analysis was the use of 171 extra datasets, which represented a sizable general population.

Chaudhuri, A.K., Sinha, D. and Thyagaraj, K.S., (2019) intended to develop a technique that would better predict and detect a patient's risk of developing breast cancer again. The UCI Repository served as the study's data source. 10 unique dependent and independent features are present throughout 286 entries. The methods employed were decision trees and discriminant analysis. The results showed that the chance of breast cancer recurrence is influenced by two crucial parameters, the extent of malignancy and the tumor breaching the lymph node.

3. Methodology

The process of data classification is a two-step procedure that incorporates steps for learning and classification. The classification stage is used to predict the class label for the provided data, while the learning step is used to build the classification model. Pre-set classes of data are utilized to construct the classifier model, which can employ any classification technique. The class label determines the predefined class, to which each tuple or sample belongs to. Training and test data are split from the dataset that is provided. Tuples used to build the model are part of the training set. The model is shown as mathematical formulas, decision trees, or classification rules. The test data is utilized to calculate the accuracy of the classifier. The data classification process is shown in figure1.

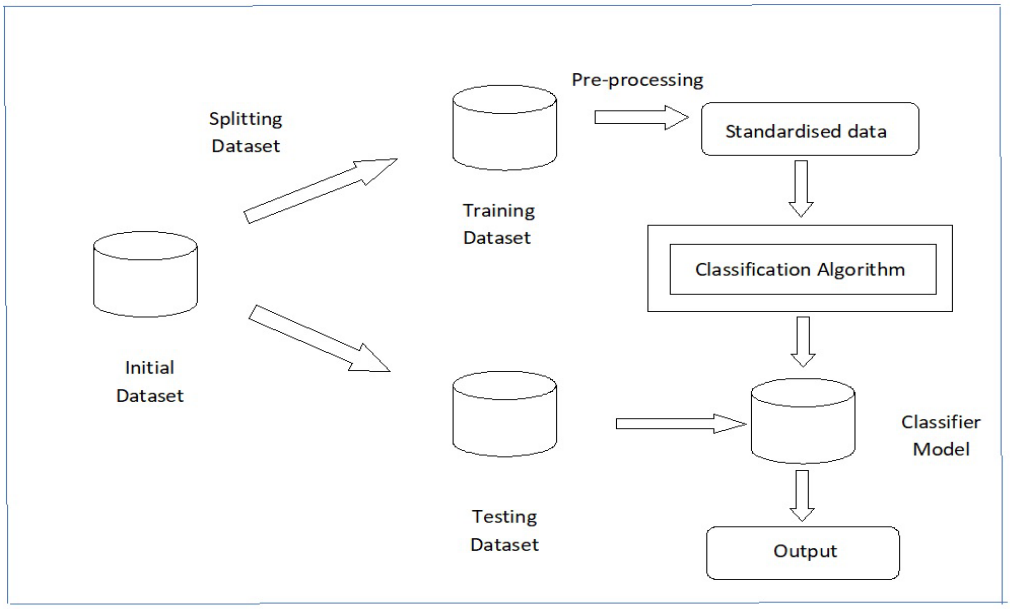

Figure1. Data Classification Process (*Source: Han, J., Pei, J. and Tong, H., 2022. Data mining: concepts and techniques. Morgan kaufmann.)*

3.1 Dataset

The dataset used for the breast cancer relapse prediction supporting Table 1 is extracted from UCI Machine Learning Repository. The data used is taken from the Institute of Oncology, University Medical Center, Ljubljana, Yugoslavia. The dataset is available publicly at https://archive.ics.uci.edu/dataset/14/breast+cancer. [Zwitter, Matjaz and Soklic, Milan]. The dataset consists of 9 attributes and one outcome class that specifies whether breast cancer will relapse or not. The description of the attributes is given in Table 1.

Table 1. Dataset of Breast Cancer Relapse

S.No.	Attribute	Value	Description
1	***Range of Age***	10 to 19 years, 20 to 29 years, 30 to 39 years, 40 to 49 years, 50 to 59 years, 60 to 69 years, 70 to 79 years, 80 to 89 years, 90 to 99 years	It specifies the age at which the initial tumor was found.
2	***Menopause***	lt40, ge40, premeno	At this age, women stop getting their periods. Here, consideration is given to the patient's menopausal condition when the disease was detected.
3	***Size of tumor***	0-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59	It specifies the dimensions of the resulting lump and is measured in millimeters (mm)
4	***inv-nodes***	(0-2), (3-5), (6-8), (9-11), (12-14), (15-17), (18-20), (21-23), (24-26), (27-29),(30-32), (33-35), (36-39)	The number of axillary nodes with symptoms is determined by a histopathological investigation.
5	***node-caps***	yes, no	It indicates whether or not the tumor has spread inside the node capsule.
6	***deg-malig***	1, 2, 3	Tumor's histological grade i.e. how closely malignant cells resemble normal cells.
7	***Breast***	left, right	It specifies which breast has cancerous cells

8	***breast-quad***	left-up, left-low, right-up, right-low, central	When the nipple acts as centre, it describes the quadrant of the breast.
9	***Irradiat***	yes, no	It is the treatment using radiation to destroy cancerous cells.
10	***Class***	no-recurrence-events, recurrence-events	It specifies outcome class i.e. whether cancerous cells will relapse or not

3.2 Pre-processing of Data

The dataset that was downloaded from the UCI Repository was not in the required format and could not be used directly for processing. The dataset contained some missing values. Hence, firstly, the dataset was treated for the missing values and replaced by the mean value for that tuple. The data normalization was performed on the given dataset to convert it into the standardized form. Further, the data was in the nominal form which was converted into its numeric form for further processing. The implementation was done in Jupyter notebook, using Python.

3.3 Data Mining Classification Algorithms

i) Logistic Regression

Logistic regression is the most frequently used Supervised Machine Learning technique. An established collection of independent factors is used to forecast the categorical dependent measure. A classified dependent variable's outcome can be estimated using logistic regression. Consequently, the result must be a distinct or categorical number. Instead of the precise values between 0 and 1, it provides the probability values that fall between those numbers. Either True or False, 0 or 1, or Yes or No, are possible outcomes.

ii) Support Vector Machine

Support Vector Machine **(**SVM) is the approach that establishes a hyperplane-based decision boundary between the various classes. Margin is the distance between classes, and the greater the margin, the greater the accuracy. The training tuples can be mapped non-linearly using a variety of kernel functions, such as polynomial, Gaussian, sigmoid, etc. The SVM algorithm seeks to find the most effective line or judgment limit that can divide n-dimensional space into categories so as to quickly classify new data points in the future. A hyperplane is a term given to this optimal decision boundary as shown in figure 2.

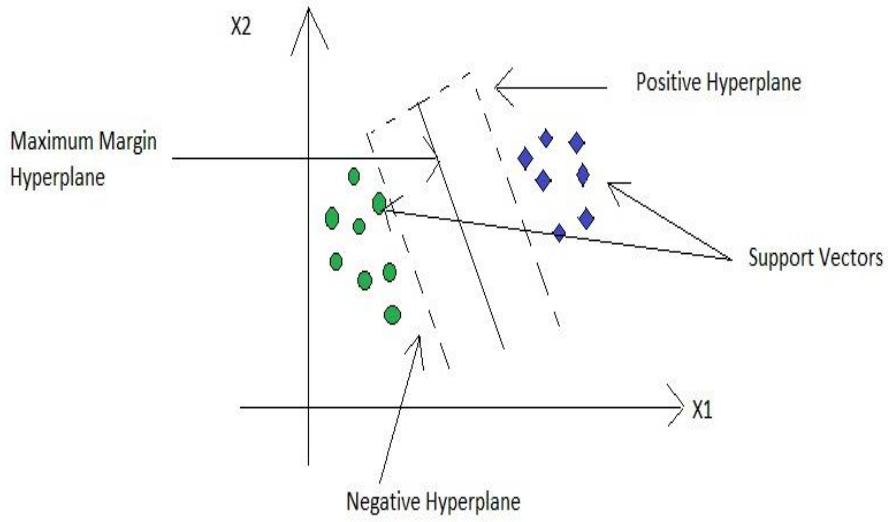

Figure 2. SVM (*Source: Han, J., Pei, J. and Tong, H., 2022. Data mining: concepts and techniques. Morgan kaufmann.)*

iii) Decision Tree

For classification and regression applications, the decision tree is a non-parametric supervised learning method. A root node, branches, internal nodes, and leaf nodes make up its hierarchically arranged structure. Both

classification and regression are accomplished using decision trees based on the supervised learning methodology. It takes a greedy approach, employing the divide-and-conquer method to build the model from the top-down approach. Each leaf node displays the output or the class label, while the interior nodes and branches reflect the dataset utilized and the classification algorithms. (Guo, J., Fung, B.C., et. al.,2017)

iv) K Nearest Neighbour (K-NN)

The supervised learning algorithm K Nearest Neighbour (often abbreviated as KNN) generates forecasts or categories about the clustering of a single value using vicinity. It can be applied to classification or regression tasks, but since it is based on the idea that adjacent similar points can be identified, it is commonly used as a classification method. The nearest neighbour is defined in terms of Euclidean distance, E(x,y) i.e.

$$E(x,y) = \sqrt{\sum_{i=0}^{n}(x_i - y_i)^2}$$

The k parameter in the k-NN algorithm determines the number of neighbours that will be looked at in order to categorise a certain query point. In this case, the value of k was taken in the range from 1 to 25, and at k=1 maximum accuracy was obtained as shown in figure 3.

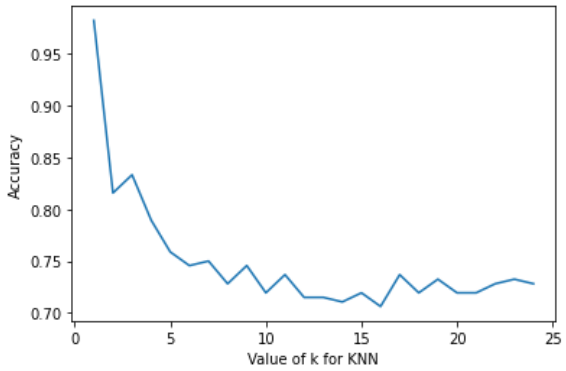

Figure 3. Accuracy at different values for k in KNN *(Source: Self)*

v) Naïve Bayes

The Naive Bayes method resolves the Bayes theorem-based categorization problems. The Naive Bayes Classifier is among the most basic and efficient categorization methods currently in use. The presumption that all characteristics of a data point under evaluation are autonomous of one another is what distinguishes a naive Bayes classifier from other classifiers. It makes it easier to develop effective algorithms with machine learning that can produce reliable predictions. Naïve Bayes will determine the likelihood of the occurrence of the object as –

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

where,

P(A|B) is Posterior probability: Likelihood of hypothesis A given that event B holds.

P(B|A) is Likelihood probability: Likelihood of observing B, given that the possibility of a hypothesis holds.

P(A): Initial probability of A

P(B): Probability that sample B is observed

vi) Random Forest

The random forest uses predictions from each tree as a starting point rather than depending just on one to determine the final output forecast. The random forest classification model utilizes different kinds of decision trees on various sections of the input data to improve the dataset's predictive performance. The greater number of trees in the forest prevents overfitting and higher accuracy.

4. Result

Following traditional data mining algorithms were implemented in the Jupyter notebook interface using Python. The dataset was partitioned into training data (80%) and testing data (20%). All the performance parameters were calculated and compared as shown in table 2. Figure 4 shows the comparison of accuracy between different algorithms and figure 5 represents precision and recall values graphically for the mining techniques used.

Table 2. Performance parameters for different classifiers

S.NO	DATA MINING CLASSIFIER	ACCURACY	PRECISION	RECALL
1	LOGISTIC REGRESSION	0.728	0.7513	0.9044
2	SUPPORT VECTOR MACHINE	0.7192	0.7569	0.8726
3	DECISION TREE	0.75	0.75	0.9554
4	K-NEAREST NEIGHBOUR	0.7587	0.7317	0.9554
5	NAÏVE BAYES	0.728	0.7513	0.9044
6	RANDOM FOREST	0.9824	0.98113	0.9936

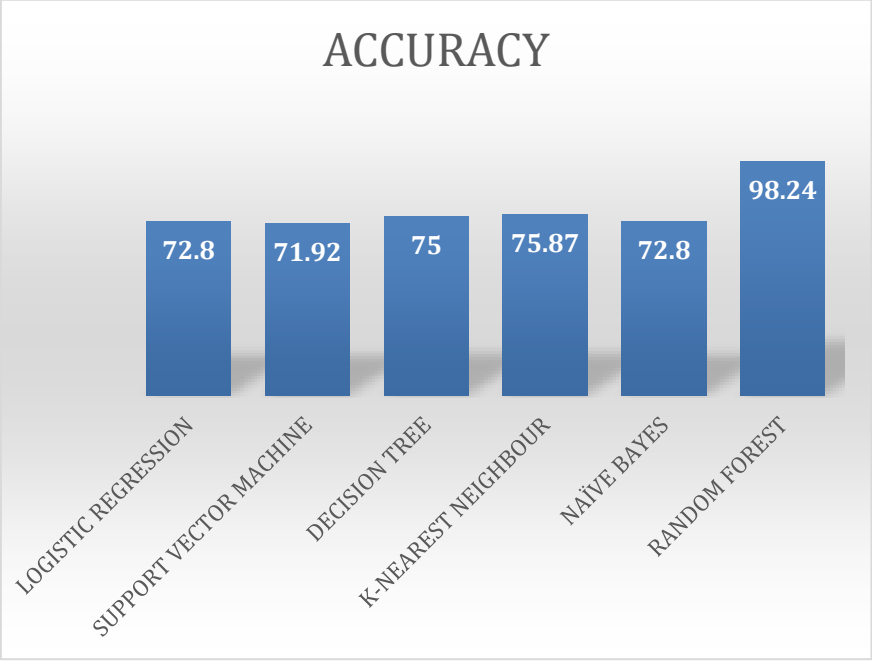

Figure 4. Accuracy for different classifiers *(Source:Self)*

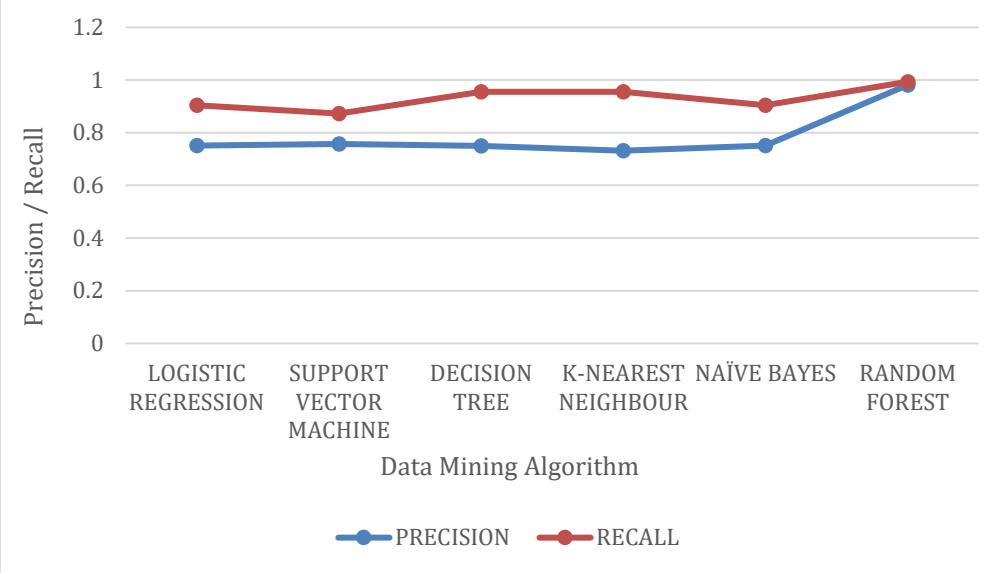

Figure 5. Precision Vs Recall value for different classifiers *(Source: Self)*

5. Conclusion And Future Scope

Taking preventative measures before breast cancer relapses is crucial because it is one of the deadliest diseases. Using conventional data mining approaches, the recurrence of breast cancer is predicted in this work. Out of the implemented algorithms, Random Forest gave the best accuracy of 98.24%. In future work, different combinations of data mining techniques and hybrid models can be implemented to increase performance efficiency. Further, by increasing the attributes of the dataset and its size, better results and flexibility can be achieved. Also, image datasets can be used along with textual data so that the model can be implemented in the real-time medical field.

Declarations

Ethical Approval- Not Applicable
Competing interests- The authors declare that they have no conflict of interest.
Authors' contributions- The first author contributed most of the work including the writing of the manuscript and implementation of the manuscript and the second author critically approved the version of the manuscript.
Funding- Not Applicable

References

[1] Abed, B.M., Shaker, K., Jalab, H.A., Shaker, H., Mansoor, A.M., Alwan, A.F. and Al-Gburi, I.S., 2016, November. A hybrid classification algorithm approach for breast cancer diagnosis. In 2016 IEEE Industrial Electronics and Applications Conference (IEACon) (pp. 269-274). Ieee.

[2] Abreu, P.H., Santos, M.S., Abreu, M.H., Andrade, B. and Silva, D.C., 2016. Predicting breast cancer recurrence using machine learning techniques: a systematic review. ACM Computing Surveys (CSUR), 49(3), pp.1-40.

[3] Ahmad, S., Ullah, T., Ahmad, I., Al-Sharabi, A., Ullah, K., Khan, R.A., Rasheed, S., Ullah, I., Uddin, M. and Ali, M., 2022. A Novel Hybrid Deep Learning Model for Metastatic Cancer Detection. Computational Intelligence and Neuroscience, 2022.

[4] Akinnuwesi, B.A., Macaulay, B.O. and Aribisala, B.S., 2020. Breast cancer risk assessment and early diagnosis using Principal Component Analysis and support vector machine techniques. Informatics in medicine unlocked, 21, p.100459.

[5] Almuhaidib, D.A., Albusayyis, F.M., Shaiba, H.A., Alzaid, M.A., Alharbi, N.G., Almadhi, R.M. and Alotaibi, S.M., 2018, April. Ensemble learning method for the prediction of breast cancer recurrence. In 2018 1st International Conference on Computer Applications & Information Security (ICCAIS) (pp. 1-6). IEEE.

[6] Alwohaibi, M., Alzaqebah, M., Alotaibi, N.M., Alzahrani, A.M. and Zouch, M., 2022. A hybrid multi-stage learning technique based on brain storming optimization algorithm for breast cancer recurrence prediction. Journal of King Saud University-Computer and Information Sciences, 34(8), pp.5192-5203.
[7] Chaudhuri, A.K., Sinha, D. and Thyagaraj, K.S., 2019. Identification of the recurrence of breast cancer by discriminant analysis. In Emerging technologies in data mining and information security (pp. 519-532). Springer, Singapore.
[8] Chaudhuri, A.K., Sinha, D., Bhattacharya, K. and Das, A., 2020. An Integrated Strategy for Data Mining Based on Identifying Important and Contradicting Variables for Breast Cancer Recurrence Research. Int. J. Recent Tech. Eng, 8.
[9] Cirkovic, B.R.A., Cvetkovic, A.M., Ninkovic, S.M. and Filipovic, N.D., 2015, November. Prediction models for estimation of survival rate and relapse for breast cancer patients. In 2015 IEEE 15th International Conference on Bioinformatics and Bioengineering (BIBE) (pp. 1-6). IEEE.
[10] Doja, M.N., Kaur, I. and Ahmad, T., 2020. Age-specific survival in prostate cancer using machine learning. Data Technologies and Applications.
[11] Ganggayah, M.D., Taib, N.A., Har, Y.C., Lio, P. and Dhillon, S.K., 2019. Predicting factors for survival of breast cancer patients using machine learning techniques. BMC medical informatics and decision making, 19(1), pp.1-17.
[12] Gayathri, B.M. and Sumathi, C.P., 2015, December. Mamdani fuzzy inference system for breast cancer risk detection. In 2015 IEEE international conference on computational intelligence and computing research (ICCIC) (pp. 1-6). IEEE.
[13] Goyal, K., Aggarwal, P. and Kumar, M., 2020. Prediction of breast cancer recurrence: a machine learning approach. In Computational Intelligence in Data Mining (pp. 101-113). Springer, Singapore.
[14] Guo, J., Fung, B.C., Iqbal, F., Kuppen, P.J., Tollenaar, R.A., Mesker, W.E. and Lebrun, J.J., 2017. Revealing determinant factors for early breast cancer recurrence by decision tree. Information Systems Frontiers, 19, pp.1233-1241.
[15] Gupta, S., Kumar, D. and Sharma, A., 2011. Data mining classification techniques applied for breast cancer diagnosis and prognosis. Indian Journal of Computer Science and Engineering (IJCSE), 2(2), pp.188-195.
[16] Kabiraj, S., Akter, L., Raihan, M., Diba, N.J., Podder, E. and Hassan, M.M., 2020, July. Prediction of recurrence and non-recurrence events of breast cancer using bagging algorithm. In 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT) (pp. 1-5). IEEE.
[17] Kate, R.J. and Nadig, R., 2017. Stage-specific predictive models for breast cancer survivability. International journal of medical informatics, 97, pp.304-311.
[18] Kaur, I., Doja, M.N., Ahmad, T., Ahmad, M., Hussain, A., Nadeem, A., El-Latif, A. and Ahmed, A., 2021. An integrated approach for cancer survival prediction using data mining techniques. Computational Intelligence and Neuroscience, 2021.
[19] Lafourcade, A., His, M., Baglietto, L., Boutron-Ruault, M.C., Dossus, L. and Rondeau, V., 2018. Factors associated with breast cancer recurrences or mortality and dynamic prediction of death using history of cancer recurrences: the French E3N cohort. BMC cancer, 18(1), pp.1-9.
[20] LANBARAN, N.M. and ÇELİK, E., 2021. Prediction of breast cancer through tolerance-based intuitionistic fuzzy-rough set feature selection and artificial neural network. Gazi University Journal Of Science, pp.1-1.
[21] Lashari, S.A., Ibrahim, R., Senan, N. and Taujuddin, N.S.A.M., 2018. Application of data mining techniques for medical data classification: a review. In MATEC Web of conferences (Vol. 150, p. 06003). EDP Sciences.
[22] Latif, M.Z., Shaukat, K., Luo, S., Hameed, I.A., Iqbal, F. and Alam, T.M., 2020, June. Risk factors identification of malignant mesothelioma: A data mining based approach. In 2020 International Conference on Electrical, Communication, and Computer Engineering (ICECCE) (pp. 1-6). IEEE.
[23] Lou, S.J., Hou, M.F., Chang, H.T., Chiu, C.C., Lee, H.H., Yeh, S.C.J. and Shi, H.Y., 2020. Machine learning algorithms to predict recurrence within 10 years after breast cancer surgery: A prospective cohort study. Cancers, 12(12), p.3817.
[24] Magboo, V.P.C. and Magboo, M.S.A., 2021. Machine Learning Classifiers on Breast Cancer Recurrences. Procedia Computer Science, 192, pp.2742-2752.

[25] Mining, D., 2019. Application of data mining techniques to predict breast cancer. Procedia Comput. Sci, 163, pp.11-18.
[26] Mojrian, S., Pinter, G., Joloudari, J.H., Felde, I., Szabo-Gali, A., Nadai, L. and Mosavi, A., 2020, October. Hybrid machine learning model of extreme learning machine radial basis function for breast cancer detection and diagnosis; a multilayer fuzzy expert system. In 2020 RIVF International Conference on Computing and Communication Technologies (RIVF) (pp. 1-7). IEEE.
[27] Mosayebi, A., Mojaradi, B., Bonyadi Naeini, A. and Khodadad Hosseini, S.H., 2020. Modeling and comparing data mining algorithms for prediction of recurrence of breast cancer. PloS one, 15(10), p.e0237658.
[28] Nicolò, C., Périer, C., Prague, M., Bellera, C., Macgrogan, G., Saut, O. and Benzekry, S., 2020. Machine learning and mechanistic modeling for prediction of metastatic relapse in early-stage breast cancer. JCO clinical cancer informatics, 4, pp.259-274.
[29] Ogundele, I.O., Popoola, O.L., Oyesola, O.O. and Orija, K.T., 2018. A review on data mining in healthcare. International Journal of Advanced Research in Computer Engineering and Technology (IJARCET).
[30] Okagbue, H.I., Adamu, P.I., Oguntunde, P.E., Obasi, E. and Odetunmibi, O.A., 2021. Machine learning prediction of breast cancer survival using age, sex, length of stay, mode of diagnosis and location of cancer. Health and Technology, 11(4), pp.887-893.
[31] Roberto Cesar, M.O., German, L.B., Paola Patricia, A.C., Eugenia, A.R., Elisa Clementina, O.M., Jose, C.O., Marlon Alberto, P.M., Fabio Enrique, M.P. and Margarita, R.V., 2020, July. Method based on data mining techniques for breast cancer recurrence analysis. In International Conference on Swarm Intelligence (pp. 584-596). Springer, Cham.
[32] Shukla, N., Hagenbuchner, M., Win, K.T. and Yang, J., 2018. Breast cancer data analysis for survivability studies and prediction. Computer methods and programs in biomedicine, 155, pp.199-208.
[33] Simsek, S., Kursuncu, U., Kibis, E., AnisAbdellatif, M. and Dag, A., 2020. A hybrid data mining approach for identifying the temporal effects of variables associated with breast cancer survival. Expert Systems with Applications, 139, p.112863.
[34] Singh, P.D., Kaur, R., Singh, K.D. and Dhiman, G., 2021. A novel ensemble-based classifier for detecting the COVID-19 disease for infected patients. Information Systems Frontiers, 23, pp.1385-1401.
[35] Tapak, L., Shirmohammadi-Khorram, N., Amini, P., Alafchi, B., Hamidi, O. and Poorolajal, J., 2019. Prediction of survival and metastasis in breast cancer patients using machine learning classifiers. Clinical Epidemiology and Global Health, 7(3), pp.293-299.
[36] Thani, I. and Kasbe, T., 2020, February. Breast Cancer: State–of-the-art causes and diagnosis. In 2nd International Conference on Data, Engineering and Applications (IDEA) (pp. 1-6). IEEE.
[37] Tseng, Y.J., Huang, C.E., Wen, C.N., Lai, P.Y., Wu, M.H., Sun, Y.C., Wang, H.Y. and Lu, J.J., 2019. Predicting breast cancer metastasis by using serum biomarkers and clinicopathological data with machine learning technologies. International journal of medical informatics, 128, pp.79-86.
[38] Walczak, S. and Velanovich, V., 2018. Improving prognosis and reducing decision regret for pancreatic cancer treatment using artificial neural networks. Decision Support Systems, 106, pp.110-118.
[39] Wang, H., Zheng, B., Yoon, S.W. and Ko, H.S., 2018. A support vector machine-based ensemble algorithm for breast cancer diagnosis. European Journal of Operational Research, 267(2), pp.687-699.
[40] Yang, P.T., Wu, W.S., Wu, C.C., Shih, Y.N., Hsieh, C.H. and Hsu, J.L., 2021. Breast cancer recurrence prediction with ensemble methods and cost-sensitive learning. Open Medicine, 16(1), pp.754-768.
[41] Zeng, Z., Yao, L., Roy, A., Li, X., Espino, S., Clare, S.E., Khan, S.A. and Luo, Y., 2019. Identifying breast cancer distant recurrences from electronic health records using machine learning. Journal of healthcare informatics research, 3(3), pp.283-299.
[42] Zhou, H., Dong, D., Chen, B., Fang, M., Cheng, Y., Gan, Y., Zhang, R., Zhang, L., Zang, Y., Liu, Z. and Zheng, H., 2018. Diagnosis of distant metastasis of lung cancer: based on clinical and radiomic features. Translational oncology, 11(1), pp.31-36.
[43] Zwitter,Matjaz and Soklic,Milan. (1988). Breast Cancer. UCI Machine Learning Repository. https://doi.org/10.24432/C51P4M.