INDIAN JOURNAL OF SCIENCE AND TECHNOLOGY

RESEARCH ARTICLE

An Expert System for Breast Cancer Prediction (ESBCP) using Decision Tree

Akhil Kumar Das[1]*, **Saroj Kumar Biswas**[2], **Ardhendu Mandal**[3]

1 Department of Computer Science, Gour Mahavidyalaya, Malda, West Bengal, 732142, Mangalbari, India
2 Department of Computer Science and Engineering, Assistant professor Stage I, NIT Silchar, Assam 788010, India
3 Department of Computer Science and Application, University of North Bengal, West Bengal, Darjeeling734013, India

OPEN ACCESS

Received: 05-04-2022

Accepted: 17-10-2022

Published: 02-12-2022

Citation: Das AK, Biswas SK, Mandal A (2022) An Expert System for Breast Cancer Prediction (ESBCP) using Decision Tree. Indian Journal of Science and Technology 15(45): 2441-2450. https://doi.org/10.17485/IJST/v15i45.756

* **Corresponding author**.

dasakhi@gmail.com

Funding: None

Competing Interests: None

Published By Indian Society for Education and Environment (iSee)

ISSN
Print: 0974-6846
Electronic: 0974-5645

Abstract

Objectives: Breast cancer is one of the major concerns in present day scenario. Detecting breast cancer at early stage increases the chances of survival. The objective of this research is to propose suitable feature selection method to improve the efficiency of breast cancer prediction at early stages to increase the survival rate. **Methods:** In this work, an expert intelligent technique has been proposed named "Expert System for Breast Cancer Prediction (ESBCP)" to detect breast cancer. To validate the results, the proposed system determines accuracy, precision, F-measure, and recall. The proposed model introduced a feature selection technique named - Undiluted Feature Set (UFS) to select the most relevant and promising features. The experimental work was carried out using Python 2.8 version in a Windows environment, taking a dataset on breast cancer from the UCI machine learning repository. There were 699 occurrences in the dataset with nine attributes and two classes. The proposed work utilized a decision tree and a new feature selection technique based on a heuristic search and the Stochastic Hill method. The experimental results were evaluated using the 10-fold Cross-Validation (CV). **Findings:** The experimental findings showed that the suggested model - ESBCP can accurately detect breast cancer at an early stage. As per the result, with simple decision tree the accuracy recorded 93.42 percent whereas ESBCP obtained 94.01 percent. It may seem that the improvement of 0.59 percent is very small, but for a large population even this mere change can have a greater impact. **Novelty:** The suggested model ESBCP and the feature selection technique - UFS have a lot of potential in the fields of medical research and bioinformatics in terms of classification capability and predictive power.

Keywords: Expert System; Decision Tree; Undiluted Feature Set; Breast Cancer; Feature Selection

1 Introduction

The advancement of medical research has a significant dependency on technological innovation, and as a result, millions of people have been saved from potentially different life-threatening diseases. Breast Cancer (BC) is one of those potentially fatal illnesses, particularly for women[1]. It begins in breast cells, which frequently develop abnormally. Cancer usually begins in the milk-producing glands - lobules or in the milk-transporting ducts. Eventually, it expands to several regions of the body - including the brain, bones, and lungs, making it a life-threatening fatal disease. The impact of breast cancer is very much severe. According to the report of 2019 from American Cancer Society (ACS), the United States of America has 3.1 million breast cancer survivors[2]. Survival of breast cancer for at least 5 years after diagnosis ranges from more than 90% in high-income countries, to 66% in India and 40% in South Africa. As per the report of WHO-2020, there were 2.3 million women diagnosed with breast cancer and 685,000 deaths globally[3]. As of the end of 2020, there were 7.8 million women alive who were diagnosed with breast cancer in the past 5 years, making it the world's most prevalent cancer. There are more lost Disability-Adjusted Life Years (DALYs) by women to breast cancer globally than any other type of cancer.

Finding any preventative techniques is essential given the seriousness of the life-threatening challenges that patients experience. Early diagnosis of Breast Cancer is essential to aid in preventative measures since it allows for proper treatment to be administered to minimize complications and reduce death rates. As a result, it is always beneficial to have an intelligent expert system to detect breast cancer based on clinical symptoms at an early stage so that the proper diagnosis and treatment can be carried out on time. Different machine learning algorithms have utilized for making a prediction model for the breast cancer dataset[4]. It's tough to come up with a comprehensive model for breast cancer prediction because it is highly non- linear. Several expert systems for identifying breast cancer are available with their merits and demerits.

In this work, our goal was to propose a model which determines breast cancer based on symptomatic features. It should be able to reduce the risk of breast cancer through early treatment by simply considering clinical symptoms, which effectively minimizes expenditure cost, and time. Consequently, this research proposes an intelligent expert system called the "Expert System for Breast Cancer Prediction (ESBCP)" to determine breast cancer based on symptomatic features. Furthermore, a tree-based classifier - Decision Tree (DT) is used for classification, which is a white box method that creates a client sensible option with valid clarity. The performance of ESBCP is compared to precision, recall, accuracy, and the F-measure with simple DT. The following is how the paper is arranged. In Section 2, introduces the related works in breast cancer detection. Sect. 3 provides the flowchart of the methodology. Sect. 4 provides the dataset description and different data pre-processing techniques employed in this research work. Result and proper discussion based on models have been explored in Section 5. Section 6 then follows with the conclusion.

1.1 Literature review

Several research works have been carried out to detect breast cancer using machine learning classifiers. Identification of the most important risk factors for breast cancer and further to detect and prevent it is also a major area of research. Many researchers have addressed these issues using the machine learning techniques. Rajaguru et al.[4] have utilized Decision Tree (DT) and k-Nearest Neighbour (k-NN) machine learning classifiers for the prediction of BC. Principle Component Analysis (PCA) for feature selection was employed by them. Although k-NN outperformed DT in terms of accuracy but the accuracy level achieved by DT was also encouraging with 91.23%. Similarly, Ghani et al.[5] also concentrated on the diagnosis of BC using four machine learning classifiers like k-NN, DT, NN and NB. They have also used recursive feature elimination method for feature selection. They found ANN is the most suitable prediction technique. Tian et al.[6] have proposed a method named P-Boosted C5.0 algorithms that combines Principal Component Analysis (PCA), a boosted C5.0 decision tree (DT) algorithm, and penalty factor. PCA was utilised to reduce the dimension of the feature subset. For classification, the boosted C5.0 decision tree (DT) approach was utilised as an ensemble classifier. A penalty factor was applied here to improve the categorization result. The proposed method was mostly used to address the problem of class imbalance in breast cancer classification while maintaining the optimal new feature subset. Khan et al.[7] have recommended a diagnosis system for the prediction of breast cancer (BC) utilising soft computing algorithms. The proposed cloud-based intelligent BCP-T1FSVM consists of two models, such as BCP-T1F and BCP-SVM. It was mainly used for the type of cancer and stage of cancer. The BCP-SVM gives the higher precision of the proposed breast cancer detection model with 97.06 % accuracy. The BCP-T1F system is used to diagnose breast cancer at an early stage with 96.56% accuracy.

A new ensemble learning technique for classifying breast cancer was recommended by Tabrizchi et al.[8]. It consisted with Multi-Verse Optimizer (MVO) and Gradient Boosting Decision Tree (GBDT) utilizing data taken for breast cancer from the WDBC and BCW. The purpose of this work was to improve classification accuracy while avoiding overfitting problems. Assegie et al.[9] have proposed a model that combines the Decision Tree (DT) and the Adaptive boosting (Adboost) method. Adaptive boosting was used to handle the biased classification problem on unbalanced dataset collected from Kaggle. According to the results, the adaptive boosting method outperformed the decision tree. Mangukiya et al.[10] have described the different machine learning methods for analysing the disease of BC. They examined seven machine learning approaches. These were Adaboost, XGboost, Decision Tree, Naive Bayes (NB), K Nearest Neighbours (k-NN), Random Forest, and Support Vector Machine (SVM). As per the experimental result, the XGboost got the highest accuracy with 98.24 percent, as well as the lowest error rate. They also employed precision, sensitivity, and specificity as performance metric for detecting breast cancer. Similar approaches have been made by Khan et al.[11] to examine the performances of different machine learning methods for the prediction of BC. These were random forest, decision tree, K-nearest neighbor, and logistic regression. They compared the accuracy of various classifiers and observed the highest accuracy of logistic regression (98.6 percent). In addition to this, they have also reported the total runtime of each technique which was approximately 2-3 minutes.

Gayakwad et al[12] have developed a new technique that gets the highest accuracy with the minimum error. The proposed model consisted of dataset selection, data processing, and research methodology. They have utilised eight machine learning methods like SVM, Decision Tree, Random Forest, Adaboost classifier, Naive Bayes, KNN, XGboost, and linear regression. In this paper, they showed that the KNN, SVM, and XGboost techniques produced better outcomes. As per the experimental study, they achieved 94 percent highest accuracy while the lowest accuracy was at 75 percent which was for decision tree. Assegie et al.[13] have suggested a machine learning model applying the support vector machine and decision tree. The support vector machine outperformed the decision tree approach in terms of precision, accuracy, and the number of misclassifications. The support vector machine had an average accuracy of 91.92 percent, while the decision tree classification model had an average accuracy of 87.12 percent. In continuation with this, Botcha et al.[14]also utilized random forest for breast cancer prediction. They used the Wisconsin Breast Cancer dataset and the error rate was at 0.0177. Similar approaches have been made by Laghmati et al.[15] to diagnosis breast cancer. They have used some data mining approaches like Binary Support Vector Machine (Binary SVM), Artificial Neuron Network (ANN), K-Nearest Neighbors (KNN), and Decision Tree (DT). According to their study, ANN was more effective for BC severity than SVM, KNN, and DT utilizing one dataset. On the other hand, SVM was better than ANN and KNN for the other dataset. The early identification of cancer utilising image mining and soft computing techniques was presented in their research. By employing the most efficient soft computing-based picture mining approaches, the linked challenges can appropriately be rectified.

Naji et al.[16] have concentrated on the diagnosis and prediction of breast cancer (BC) using different machine learning methods like support vector machine, Random Forests, Logistic Regression, Decision Trees, and K-NN. They compared the accuracy of various strategies and noticed SVM was the most suitable prediction technique with highest accuracy recorded at 97.2%.

Sivapriya et al.[17] have utilized NB, SVM, Random Forest, and LR for predicting BC. They examined the accuracy of various techniques and came to conclusion that SVM was the most appropriate prediction technique with the highest accuracy recorded at 99.76%.

Ak et al.[18] used KNN, NB, DT, RF, SVM and LR to classify the breast cancer dataset and got the best result by utilizing LR with 98.1% accuracy. The purpose of the study was to conduct a comparison of data visualisation the BC detection and diagnosis. Hazra et al.[19] have investigated the best method for predicting breast cancer using two classifiers such as ANN and DT. In this study, ANN came out to be the best classifier with 98.55% accuracy. Abdulrahman et al.[20] made a comparison among ANN, SVM, DT, RF, GNB and KNN to obtain the best results for BC prediction. As per the study, SVM and RF found out to be the best performer. It was evident from the literature that breast cancer has been considered as a serious issue and diverse strategies have been proposed for its diagnosis and prevention.

Table 1. Accuracies of the previous works and their limitations

Reference	Year	Methodology	Limitation	Accuracy of Decision Tree (%)
[4]	2019	**DT** and KNN	Proper Data Pre-processing approach missing	91.23
[5]	2019	**DT**, KNN, NB and NN	Dataset was very small to make proper decision	71.43
[9]	2021	Adaptive boosting and DT	Feature selection technique has not been utilized.	88.8

Continued on next page

Table 1 continued

(12)	2021	KNN, SVM, XGboost , NB, AD, SVM, DT etc.	Feature selection techniques were not used properly.	75
(13)	2020	SVM and DT	Data Pre-processing as well as feature selection techniques were not used.	87.12
(18)	2020	Logistic Regression, k-NN, SVM, NB, DT, RF	Data Pre-processing methods were not used.	93.85 (avg)
(19)	2020	ANN and DT	Data Pre-processing as well as feature selection techniques were not used properly.	94
(20)	2022	Vector Support equipment, Nave Bayes Classifier, Artificial Neural Networks, K-NN, Random Forests, and DT	Feature selection technique has not been utilized.	93

Research Gap: Even though, the researchers have applied the decision tree method to predict breast cancer, still scope of improvement exists to make it more accurate. The proposed work introduced a novel feature selection strategy – UFS to increase breast cancer prediction accuracy.

The following are the primary achievements of this research task:

- This research paper focused on two key points:
- Utilizing a tree based classifier - a decision tree;
- A novel and efficient feature selection method - UFS.

2 Methodology

Consequently, this research proposes the use of an intelligent expert system called the Expert System for Breast Cancer Prediction (ESBCP) to determine BC based on symptomatic features. There are many expert systems for detecting breast cancer on the market. But the proposed ESBCP is more efficient and an early prediction of BC. ESBCP is a fruitful model because it takes only clinical symptoms, so it reduces the expenditure and time; moreover, it reduces the danger of BC by properly treating it at an early stage. Besides, a BC classifier is utilized for classification, which may well be a white box strategy and produces a reasonable choice with legitimate clarification and thinking. The execution of ESBCP is compared to Precision, the F-Measure, Accuracy, and Recall with straightforward BC.

The ESBCP consists of three phases. These are i) Data description & data pre-processing, ii) Feature Selection and iii) Decision Making. The raw dataset of breast cancer is frequently incomplete, inconsistent, deficient in specific behaviours or trends, and prone to numerous errors. In the data pre-processing phase, the raw dataset of breast cancer is converted into an appropriate format that can be understood. The next phase, i.e., feature selection, evaluates the best feature from the breast cancer dataset. The Decision-Making phase produces the rules from the best features of the breast cancer dataset. These rules are generated using the decision tree and make the decision about breast cancer. Figure 1 depicts a schematic illustration of the proposed ESBCP.

2.1 Data Description

The breast cancer dataset contains 699 cases with 11 attributes, including the target attribute, which is classified as ’Benign(2)’ or ’Malignant(4)’. ’id’, ’Clump Thickness’, ’Uniformity of Cell Size’, ’Uniformity of Cell Shape’, ’Marginal Adhesion’, ’Single Epithelial Cell Size’, ’Bare Nuclei’, ’Bland Chromatin’, ’Normal Nucleoli’, ’Mitoses’, and ’Class’ were the attributes found in the dataset. The dataset contained 458 and 241 benign and malignant cases, respectively. 16 rows out of the 699 records contained missing values.

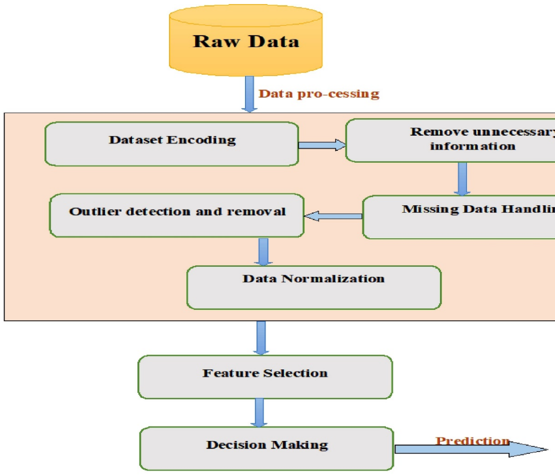

Fig 1. Expert System for Breast Cancer Prediction (ESBSP)

Table 2. Distribution of Mean and Standard Deviation of the raw dataset

Central Tendency Attribute	Clump Thickness	Uniformity of Cell Size	Uniformity of Cell Shape	Marginal Adhesion	Single Epithelial Cell Size	Bare Nuclei	Bland Chromatin	Normal Nucleoli	Mitoses
Mean	4.417740	3.134478	3.207439	2.806867	3.216023	1.417740	3.437768	2.866953	1.589413
Std	2.815741	3.051459	2.971913	2.855379	2.214300	2.499862	2.438364	3.053634	1.715078

2.2 Data pre-processing

The raw dataset of BC is prepossessed in the data pre-processing phase to remove any irrelevant or extraneous features. It includes the following sub-phases: Dataset encoding, removing unnecessary information, Missing Data Handling, Outlier detection and removal, and Normalized Data.

2.2.1 Dataset Encoding

Among the 11 attributes, Bare Nuclei was the only attribute with categorical feature (object data type). Using label encoding the feature was encoded in six (0-5) labels for better processing. The target attribute class was categorised with two values – '2' for benign and '4' for malignant. It was also converted to 0 and 1 for benign and malignant respectively.

2.2.2 Remove unnecessary information

The dataset was collected from the UCI repository. In this step, unnecessary elements present in the dataset - like 'id' were deleted in order to create a uniform data collection.

2.2.3 Missing Data Handling

The BC dataset contained 16 rows with missing values which was presented in the form of '?'. Several techniques are present to handle these missing values like imputation with the mean, mode etc. For simplicity, we have removed them from the dataset. After deleting them, the final dataset consists of 683 records out of which 444 and 239 were benign and malignant. Distribution of data after deletion can be visualized from the following table.

Table 3. Distribution of Mean and standard deviation after removing the missing data

Central Tendency Attribute	Clump Thickness	Uniformity of Cell Size	Uniformity of Cell Shape	Marginal Adhesion	Single Epithelial Cell Size	Bare Nuclei	Bland Chromatin	Normal Nucleoli	Mitoses
Mean	4.442167	3.150805	3.215227	2.830161	3.234261	1.216691	3.445095	2.869693	1.603221
Std	2.820761	3.065145	2.988581	2.864562	2.223085	2.151154	2.449697	3.052666	1.732674

2.2.4 *Outlier detection and removal*

Outliers are the elements which causes difficulties for learning and prediction. Detection and removal of outliers present in the dataset are one of the challenging issues [(21)].In our work we have utilized z-score to find out the outliers present in the dataset. Absolute value of z-score with less than 3 was considered for which 73 records were identified as outliers.

2.2.5 *Data Normalization*

In this step, dataset have been converted the features in such a way that every feature equally contributes. It is mostly used to organise and analyse enormous amounts of data. It is also converting the data from one format to another to enable further processing at this stage. Standardization scales each input variable separately by subtracting the mean and dividing by the standard deviation to shift the distribution to have a mean of zero and a standard deviation of one.

2.3 Feature selection

The relevant and crucial features have been selected utilizing a proposed feature selection method named Undiluted Feature Set (UFS). In this technique, a heuristic search method and stochastic hill climbing were utilised to select the most significant and promising features. By deleting the irrelevant and redundant features from the existing feature set, the significant features were identified. A fitness proportionate selection technique was utilised for eliminating a feature, with a probability of selecting a feature according to its score value. The systematic view of the UFS feature selection approach is depicted in Figure 2.

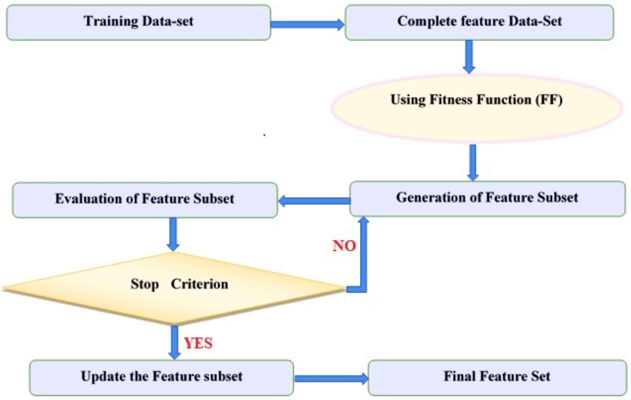
Training Data-set
Complete feature Data-Set
Using Fitness Function (FF)
Generation of Feature Subset
Evaluation of Feature Subset
NO
Stop Criterion
YES
Update the Feature subset
Final Feature Set

Fig 2. Undiluted Feature Set (UFS)

2.4 Decision Making

To create Decision Tree(DT), several methods can be utilized to split a data set based on specific factors. Decision Tree is a non-parametric supervised learning method for classification and regression. Here, the DT was used to predict and classify the class of data. This tree adopts the C 4.5 algorithm. At this stage, a sequence of decision rules is generated using the DT. These rules are used to identify the class of breast cancer.

Classification has been done as per the following way:

$$\text{Splitinfos (Z)} = -\sum \frac{|z_j|}{|z|} * \log_2 \frac{|z_j|}{|z|} \tag{1}$$

$$GainRatio(S) = Gain(S)/SplitinfoS(Z) \tag{2}$$

$$GainS(Z) = Info(Z) – InfoS(Z) \tag{3}$$

$$Info(Z) = -\sum Pj * log2Pj \tag{4}$$

$$\text{Info}_S(Z) = -\sum \frac{|z_j|}{|Z|} * \text{Info}\,(Z_\text{i}) \tag{5}$$

Where, Z → dataset ; (Z1 . . . Zj) →patterns; C → different classes Pj→ Probability of a pattern; $\frac{|z_j|}{|z|}$ → weight of j$^{\text{th}}$ partition;S→ attribute ; SplitinfoS(Z) → potential information in (1); GainS(Z) → expected reduction in entropy in (3); Info(Z) → probabilistic measure of uncertainty in (4); InfoS(Z) → information of uncertainty in (5);

3 Results and Discussion

The proposed expert model ESBCP was implemented using Python programming language. The final dataset was split in 90:10 ratios as training and testing purpose. The suggested approach ESBCP was compared to a simple decision tree to validate the findings produced. The classical way to determine the efficiency of any model is to compare it with the baseline simple decision tree model (Simple DT). In this work, we had utilized Undiluted Feature Set (UFS) to select the most significant and promising features by deleting the irrelevant and redundant features from the existing feature set. A fitness proportionate selection technique was utilised for eliminating a feature, with a probability of selecting a feature according to its score value. Here Figure 3 displays the accuracy comparison of simple decision tree (DT) and the proposed model-ESBCP. As per the obtained result, the accuracy of ESBCP is 0.59% higher than that of the basic decision tree for the breast cancer dataset, demonstrating ESBCP's ability to enhance the performance of the proposed model. The confusion matrix is mainly utilized to describe the performance of this ESBCP model. It's been given below:

This ESBCP model is used to determine the precision, recall, accuracy, and F-measure using the following equations.

$$Accuracy\ =\ (TP+TN)/\ (TP+FP+TN+FN) \tag{6}$$

$$Precision(Pr)\ = TP/(TP+FP) \tag{7}$$

$$Recall(Re) = TP/(TP+FN) \tag{8}$$

$$F-measure =\ (2\ *\ Pr\ *\ Re)\ /\ (Pr\ +\ Re) \tag{9}$$

Where, TP →True Positive, TN → True Negative, FP → False Positive, FN → False Negative.

Table 4. Confusion matrix

		Predicted: NO	Predicted: YES
Actual	NO	TN	FP
	YES	FN	TP

Table 5. Performance Comparison with the Proposed Model (ESBCP)

	Accuracy of Decision Tree (%)
(4)	91.23
(5)	71.43
(9)	75
(12)	88.8

Continued on next page

Table 5 continued

(13)	87.12
(18)	93.85
(19)	94
(20)	93
Simple DT	93.42
Proposed Model- ESBCP	94.0**1**

Compares the performance with the existing state-of-art models. It is observed that the existing models failed to utilize proper data pre-processing and feature selection techniques to get better insights. As per the result obtained, the proposed ESBCP system is better with respect to accuracy than the above mentioned approaches.

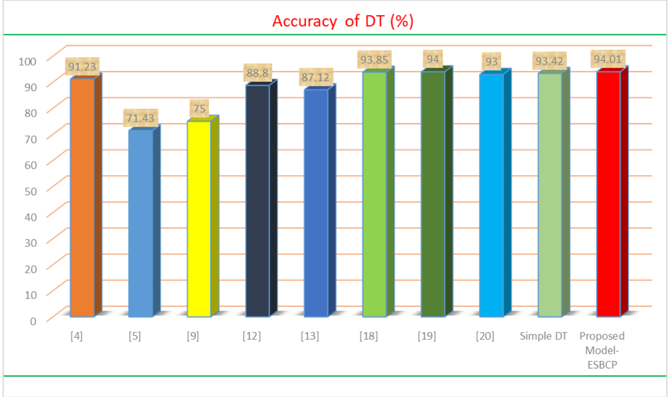

Fig 3. Performance comparison with the state-of-art models

To overcome the limitations of accuracy, precision performance metric was utilized in this work. It determines the proportion of positive predictions i.e. actually correct. In this case, out of 239 malignant cases, 94 percent correct prediction was recorded, whereas for simple DT, it merely recognized 87 percent correctly. For any transparent expert system, it is very much desirable to detect genuine malignant cases positively. The proposed model ESBCP, enhance the performance by utilizing UFS to consider relevant feature set. It is observed that the higher amount of precision off course, enhance the performance of the proposed model but cases still exists where actual positive malignant cases were predicted incorrectly. To emphasis on this, recall was considered. As per the result, 94 percent recall was recorded whereas simple DT was found to be at 90 percent. Figures 4 and 5 showed the precision and recall comparison with simple decision tree (DT) and ESBCP.

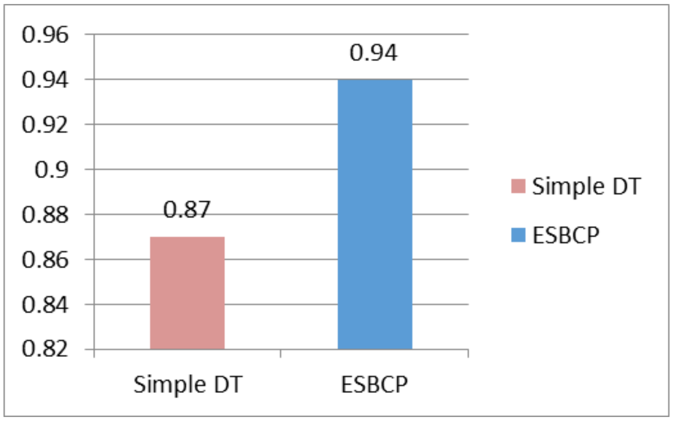

Fig 4. Precision comparison graph for BC

For any models better performance, it is very much essential to understand the raw data properly. To get the better insight, outliers and missing values were eliminated from the dataset. The dataset was normalized and balanced accordingly. After the

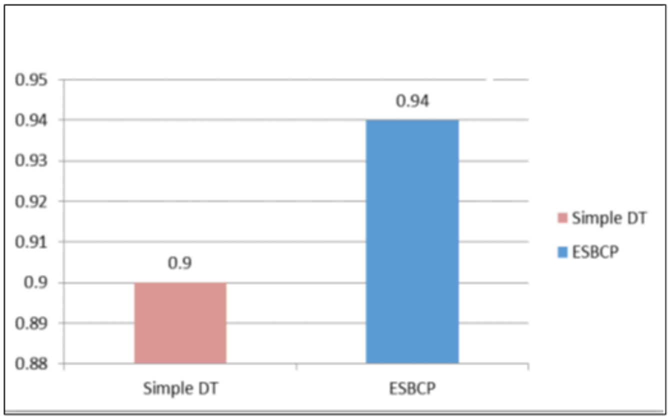

Fig 5. Recall comparison graph for BC

data preprocessing, features of the dataset was selected based on the computation by the UFS which utilized a heuristic search method and stochastic hill climbing method. The proposed model outperformed simple DT in all categories.

Although the model only achieved 0.59% more accurate than simple decision tree, but for huge amount of data this mere change will also reflect a greater impact on the scale of performance. Since breast cancer is the world's most life-threatening fatal disease among women, even this 0.59% accuracy improvement could also save several people's life.

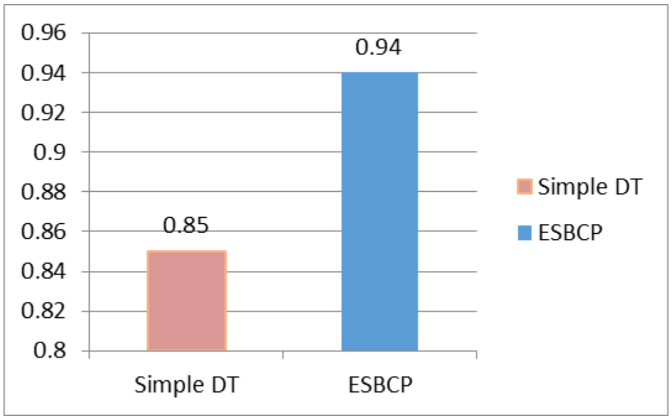

Fig 6. F- Measures comparison graph for BC

4 Conclusion

The proposed model uses a decision tree and the UFS algorithm, implying that the lazy learning method beats the decision-tree strategy as shown in the abobe Table 1. A feature selection model named Undiluted Feature Set (UFS) has been developed for this purpose. In this feature selection model, the data pre-processing phase takes the raw dataset of breast cancer and preprocessed it by deleting the irrelevant features using the Remove unnecessary information, Missing Data Handling. and Normalized Data steps. The relevant data is used for feature selection. Finally, the dataset was split in two parts for training (90%) and testing (10%). Eventually after training, the model was tested and found to me more accurate than its close counterpart by 0.59% w.r.t. accuracy. Therefore, ESBCP employs symptomatic features to diagnose breast cancer, saving time and money while also identifying the high risk of breast cancer at an early stage. The performance of the expert model was also validated using performance measures such as recall, f-measure, and precision and further compared with the simple decision tree and the proposed model ESBCP. Even this 0.59% increase in accuracy has the potential to save the lives of countless individuals considering breast cancer is the most lethal illness that affects women worldwide.

In the future, the performance of the ESBCP model can be enhanced by learning it with a different set of data and then utilizing various preprocessing techniques to remove all irrelevant and extraneous data. Other machine learning classifiers like

Artificial Neural Network (ANN), Support Vector Machine (SVM), k-Nearest Neighbor (k-NN), Logistic Regression (LR), etc. can be utilized to apprehend the necessary changes for further scope of improvement. Further research required to observe necessary requirements to utilise the feature selection technique UFS to greater effect for ensemble and hybrid algorithms also.

References

1) Das AK, Biswas SK, Mandal A. Transparent Decision Support System for Breast Cancer (TDSSBC) to Determine the Risk Factor. *Lecture Notes in Electrical Engineering*. 2021;p. 265–274. Available from: https://doi.org/10.1007/978-981-16-5078-9_23.
2) Das AK, Biswas SK, Bhattacharya A, Alam E. Introduction to Breast Cancer and Awareness. *2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS)*. 2021;1:227–232. Available from: https://doi:10.1109/ICACCS51430.2021.9441686.
3) Lei S, Zheng R, Zhang S, Wang S, Chen R, Sun K, et al. Global patterns of breast cancer incidence and mortality: A population-based cancer registry data analysis from 2000 to 2020. *Cancer Communications*. 2021;41(11):1183–1194. Available from: https://doi.org/10.1002/cac2.12207.
4) Rajaguru H, R SCS. Analysis of Decision Tree and K-Nearest Neighbor Algorithm in the Classification of Breast Cancer. *Asian Pacific Journal of Cancer Prevention*. 2019;20(12):3777–3781. Available from: https://doi:10.31557/APJCP.2019.20.12.3777.
5) Ghani MU, Alam TM, Jaskani FH. Comparison of Classification Models for Early Prediction of Breast Cancer. *International Conference on Innovative Computing (ICIC)*. 2019;2019:1–6. Available from: https://doi:10.1109/ICIC48496.2019.8966691.
6) Tian JXX, Zhang J. Breast cancer diagnosis using feature extraction and boosted C5.0 decision tree algorithm with penalty factor. *Mathematical Biosciences and Engineering*. 2022;19(3):2193–2205. Available from: https://doi:10.3934/mbe.2022102.
7) Khan F, Khan MA, Abbas S, Athar A, Siddiqui SY, Khan AH, et al. Cloud-Based Breast Cancer Prediction Empowered with Soft Computing Approaches. *Journal of Healthcare Engineering*. 2020;2020:1–16. Available from: https://doi.org/10.1155/2020/8017496.
8) Tabrizchi H, Tabrizchi M, Tabrizchi H. Breast cancer diagnosis using a multi-verse optimizer-based gradient boosting decision tree. *SN Applied Sciences*. 2020;2(4):1–9. Available from: https://doi.org/10.1007/s42452-020-2575-9.
9) Assegie TA, Tulasi RL, Kumar NK. Breast cancer prediction model with decision tree and adaptive boosting. *IAES International Journal of Artificial Intelligence (IJ-AI)*. 2021;10(1):184. Available from: https://doi:10.11591/ijai.v10.i1.pp184-190.
10) Mangukiya M, Vaghani A, Savani M. Breast Cancer Detection with Machine Learning. *International Journal for Research in Applied Science and Engineering Technology*. 2022;10(2):141–145. Available from: https://doi.org/10.22214/ijraset.2022.40204.
11) Khan M, Islam M, Sarkar S, Ayaz S, Ananda FI, Tazin MK, et al. Machine Learning Based Comparative Analysis for Breast Cancer Prediction. *Journal of Healthcare Engineering*. 2022;2022:1–15. Available from: https://doi.org/10.1155/2022/4365855.
12) Gayakwad G. Breast Cancer Detection Using Machine Learning Classifier. *International Journal of Multidisciplinary and Current Educational Research (IJMCER)*. 2021;3(4).
13) Assegie TA, J SS. A Support Vector Machine and Decision Tree Based Breast Cancer Prediction. *International Journal of Engineering and Advanced Technology*. 2020;9(3):2972–2976. Available from: https://doi:10.35940/ijeat.A1752.029320.
14) Botcha VM, Kolla BP. Predicting Breast Cancer using Modern Data Science Methodology. *International Journal of Innovative Technology and Exploring Engineering*. 2019;8(10):4444–4446. Available from: https://doi:10.35940/ijitee.J1077.0881019.
15) Laghmati S, Tmiri A, Cherradi B. Machine Learning based System for Prediction of Breast Cancer Severity. *2019 International Conference on Wireless Networks and Mobile Communications (WINCOM)*. 2019;2019. Available from: https://doi:10.1109/WINCOM47513.2019.8942575.
16) Naji MA, Filali SE, Aarika K, Benlahmar EH, Abdelouhahid RA, Debauche OA. Machine Learning Algorithms For Breast Cancer Prediction And Diagnosis. *Procedia Computer Science*. 2021;191:487–492. Available from: https://doi.org/10.1016/j.procs.2021.07.062.
17) Sivapriya J, Kumar A, Sai SS, Sriram S. Breast cancer prediction using machine learning. *International Journal of Recent Technology and Engineering (IJRTE)*. 2019;8(4):4879–4881. Available from: https://doi:10.35940/ijrte.D8292.118419.
18) Ak MF. A Comparative Analysis of Breast Cancer Detection and Diagnosis Using Data Visualization and Machine Learning Applications. *Healthcare*. 2020;8(2):111. Available from: https://doi.org/10.3390/healthcare8020111.
19) Hazra R, Banerjee M, Badia L. Machine Learning for Breast Cancer Classification With ANN and Decision Tree. *11th IEEE Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*. 2020. Available from: https://doi:10.1109/IEMCON51383.2020.9284936.
20) Abdulrahman BF, Hawezi RS, Sm MR, Kareem SW, Ahmed ZR. Comparative Evaluation of Machine Learning Algorithms in Breast Cancer. *Qalaai Zanist Journal*. 2022;7(1):878–902. Available from: https://doi.org/10.25212/lfu.qzj.7.1.34.
21) Boukerche A, Zheng L, Alfandi O. 2020. Available from: https://doi.org/10.1145/3381028.