TPM Vol. 32, No. S8, 2025
ISSN: 1972-6325
https://www.tpmap.org/

EARLY BREAST CANCER PREDICTION USING MACHINE LEARNING FRAMEWORK

DR. MOUSOOMI BORA[1*], DR. RUPAM BARUAH[2], DR. DIGANTA BAISHYA[3]

[1*] DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING, THE ASSAM KAZIRANGA UNIVERSITY, JORHAT, ASSAM, INDIA, EMAIL mousoomi@kzu.ac.in
[2]DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING, JORHAT ENGINEERING COLLEGE, JORHAT, ASSAM, INDIA, EMAIL rupam.baruah.jec@gmail.com
[3]DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING, JORHAT ENGINEERING COLLEGE, JORHAT, ASSAM, INDIA, EMAIL dbaishya@gmail.com

Abstract. Breast cancer is considered as the second most diagnosed cancer in women caused by various factors like social, economic, or psychological. Although, often relies on mammography and biopsy reports, Machine learning and Artificial Intelligence have played a significant role in breast cancer research. This study highlights the importance of Machie learning in breast cancer prediction and proposes a Machine learning framework that comprises three models based on one of the popular Machine Learning algorithms - Support vector machine (SVM). An empirical study is performed using 'linear', 'poly' and 'rbf' kernels to evaluate the performance of Support Vector Machine in Model 1. To enhance the prediction performance, feature engineering techniques are employed to reduce model's complexity as well as computation time in Model 2. As a base learner, Support Vector Machine is attempted to improve its prediction performance using ensemble learning followed by 10-fold cross-validation in Model 3. Results show that the Model 3 has greatly improved its accuracy up to 99.34% compared to Model 1 and Model 2 while SVM with 'poly' kernel is used as one of the diverse models of stacking, a highly accurate and balanced ensemble model.

Keywords: Support Vector Machine (SVM), Kernel, Feature Engineering, Cross-validation, Ensemble Learning

1. INTRODUCTION

One of the leading causes of deaths among women world-wide is breast cancer. The research shows that early prediction of this type of cancer can improve the death rate to a greater extent. The mutation or abnormal changes in genes of body cells is considered as the major cause of cancer. As the genes are liable for regulating the growth of cells, abnormal changes of genes may have the ability to divide the cells in an uncontrolled manner and produce more and more cells and form a tumor. Regular breast cells become cancerous because of Deoxyribonucleic acid (DNA) mutations in the genes. The tumor can be benign or malignant based on degree of DNA mutation. 'Benign' tumors are not considered as cancerous, however 'Malignant' tumors are graded as cancerous. Breast cancer is a type of cancer where malignant tumors are developed in the breast cells of the lobules which are milk producing glands. Researchers have been exploring all possible ways of predicting cancer in its preliminary stage. Machine Learning (ML) has become one of the indispensable technologies in reducing number of death rates among the women in a greater extent. Decision Tree, Random Forest, Support Vector Machine (SVM), Linear Regression etc. are the ML algorithms used in predicting diseases including breast cancer. There are multiple causes that a woman is diagnosed as a breast cancer patient. The risk factors are primarily age, overweight, life style, lack of exercise, family history, genetics, radiation to chest or face before 30 years of age, certain changes in breast, race or ethnicity, history of pregnancy, menstruation history younger than age 12, use of Hormone Replacement Therapy, alcohol consumption, smoking etc. High level research work and awareness of breast cancer among women have helped the health care practitioners in their diagnosis as well as treatment of the disease. It has gradually decreased the total death counts associated with breast cancer and capable of increasing number of survivalists.

2. RELATED WORK

Factors like early detection, technological advancement in the medical field and better understanding of the disease are helping doctors in prognosis of Breast Cancer. ML based frameworks have been adapted by researchers so that they can contribute to the society by reducing death rates of breast cancer diagnosed patients. It is such a humble step taken by the researchers and technologists that they have come forward to work hand in hand with the health care professionals in early prediction and detection of breast cancer. A standard Wisconsin Dataset for Breast Cancer analysis (WDBC) is used in this study to implement a ML framework as a part of breast cancer research. The literature confers a number of ML algorithms such as Logistic Regression, Naïve Bayes, Decision Trees, Artificial Neural Networks, K-Nearest Neighbor, SVM and so on [1-7]. BCP-T1F-SVM is an expert system that determines various stages and types of cancer. It is a hybrid model of fuzzy logic and SVM and capable of predicting breast cancer with higher precision rate [5]. It has been realized that SVM has outperformed many of the existing algorithms in predicting breast cancer [8]. Feature selection and

dimensionality reduction methods are used with the ML algorithms with an attempt to improve models' prediction accuracy. Backward feature elimination is employed by researcher in Multiple SVM-RFE (MSVM-RFE) to improve the performance of existing SVM-RFE. Results show that MSVM-RFE yields more accurate result compared to the SVM-RFE [6]. It is suggested that combining multiple classifiers often provide better performance than single classifier. It is an active research field of pattern classification [9]. Ensemble ML with different tuning parameters improves classification performance with emphasis on sensitivity [10]. Random forest while used as a weak learner in AdaBoost greatly improves its accuracy and stability by reducing overfitting problems in breast cancer survivability prediction model [11]. Ensemble of Linear-SVM based on the bagging and Ensemble of RBF-SVM based on the boosting can be used as better approaches for small-scale datasets [12]. Stacking ensemble method with f-test feature selection is an efficient and consistent approach to predict breast cancer [13].
In this study, keeping accuracy as the major performance factor, we moved forward and emphasized more towards the improvement of model's accuracy using SVM as the primary component of the proposed framework in early prediction of breast cancer.

3. Description of dataset

A standard UCI ML Breast Cancer Wisconsin (diagnostic) dataset is downloaded from UCI machine learning repository to carry out this study. The dataset contains 569 instances with 30 features where features are extracted from a digitized image of fine needle aspirate (FNA) of a breast mass. The characteristics of the cell nuclei are represented by the features available in the images. The dataset is used to implement the proposed approach as a binary classifier having two target classes – Malignant with 212 instances and Benign with 357 instances. An overview of the dataset is displayed in Table 1.

Table 1 Feature summary of Breast Cancer Wisconsin (diagnostic) dataset

Sl. No.	Information of features
1	ID Number
2	Diagnosis (M = malignant, B = benign)
3	Radius (mean of distances from center to points on the perimeter)
4	Texture (standard deviation of gray-scale values)
5	Perimeter
6	Area
7	Smoothness (local variation in radius lengths)
8	Compactness (perimeter^2 / area - 1.0)
9	Concavity (severity of concave portions of the contour)
10	Concave points (number of concave portions of the contour)
11	Symmetry
12	Fractal dimension (coastline approximation – 1)

The first and second features in Table 1 are counted as integer variables. On the other hand, remaining ten features provide real values with three types of measures i.e. mean, standard error and worst/ largest value computed for each feature taken from the image. The ID column is removed from the dataset being an insignificant input in breast cancer prediction. Diagnosis column is used as the target feature leaving other 30 features to act as input features for the ML models used in breast cancer prediction. The overall size of the dataset includes 569 instances and 30 input features. The input dataset is divided into two parts – Training and Test data in 80:20 ratios. This way 455 records are used for training and remaining 114 records are used for testing the ML models with 30 available features.

4. METHODOLOGY WITH THEORETICAL BACKGROUND

The primary objective of this study is to predict breast cancer among females in early days based on information collected from their mammogram and biopsy reports. The proposed ML framework for early prediction of breast cancer is portrayed in Fig. 1.

TPM Vol. 32, No. S8, 2025
ISSN: 1972-6325
https://www.tpmap.org/

Open Access

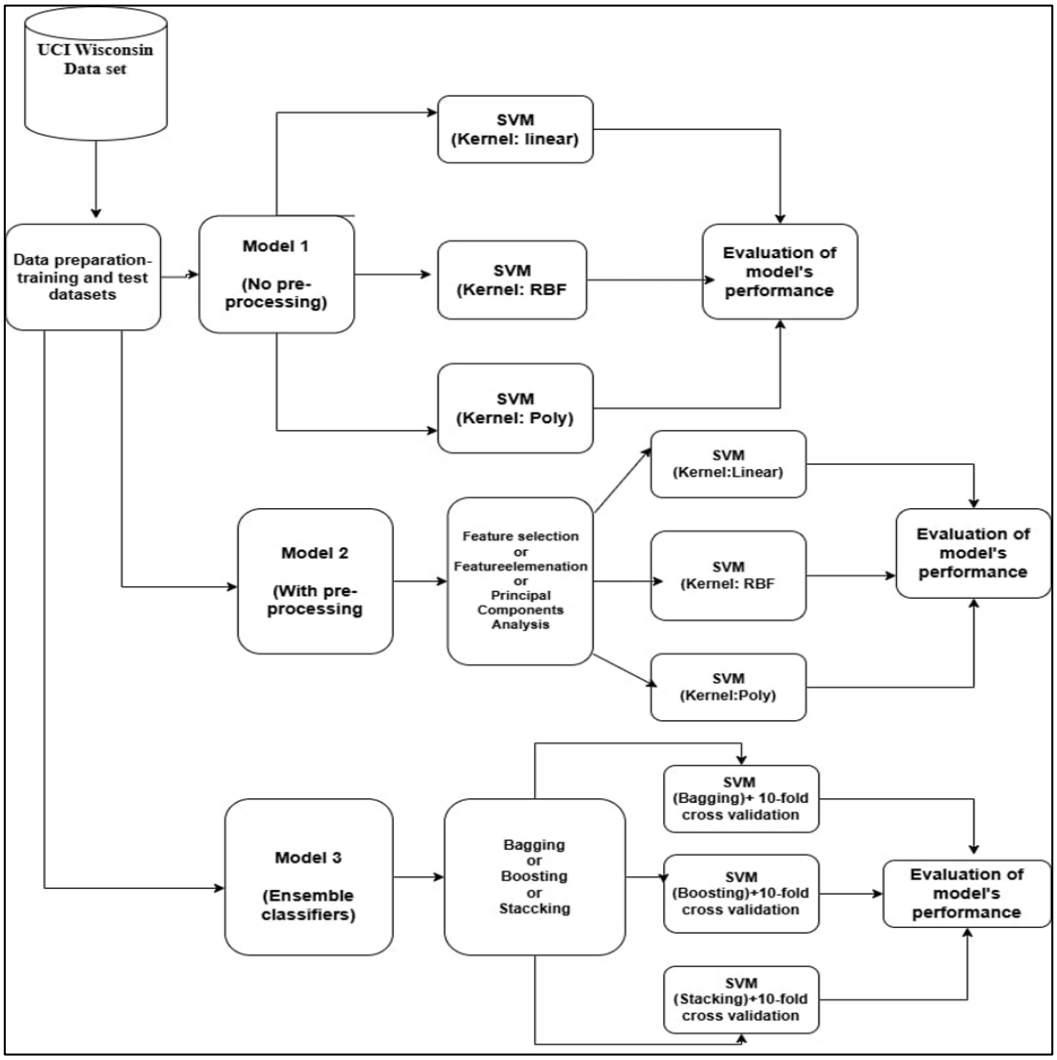

Fig. 1 Steps used in Machine learning framework used in Breast Cancer prediction

As sown in Fig. 1, the framework is composed of three ML models. The models are based on kernel SVM. SVM is a supervised ML algorithm and can be used for both classification and regression problem. Though, SVM is mostly used in solving classification problems. SVM is considered as one of the best ML algorithms due to its ability to handle high dimensional datasets. SVM can be also used to perform non-linear classification using kernel function. A dataset having small training dataset also can be trained successfully using SVM. Considering all these merits, SVM is selected as the base line prediction model under this study. SVM tries to find the best decision boundary, knowns as hyperplane to separate the training data set into predefined number of classes. The closest data points to the hyperplane are called support vectors. The optimal hyper-plane in SVM helps to maximize the distances between nearest data points from either class. The primary objective of SVM is to maximize the distance between support vectors and hyperplane called margin. For wider margin between the support vectors from either side, SVM provides better classification accuracy. Kernels in SVM are mathematical function that transforms existing input features to some new features. The new features created by kernel are the key for SVM to find the nonlinear decision boundary. kernel acts as a hyper parameter in SVM with its variants such as linear, poly, radial basis function(rbf), and sigmoid.

The model 1 of the proposed ML framework is trained on the training dataset having 30 input features without employing any pre-processing steps. The model 2 has explored feature engineering techniques. The model 3 has implemented ensemble learning and 10-fold cross validation. The steps used in Model 1, Model 2 and Model 3 are described as below.

4.1 Model 1: Kernel SVM with no pre-processing

As a preliminary study, a good number of ML classifiers are implemented on a standard Wisconsin Dataset for Breast Cancer analysis (WDBC) and performance of each classifier is measured using accuracy score. Model 1 is based on kernel SVM where three types of kernels are considered – 'linear', 'poly' and 'rbf'. SVM with 'linear' kernel helps in transforming input features present in a linearly separable dataset. SVM with 'poly' kernel is used in non-linearly separable dataset. SVM with 'rbf' kernel transforms a dataset into an infinite number of dimensional spaces. The performance of Model 1 with respect to three kernel types will be discussed in next section of results and discussion.

4.2 Model 2: Kernel SVM with pre-processing

Feature engineering is employed in Model 2 as a pre-processing step. The objective of applying feature engineering on WDBC dataset is to reduce the computation time and overfitting problem faced by the ML model. Three types feature engineering is employed in the dataset prior to train the model based on kernel SVM.

4.2.1 Pre-processing using feature selection method

Feature selection is a dimensionality reduction technique which is applied to improve the performance of classification or prediction model by discarding insignificant or less contributing features from the dataset. Model 2 uses 'Select From Model' as one of the feature selection methods where features are selected based on the mean of the importance of all the features available in the dataset. As a result of feature selection, total number of features have reduced to 9 from original dataset of 30 input features. Table 2 lists the name of the features as recommended by the feature selection method i.e. 'Select From Model'.

Table 2 Features are selected by feature selection method ('Select From Model')

Sl. No	Feature Name
1.	Mean radius
2.	Mean concavity
3.	Mean symmetry
4.	Texture error
5.	Worst smoothness
6.	Worst compactness
7.	Worst concavity
8.	Worst concave points
9.	Worst symmetry

4.2.2 Pre-processing using feature elimination method

Recursive Feature Elimination (RFE) is used as one of the feature elimination methods in Model 2. In RFE, features are recursively eliminated and only the selected features are used to build the prediction model. RFE takes two function parameters – the estimator and number of features to select. Based on the type of estimator or ML classifier, RFE recursively removes one feature at a time. The least performing features are eliminated first and keep on removing the less performing features one at a time. With RFE, Model 2 is expected to perform better with selected features in the dataset compared to the model built on all the available features in the original data set.

4.2.3 Pre-processing using feature extraction method

Principal Component Analysis (PCA) is an orthogonal feature extraction technique that speeds up the training process in a ML model implemented on a large feature vector. By linear combinations of input features, PCA transforms the input dataset to already defined number of principal components which are not corelated to each other. Most of the information within the original features is compressed into the first principal component. In PCA, maximum information is extracted from the original features and stored in the first principal component. While majority of remaining information is stored in the second principal component and so on. It is hard to interpret the principal components generated by PCA and don't have any real meaning since they are constructed as linear combinations of the original features. The outcome of the application of PCA will be greatly effective if we scale down the features in the dataset before its application on the dataset.

4.3 Model 3: Kernel SVM with ensemble learning and 10-fold cross-validation

Model 3 is based on ensemble ML. Ensemble ML techniques are used to improve the performance of predictive models where multiple ML models are strategically constructed to solve bias and variance problems. Ensemble learning is the art of combining a diverse set of learners together to improve the stability and predictive capacity of the model. There are two types of ensemble learning methods - sequential ensemble and parallel ensemble methods. In sequential ensemble methods, the base learners are stitched together in a sequential manner. Where as in parallel ensemble methods, base learners are connected to each other parallelly. Different types of ensemble learning techniques are explored by researchers for predictive analytics tasks such as - bagging, boosting and stacking.

Bagging or bootstrap aggregation is used to reduce the variance of a classification or regression type ML model by averaging the predicted outcome of several ML models. Bagging implements similar learners on small subsets of training data and thereafter produces its final outcome based on majority voting for output produced by classification and averaging for regression output. It is a parallel ensemble technique where all input features are included. In this approach, random samples of data from the training dataset are selected enabling duplication of sample instances. Boosting is an iterative technique which fine-tunes the weight of an observation, based on the outcome produced by previous model. Boosting is a sequential ensemble technique that typically works by combining several weak learners. Unlike bagging, boosting focuses on reducing bias rather than variance where each new model tries to correct the errors made by the previous models. Boosting type of ensembling method develops a strong prediction model by combining the predictions of weak learners, and thus improving model's performance over time. Unlike bagging and boosting, stacking combines different types of models rather than the same type to enhance predictive performance of ML models. Multiple base models are trained independently on the same training data in stacking. The predicted output from each independent model forms a new set of input features for the meta model, representing the view of independent base models. The meta-model in stacking learns the best way to integrate the outputs from diverse base models leading to a more balanced and accurate result.

Model 3 is validated using 10-fold cross-validation technique. Cross-validation is a technique where the ML model is trained using subsets of the dataset and then validated using the complementary subset of the data-set. The basic resolution of cross-validation is to evaluate how the model would perform in an unknown dataset. Cross-validation technique is used to detect overfitting. i.e. failing to generalize the model on similar datasets. Cross - validation affords more precise

estimate of out-of-sample accuracy. Using this validation technique, efficient and well-organized use of data is possible as every observation is used for both training and testing.

5. EXPERIMENTAL RESULTS AND DISCUSSION

Throughout this research, UCI ML Breast cancer Wisconsin (diagnostic) dataset is used by all the three models of proposed ML framework. The ML framework has explored various steps of ML and implemented those on the dataset under study to evaluate model's performance as discussed below.

5.1 Independent machine learning algorithms in breast cancer prediction

As a preliminary study, nine individual ML algorithms are trained on the training dataset of the WDBC dataset and tested on the unseen test data. Table 3 shows the accuracy of each independent ML classifier on the same dataset.

Table 3 Breast cancer prediction accuracy for existing machine learning algorithms

Algorithm	Accuracy
Logistic Regression	92.98
K-Nearest Neighbor	92.39
Gaussian Naïve Bayes	91.22
Decision Tree	86.54
Random Forest	89.47
SVM with 'linear' kernel	91.22
SVM with 'poly' kernel	92.26
SVM with 'rbf' kernel	85.96
Artificial Neural Network (ANN)	92.39

After implementing each of the algorithm as listed in Table 3, from Logistic Regression to Artificial Neural Network (ANN), the accuracy score is found between 65.96 % to 92.98 %.

5.2 Kernel SVM without any pre-processing steps

Table 4 shows the performance of SVM on three types of kernels – linear, poly and rbf.

Table 4 Breast cancer prediction accuracy for 3 different kernels in SVM

SVM with different type of Kernels	Accuracy
SVM using 'linear' kernel	**0.96**
SVM using 'poly' kernel	**0. 63**
SVM using 'rbf' kernel	**1.0**

From Table 4, it has been found that accuracy score of SVM with 'Poly' kernel is minimum. On the other hand, the accuracy of SVM with 'rbf' kernel is maximum while implemented on the same WDBC dataset for breast cancer prediction.

5.3 SVM with feature selection

As a part of Model 2, feature selection using "Select from Model" in SVM while implemented on WDBC dataset, number of input features are reduced and model is built on the dataset with reduced features. Table 5 shows the outcome of Model 2 in terms of accuracy and training time with and without feature selection on the dataset under study.

Table 5 Accuracy and training time with and without feature selection ('Select From Model')

Without feature selection	Accuracy	Training Time	With feature selection	Accuracy	Training Time
SVM ('linear') + "Select From Model"	**92.61**	**3.9 S**	SVM ('linear') + "Select From Model"	**91.31**	**196 MS**
SVM ('poly') + "Select From Model"	**93.22**	**2.1 S**	SVM ('poly') + "Select From Model"	**93.5**	**150 MS**
SVM ('rbf') + "Select From Model"	**93.82**	**3.6 S**	SVM ('rbf') + "Select From Model"	**93.22**	**186 MS**

As shown in Table 5, the accuracy and training time of the model are computed for 30 iterations and it is observed that no significant improvement is observed in accuracy of Model 2. However, total training time is reduced after feature selection is applied using 'Select From Model'.

5.4 SVM with feature elimination

This experiment is used to evaluate the performance of SVM using another preprocessing step called Recursive Feature Elimination (RFE) (Table 6).

TPM Vol. 32, No. S8, 2025
ISSN: 1972-6325
https://www.tpmap.org/

Open Access

Table 6 Accuracy and training time with and without feature elimination ('RFE')

Without feature elimination	Accuracy	Training Time	With feature elimination	Accuracy	Training Time
SVM ('linear') + RFE	**93.61**	**1.75 s**	SVM ('linear') + RFE	**93.31**	**161 ms**
SVM ('poly') + RFE	**95.22**	**2.1 s**	SVM ('poly') + RFE	**94.95**	**156 ms**
SVM ('rbf') + RFE	**93.82**	**3.6 s**	SVM ('rbf') + RFE	**93.22**	**196 ms**

As shown in Table 6, it is observed that no improvement is observed in accuracy of Model 2 after implementing feature elimination step. However, total training time is shown to be reduced after RFE is applied as one of the feature elimination techniques.

5.5 SVM with feature extraction

SVM while integrated with PCA as a feature extraction technique enables both the classes to be linearly separable up to some extent where benign class is spread out as compared to the malignant class (Fig. 2). Table 7 shows the results after applying PCA as a pre-processing step for kernel SVM.

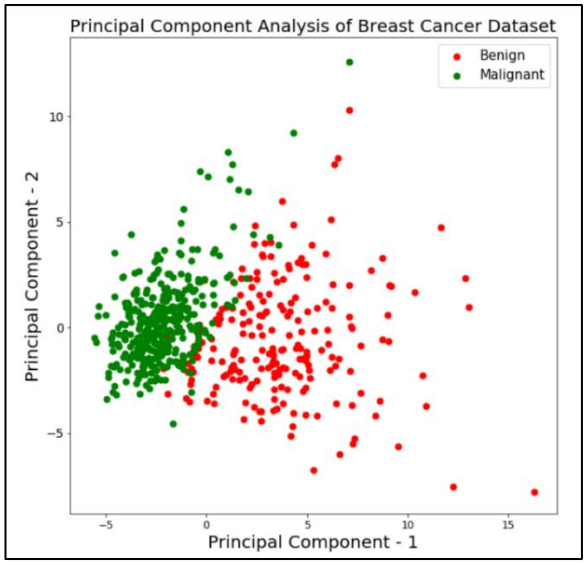

Fig. 2 Principal Component Analysis (PCA) of Breast cancer dataset

Table 7 Accuracy of kernel SVM with and without feature extraction

Without feature extraction	Accuracy	With feature extraction	Accuracy
SVM ('linear') + PCA	**90.10**	SVM ('linear') + PCA	**92.10**
SVM ('poly') + PCA	**92.18**	SVM ('poly') + PCA	**92.98**
SVM ('rbf') + PCA	**71.45**	SVM ('rbf') + PCA	**81.05**

The results show that SVM with 'poly' kernel shows accuracy enhancement over 'rbf' and 'linear' kernels.

5.6 SVM with ensemble learning and 10-fold cross-validation

Ensemble learning and 10-fold cross-validation are used to improve the accuracy of SVM. Table 8 shows the performance of Model 3 while bagging, boosting and stacking are used. A stack of weak learners – Logistic Regression, Kernel SVM and Decision tree is used. The predicted outcomes of weak learners are averaged in stacking approach on the basis of majority vote.

Table 8 Accuracy of SVM with ensemble learning and 10-fold cross-validation

SVM with Bagging + 10- fold cross validation	**Accuracy**
SVM ('linear) with Bagging and 10- fold cross validation	94.64
SVM ('rbf) with Bagging and 10- fold cross validation	95.04
SVM ('poly') with Bagging and 10- fold cross validation	96.04
SVM with Boosting + 10- fold cross validation	**Accuracy**
SVM ('linear') with Boosting and 10- fold cross validation	94.94
SVM ('rbf') with Boosting and 10- fold cross validation	95.38
SVM ('poly') with Boosting and 10- fold cross validation	96.88
SVM with Stacking + 10- fold cross validation	**Accuracy**
Voting Classifier: Ensemble of - LogisticRegession, SVM ('linear'), Decison Tree and 10- fold cross validation	96.92

Voting Classifier: Ensemble of - LogisticRegession, SVM ('rbf'), Decison Tree and 10- fold cross validation	98.46
Voting Classifier: Ensemble of - LogisticRegession, SVM('poly'), Decison Tree, Random Forest and 10- fold cross validation	99.34

As shown in Table 8, SVM with 'poly' kernel while used with bagging and 10-fold cross validation provides better accuracy compared to 'linear' and 'rbf' kernel of SVM. Similarly, SVM with 'poly' kernel with Boosting and Stacking shows better performance. Out of bagging, boosting and stacking, SVM with 'poly' kernel outperforms other ensemble models experimented in this study having accuracy score up to 99.34%.

6.CONCLUSION AND FUTURE WORK

An empirical study is performed in this paper to analyse the performance of proposed ML framework while dealing with breast cancer prediction. Among several ML algorithms, it is found that SVM provides the best prediction accuracy compared to the remaining algorithms. Thus, SVM is chosen for the rest of the research work in early prediction of breast cancer. The study analyses the performance of three ML models as a part of ML framework. Model 1 is developed without any pre-processing step. Model 2 has used pre-processing steps like feature selection, feature elimination and feature transform. Model 3 is designed using ensemble learning methodology. Although, application of feature selection, feature elimination or feature extraction steps have not shown much differences in performance in terms of accuracy, training time is reduced in Model 2 after applying these feature engineering techniques viz. Select From Model, RFE and PCA. SVM using 'poly' kernel shows better performance compared to 'linear' and 'rbf' kernels in both Model 2 and Model 3. 10-fold cross-validation has helped to improve model's accuracy in Model 3 especially in stacking compared to bagging or boosting technique. Throughout this research work, it has been found that Model 3 i.e. SVM as one of the base classifiers of Stacking based ensembled model has shown significant improvement in its prediction accuracy compared to Model 1 and Model 2. As a future work, Deep Learning techniques along with advanced optimization techniques can be used to accomplish better outcomes in the field of breast cancer prediction in its early stage.

7.Acknowledgment
The work was supported by Virtualization and Cloud Computing Laboratory, The Assam Kaziranga University. The authors greatly acknowledge Jorhat Engineering College (Assam, India) for other technical supports.

Availability of data and material (data transparency)
Breast Cancer Wisconsin (Diagnostic) Data Set
http://mlr.cs.umass.edu/ml/datasets/Breast+Cancer+Wisconsin+(Diagnostic)
1995

REFERENCES

1. Ali A., Khan U., Kim M. (2009). A survey of prediction models for breast cancer survivability. In Proceedings of the International Conference on Interaction Sciences: Information Technology, Culture and Human, Seoul, Korea, pp.1259 – 1262.
2. Aruna S., Rajagopalan S.P., Nanda K.L.V. (2011). Knowledge based analysis of various statistical tools in detecting breast cancer. In Proceedings of the International Conference on Computer Science, Engineering and Applications, Chennai, India, pp. 37– 45.
3. Cruz, J.A., Wishart, D.S. (2006). Applications of machine learning in cancer prediction and prognosis, Cancer Informatics, 2, 59–78. https://doi.org/10.1177/117693510600200030
4. Gayathri, B.M., Sumathi, C.P., & Santhanam, T. (2003). Breast cancer diagnosis using machine learning algorithms - a survey. International Journal of Distributed and Parallel Systems, 4(3), 105–112.
5. Khan F., Khan M. A., Abbas S., Athar A., Siddiqui S. Y., Khan A. H., Saeed M. A., and Hussain M. (2020). Cloud-Based Breast Cancer Prediction Empowered with Soft Computing Approaches
6. Nurul, N.M. H., Nies, H.W., Chan, W. H., Mohd, S. M., Safaai, D., Shahreen, K. (2017). Improved Support Vector Machine Using Multiple SVM-RFE for Cancer Classification. International Journal on Advanced Science, Engineering and Information Technology, 7, 1589-1594. https://doi.org/10.18517/ijaseit.7.4-2.3394
7. Shajahaan, S.S., Shanthi, S., Manochitra, V. (2013). Application of data mining techniques to model breast cancer data. International Journal of Emerging Technology and Advanced Engineering, 3(11), 362– 369.
8. Senturk, Z.K., Kara, R. (2014). Breast cancer diagnosis via data mining: performance analysis of seven different algorithms. Computer Science & Engineering: An International Journal, 4(1), 35–46.
https://doi.org/10.5121/cseij.2014.4104
9. Kittler, J., Hatef, M., Duin, R.P.W., Matas, J. (1998). On combining classifiers. IEEE Transactions on Pattern Analysis and Machine Intelligence, 20(3), 226–239. https://doi.org/10.1109/34.667881
10. Djebbari, A., Liu, Z., Phan, S., & Famili, F. (2008). An ensemble machine learning approach to predict survival in breast cancer. International Journal of Computational Biology and Drug Design, 1(3), 275-333.
https://doi.org/10.1504/ijcbdd.2008.021422

11. Jaree T., Guandong X., Yanchun Z. (2008). AdaBoost algorithm with random forests for predicting breast cancer survivability. In Proceedings of the International Joint Conference on Neural Networks, Piscataway, New Jersey, pp. 3062 – 3069
12. Huang, M.-W., Chen, C.-W., Lin, W.-C., Ke, S.-W., & Tsai, C.-F. (2017). SVM and SVM Ensembles in Breast Cancer Prediction. PLOS ONE, 12(1), e0161501. https://doi.org/10.1371/journal.pone.0161501
13. Dhanya R., Irene R. P., Sai S. A., Sivakumar M., Nair J. J. (2020). F-test feature selection in Stacking ensemble model for breast cancer prediction. Procedia Computer Science, Volume 171, 2020, Pages 1561-1570. https://doi.org/10.1016/j.procs.2020.04.167