



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

IJIRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 13, Special Issue 1, April 2024

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.423

 9940 572 462

 6381 907 438

 ijirset@gmail.com

 www.ijirset.com

Prediction of Breast Cancer Using Machine Learning

Dr. S. Xavier Arockiaraj¹, R. Abarna Lakshmi², B. Elfalin Shruthi³, A. Jencisusan⁴,
S. Uma Devi⁵

Associate Professor, Department of Electronics and Communication Engineering, Adhiyamaan College of Engineering, Krishnagiri, Tamil Nadu, India. ¹

U.G Scholars, Department of Electronics and Communication Engineering, Adhiyamaan College of Engineering, Krishnagiri, Tamil Nadu, India. ^{2, 3, 4, 5}

ABSTRACT: Breast cancer is one of the most prevalent forms of cancer affecting women so, it is necessary to detect BC in the early stage for effective treatment. The dataset obtained by SEER explore the prognostic value of various genetic factors and clinical features in predicting patient survival outcomes. The proposed work starts with preprocessing the dataset, which includes cleaning, normalization to ensure data quality. After preprocessing, The selected features were trained by mc algorithm such as support vector machine, Kernel Naïve Byes, Gaussian Naïve Byes, Neural Network, Ensemble Boosted Trees are applied to build efficient models. Finally, the performance metrics of each model is evaluated. The performance metrics includes Accuracy, True Positive Rate (TPR), Positive Predictive Value (PPV), Area Under the ROC Curve (AUC) and Confusion Matrix (CM) for both validation and testing. Among 5 Machine learning algorithms Ensemble Boosted Trees gives the accuracy of 90.4% which is found to be superior than other state of art of machine learning algorithms available in the literature.

KEYWORDS: - Breast Cancer, Confusion matrix, Ensemble boosted trees, Machine Learning, Survival Prediction

I. INTRODUCTION

Breast cancer is another type of cancer that develops in the cells of the breasts. This cancer also mainly occurs in women and leads to the highest death rate among those women. So, detecting this cancer in order to cure it in the initial stage itself is much needed. Radiologists use the most reliable method called Mammography to detect breast cancer [1]. The disease's mortality rate climbed drastically from 0.96 per 100,000 in 1995 to 4.33 per 100,000 in 2010[9]. Breast cancer remains one of the most prevalent and deadly forms of cancer among women worldwide. Timely diagnosis is crucial for effective treatment and improved survival rates. With advancements in technology, particularly in the realm of machine learning, there is a growing interest in leveraging data-driven approaches for early detection and diagnosis. In this study, the application of machine learning algorithms for the early diagnosis of breast cancer using the Surveillance, Epidemiology, and End Results (SEER) dataset is being focused. SEER provides a rich repository of clinical, demographic, and tumor characteristics data collected from various cancer registries across the United States. By harnessing this extensive dataset, the aim is to develop predictive models which is capable of identifying accurately at the risk of developing BC in individuals.

Machine learning algorithms are used to enable computers to learn from data and make predictions or decisions without being explicitly programmed. This capability allows them to uncover patterns, trends, and insights from large datasets, automate tasks, optimize processes, and improve decision-making in various fields such as healthcare, finance, marketing, and more. Without machine learning algorithms, organizations face challenges in efficiently analyzing large volumes of data, identifying patterns and insights, and making accurate predictions or decisions. Manual analysis of data can be time-consuming, error-prone, and limited in scalability. Moreover, traditional

approaches may struggle to handle complex, unstructured data types such as text, images, and audio. Additionally, without machine learning, businesses may find it difficult to automate repetitive tasks, personalize user experiences, detect anomalies or fraud, and optimize processes. To achieve this aim, machine learning algorithms were employed such as ensemble boosted trees, SVM, kernel naive Bayes, neural networks, and Gaussian naive Bayes. Each algorithm brings its unique strengths and characteristics to the task, allowing us to compare their performance and identify the most effective approach for breast cancer prediction. Furthermore, to ensure robustness and reliability, a rigorous methodology for model evaluation is adopted. The dataset is splitted into training and testing sets using a stratified k-fold cross-validation approach with $k=5$, allocating 90% of the data for training and 10% for testing. This strategy helps mitigate overfitting and ensures that the models generalize well to unseen data. The motivation of the work is of twofold. First, to develop accurate and reliable ML algorithms for early BC detection. Secondly, by helping clinicians in their decision making process of identifying the individuals with BC.

II. RELATED WORKS

Breast cancer develops when cells in the breast start to proliferate uncontrollably. Typically, breast cancer cells produce an X-ray-observable tumor or a lump that can be felt. Breast cancer primarily affects women, while it can also occasionally affect men [2,3]. A number of tests, such as screening tests including mammograms, breast ultrasounds, breast MRIs, etc., are used to diagnose breast cancer. Compared to other cancers, a greater number of women are receiving a diagnosis of breast cancer. Between 2008 and 2017, there was a yearly 0.5 percent increase in female breast cancer cases.

In 2020, 6,85,000 people died worldwide from breast cancer and 2.3 million women received a diagnosis [4]. [5] created a study centered on determining the efficacy of multi-layer perceptrons (MLP) to predict the course of breast cancer using a set of four biomarkers specific to breast cancers. This MLP technique explains the ANN procedure to determine the accuracy of features in the nodal status prediction for breast cancer prognosis. The model is assessed using simple k -fold cross validation and stratified cross validation. 46 patient records for benign breast cancer were the source of data used in the analysis. The final structure is a three-layered network that is trained using the scaled conjugate gradient (SCG) algorithm. Better accuracy of around 63% was demonstrated by the 2-fold cross validation, which also showed improved reliability and sensitivity. [6] created a model using transfer learning and a CNN foundation. CNN is trained from scratch in the first approach. In contrast, the second approach uses a pre-trained VGG-16 model to extract features from mammograms and use those data to construct a neural network classifier. Lastly, the third method uses back-propagation (fine-tuning) to update weights in the last layers of the pre-trained VGG-16 model in order to identify the aberrant regions. The finetuning model is judged to be appropriate for study despite having an accuracy that is just 0.008 more accurate than the second approach. This is because the finetuning model requires significantly less time to complete than the other two ways. Subsequently, the third technique uses 10-fold cross validation on the DDSM database to categorize normal vs abnormal, malignant vs normal, and benign vs normal cases and met 95% accuracy for abnormal instances compared to normal cases.

suggested a sophisticated CNN feature extraction method, similar to Residual Networks (ResNet), GoogLeNet, and Visual Geometry Group Network (VGGNet). Later on, it is applied to the average pooling-based classification of pictures into benign and malignant categories. In order to produce 8000 photos, two datasets were examined, and data augmentation was done first. 6000 of the 8000 photos were used for training, while 2000 images were used for testing. The evaluation process indicates that the GoogLeNet, VGGNet, and ResNet designs have accuracy values of 93.5%, 94.15%, and 94.35%, in that order. However, the combined model's accuracy of 97.525 percent indicates that it outperforms the individual approaches.

To forecast breast cancer using pictures from mammograms. Images with anomalies are categorized as benign and malignant in this approach as well. In this case, thresholding is done after applying a 2D median filter. The thresholding technique produces divided binary photos that were used to extract features, after which test data is fed into the model to provide labels for classification. As classification models, SVM, Random Forest Classifier, Gradient Boost, Decision Tree, K-Nearest Neighbor (KNN), AdaBoost, and Logistic Regression are used. Every method's accuracy is examined, and SVM received the highest accuracy score of 90%. Additionally, SVM displayed the best test score, F-score, recall,

and precision. Gradient Boost classifier, on the other hand, produced the lowest accuracy, at 52%. Consequently, it may be said that algorithms having high precision and recall shows high accuracy.[10] used Decision Tree, k-Nearest Neighbour, and Naïve Bayes, three categorization techniques, for the various datasets. The authors additionally examine the error rate evaluation metrics. The implementation was concentrated on a particular kind of dataset attribute.

III. METHODOLOGY

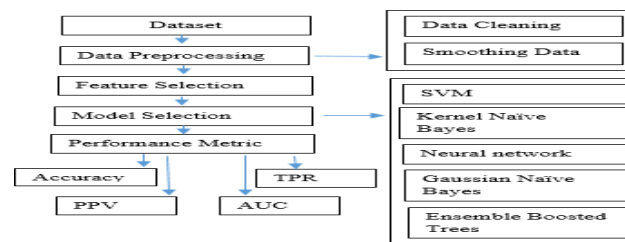


Figure 1: Proposed Work Flow

A. Dataset

Public health experts and academics studying cancer can benefit greatly from the Surveillance, Epidemiology, and End Results (SEER) program. It is kept up to date by the National Cancer Institute (NCI), which gathers information on cancer survival and incidence from a number of community-based cancer registries that account for around 34.6% of the US population. SEER statistics offer insights into cancer trends, treatment outcomes, and inequities by containing information on cancer cases diagnosed within particular geographic areas and demographic groupings. Typically, the SEER dataset contains the following columns: Patient characteristics: Age at diagnosis, gender, marital status, race/ethnicity, and socioeconomic status. Primary site, histology, grade, stage, size, and disease extent are all features of tumors. Details about the diagnosis and course of treatment include the date of the diagnosis as well as information on radiation, chemotherapy, surgery, hormone therapy, immunotherapy, and other treatments received. Additional information: The vital status, reason for death, length of survival, recurrence, and future malignancies. Geographical data includes county, state, urban/rural classification, and SEER registry. A distinct cancer case, complete with comprehensive details about the patient, tumor, diagnosis, therapy, and follow-up, is represented by each row in the SEER dataset. The de-identified dataset safeguards patient privacy while offering rich information useful for clinical research, epidemiological investigations, and health policy analysis. Through the authorized SEER data users' program or the SEER website (seer.cancer.gov), researchers and public health experts can access SEER data.

B. Data Preprocessing

Data preparation, which involves cleaning and transforming raw data into a format appropriate for analysis, is an essential stage in the data analysis pipeline. Handling missing values, eliminating outliers, standardizing or normalizing numerical features, and encoding categorical variables are some of the processes that are usually included in the process. Depending on the type and degree of missingness as well as the particular requirements of the study, handling missing data entails choosing whether to impute missing values or eliminate them completely. To avoid unnecessarily impacting the study results, outliers—data points that differ significantly from the rest of the dataset—may need to be found and either corrected or eliminated.

Data Cleaning- In the framework of the SEER dataset, data cleaning is the methodical process of locating and fixing mistakes, inconsistencies, and missing values in the gathered data. This entails activities like eliminating redundant entries, fixing typos, imputes missing data, and verifies data accuracy and dependability by comparing it to predetermined standards. The SEER dataset attempts to improve the quality of the data through stringent data cleaning techniques, making it appropriate for meaningful analysis and interpretation in epidemiology and cancer research

investigations.

Smoothing Data- The technique of lowering noise or fluctuation in a dataset to highlight underlying trends or patterns is known as "smoothing data." Many statistical methods, including spline interpolation, kernel smoothing, and moving averages, can be used to accomplish this. By reducing short-term fluctuations and emphasizing long-term patterns, smoothing makes data easier to understand and more analytically useful. Smoothing techniques can be utilized to discover overall patterns and variances in cancer incidence or survival rates over time in the SEER dataset. This can yield useful insights for cancer research and public health decision-making.

C. Feature selection

In machine learning and statistics, feature selection is the process of identifying and choosing the most pertinent features (variables) from the original dataset in order to construct predictive models. By decreasing the dimensionality of the dataset and removing superfluous or unnecessary features that could cause noise or overfitting, feature selection aims to enhance model performance. There are several approaches to feature selection, such as wrapper, embedding, and filter approaches. Filter approaches choose features regardless of the selected model and assess the relevance of features using statistical measures like mutual information, correlation, or significance tests. Wrapper approaches analyze the performance of various feature subsets using a given evaluation metric, such as accuracy or cross-validation score, by applying a particular machine learning algorithm. While these techniques can be computationally costly, they frequently produce better feature subsets that are customized for the particular model. Feature selection is integrated into embedded techniques such that feature subset and model parameters are optimized at the same time during the model training process. Examples include decision tree-based algorithms like Random Forests, which naturally carry out feature selection during training, and regularization strategies like Lasso regression. In proposed method we are using cross validation $K=5$, training 90 and 10 for testing.

D. Model Selection

Model selection is the process of determining which machine learning algorithm or model architecture, based on performance and generalization capabilities, is best suited for a particular task. It entails testing several candidate models on training data using methods like holdout validation or cross-validation, then choosing the model that performs the best on a different validation or test dataset in terms of accuracy, precision, recall, or F1 score. When choosing a model, various criteria are taken into account, including the model's complexity, computational efficiency, interpretability, and the particular needs of the given situation. It is crucial for developing dependable and successful predictive models that, while preventing overfitting and guaranteeing robust performance on unknown data, correctly capture the underlying patterns in the data. In this paper 5 different models namely SVM, Kernel Naïve Bayes, Gaussian Naïve Bayes, Neural network and Ensemble Boosted Trees are used.

a) *Support Vector Machine (SVM)*: SVM is an effective supervised learning method that may be applied to regression and classification problems. It maximizes the margin between classes by identifying the ideal hyperplane in a high-dimensional feature space that best divides data points of various classes. Using kernel functions, SVMs can handle both linear and non-linear classification problems, which makes them adaptable to a wide range of datasets.

b) *Kernel Naive Bayes*: An improvement on the Naive Bayes classifier, Kernel Naive Bayes uses kernel techniques to manage non-linear decision boundaries. It maps data into higher-dimensional spaces, where linear separation may be attained, by applying the Bayes theorem under the presumption of feature independence. When paired with suitable kernel functions, Kernel Naive Bayes can outperform other algorithms in certain difficult classification tasks, even though it is simple and computationally efficient.

c) *Neural Network*: Artificial Neural Networks (ANN), commonly referred to as Neural Networks, are a class of deep learning algorithms that draw inspiration from the architecture and operations of the human brain. They

are made up of interconnected layers of artificial neurons that use weighted connections to process input data and then apply non-linear transformations to produce predictions as an output. Neural networks are a powerful tool for a variety of tasks, including image recognition, natural language processing, regression, classification, and identification of intricate patterns and correlations in massive datasets.

d) Gaussian Naive Bayes: A variation of the Naive Bayes classifier that makes the assumption that features have a Gaussian distribution is called Gaussian Naive Bayes. It functions by estimating the mean and variance of each feature for each class, which is especially useful for continuous-valued features. The probability of class membership for new instances is then calculated using the Bayes theorem. Gaussian Naive Bayes can achieve unexpectedly good results on many classification tasks, especially when the feature distributions near Gaussian distributions, despite its oversimplified assumption of feature independence.

e) Ensemble Boosted Trees: Multiple decision trees are combined to create a powerful prediction model using ensemble boosted trees, such as Gradient Boosting Machines (GBM) and XGBoost. Boosting involves training each tree on the residuals of the previous ensemble and then adding additional trees one after the other in order to fix the errors of the previous ones. A very reliable and accurate model that can successfully represent intricate interactions and nonlinear correlations in the data is produced by this iterative approach. Boosted trees are extensively employed across multiple fields for functions like classification, regression, and ranking. They frequently attain cutting-edge outcomes in predictive modelling contests and practical uses. In proposed method Ensemble Boosted Trees gives the highest accuracy of 90.4%

IV. EXPERIMENTAL RESULTS

Performance metrics are measurements that are used to assess how well a machine learning model predicts the future based on data that hasn't yet been observed. Accuracy, AUC, PPV, TPR, and CM are examples of common performance measures that change based on the kind of task, such as classification, regression, or clustering. These metrics help in model selection, tuning, and comparison by offering insightful information about the behavior of the model, its capacity to accurately categorize or predict outcomes, and its ability to generalize to new data.

a) TPV: The performance statistic known as True Positive Rate (TPR), which is often referred to as sensitivity or recall, quantifies the percentage of real positive cases in the dataset that the model correctly recognized as positive examples. The ratio of genuine positive forecasts to the total number of real positive occurrences is used to compute it. When it comes to situations like fraud detection or medical diagnostics, when accurately identifying positive instances is crucial, TPR is very helpful.

b) PPV: The performance indicator known as Positive Predictive Value (PPV), or precision, calculates the percentage of actual positive predictions among all positive predictions the model makes. The ratio of actual positive forecasts to all positive predictions is used to compute it. When the cost of false positives is substantial, PPV is helpful in determining the model's reliability for positive predictions.

c) AUC: A performance indicator called Area Under the ROC Curve (AUC) assesses how well a binary classification model can discriminate between classes across all potential classification thresholds. It computes the area under the curve by plotting the True Positive Rate (TPR) versus the False Positive Rate (FPR) at different threshold values. Higher values indicate greater performance. AUC provides a single scalar value that represents the overall discriminative ability of the model. It is frequently used to assess and compare the effectiveness of various classifiers, particularly when working with datasets that are unbalanced or when the trade-off between TPR and FPR needs to be taken into account.

d) Confusion Matrix: In machine learning, a confusion matrix is a tool used to assess a classification model's performance. By comparing predicted and true labels, a classification algorithm's performance is summarized in this matrix. Typically, it takes the form of a square matrix, with rows denoting the actual classes and columns the

anticipated classes. TP (True Positive): The model correctly predicts the positive class. TN (True Negative): The model correctly predicts the negative class. FP (False Positive): The model incorrectly predicts the positive class when it's actually negative (Type I error). FN (False Negative): The model incorrectly predicts the negative class when it's actually positive (Type II error).

1) Data Analysis

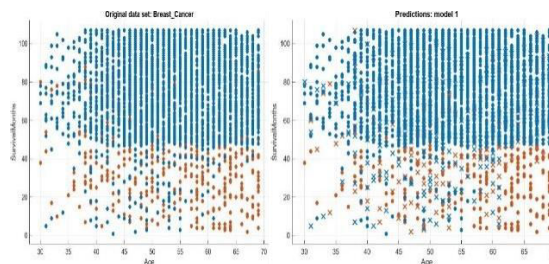


Figure 2: Age vs Survival months

Age vs. Survival Months: Analyzing the relationship between a patient's age at diagnosis and survival months can identify trends in how age affects how long a patient lives after receiving a cancer diagnosis. When additional variables like cancer stage and treatment modalities are taken into account, such as age, can be taken into account to determine whether younger or older patients typically have better survival outcomes. For the purpose of prognosis prediction and tailored treatment planning, it is essential to comprehend the correlation between age and survival months.

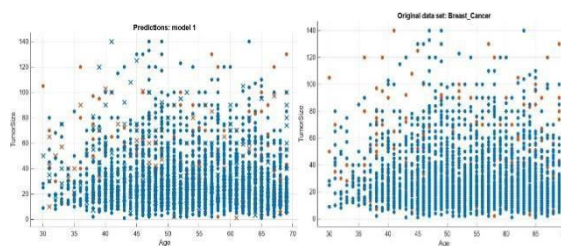


Figure 3: Age vs Tumor size

Age vs. Tumor Size: Determining whether age affects tumor growth rates and aggressiveness can be accomplished by looking into the relationship between age and tumor size. It can shed light on whether younger patients tend to have larger or smaller tumors, and whether age has an impact on the biology and progression of tumors. This comparison is necessary to identify the best course of action because tumor size frequently affects the choice of therapies, including radiation therapy, chemotherapy, and surgical resection.

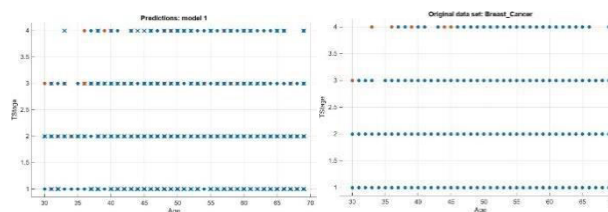


Figure 4: Age vs Tumor stage

Age vs. Tumor Stage: Determining whether age affects the chance of receiving an advanced cancer diagnosis requires examining the link between age and tumor stage. Age-related differences in cancer screening habits, healthcare access, and disease awareness can be found with the aid of this approach. Reducing disparities in cancer outcomes across age groups can be accomplished by developing tailored therapies and improving early detection initiatives by taking into account the impact of age on tumor stage.

2) Ensemble Boosted Trees Validation

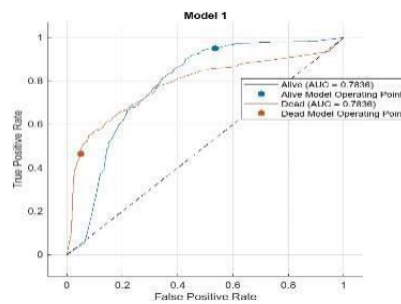


Figure 5: AUC validation

AUC is a crucial criterion for evaluating how well a model can generalize to new data. Model performance and robustness are objectively estimated by assessing AUC on validation datasets that were not utilized for model training. A high AUC value on the validation set indicates that the model can generalize to new examples with effectiveness and has picked up useful patterns from the training set. On the other hand, overfitting or a lack of generalization may be indicated by a noticeably lower AUC on the validation set in comparison to the training set

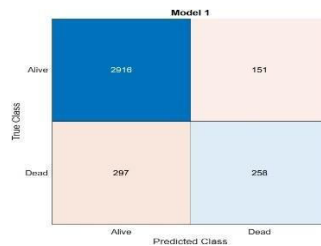


Figure 6: Confusion matrix validation

As a diagnostic tool, the confusion matrix evaluates how the model behaves with unknown data. It offers insights into possible areas of improvement, such as identifying particular types of errors (e.g., false positives or false negatives) and comprehending the model's strengths and weaknesses across different classes, by comparing the model's predictions with the true class labels in the validation set.

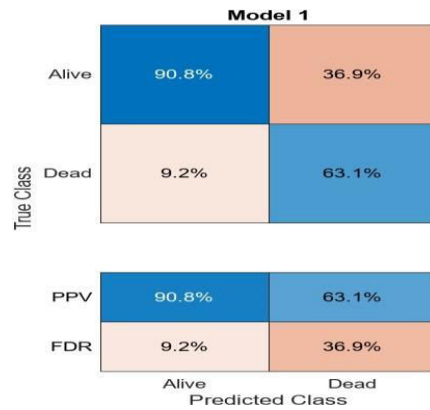


Figure 7: PPV validation

A crucial criterion for evaluating the model's performance on hypothetical data is PPV. Through the use of validation datasets not utilized for model training, PPV offers an objective assessment of the model's accuracy and dependability in practical situations. Confidence in the model's deployment is increased by a high PPV on the validation set, which indicates that the model retains its predictive accuracy and generalizes well to new instances.

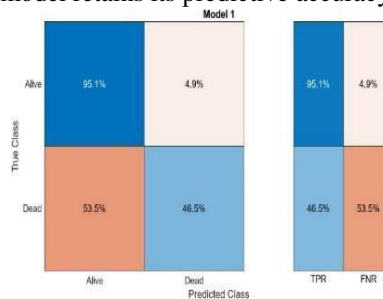


Figure 8: TPR validation

TPR is a crucial metric for assessing how well the model performs with unknown data. To provide an impartial measure of the model's sensitivity and generalizability to new cases, TPR is evaluated on validation datasets that were not utilized during model training. A high TPR on the validation set indicates the model's robustness and capacity for generalization, indicating that it can successfully find positive examples in fresh data.

3) Ensemble Boosted Trees Testing

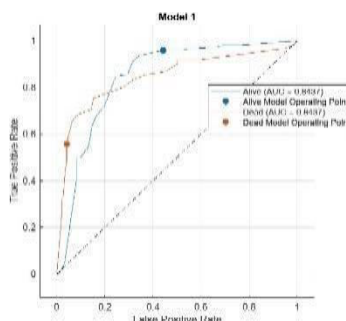


Figure 9: AUC Testing

Plotting the True Positive Rate (TPR) versus the False Positive Rate (FPR) at different threshold values yields the AUC. This curve, also referred to as the Receiver Operating Characteristic (ROC) curve, shows how sensitivity—the capacity to recognize positive instances—and specificity—the capacity to recognize negative instances—are traded off. The area under this ROC curve is represented by the AUC value; a greater AUC indicates that the model performs better overall in differentiating between the two classes.

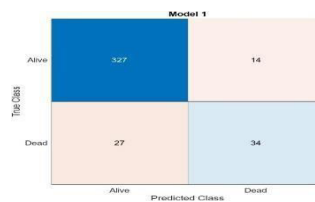


Figure 10: Confusion matrix Testing

The model's predictions for various classes are compiled into a confusion matrix, which is usually divided into four categories: True Positives (TP), False Positives (FP), True Negatives (TN), and False Negatives (FN). The symbols TP, FP, TN, and FN stand for cases that are correctly identified as positive, mistakenly classified as positive, correctly classified as negative, and incorrectly classified as negative, respectively.

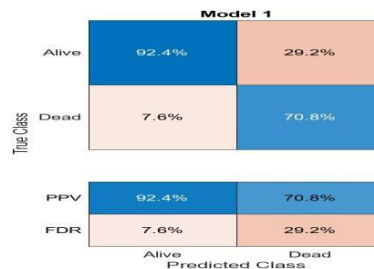


Figure 11: PPV testing

The ratio of actual positive predictions to all positive predictions is used to compute PPV. It assesses how well the model detects positive cases while reducing the amount of false positives. A high predictive power value (PPV) suggests that the model's positive predictions are reliable and likely to be accurate, which makes it appropriate for applications where accuracy is critical, like fraud detection or medical diagnosis.

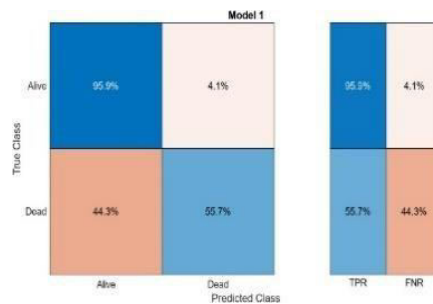


Figure 12: TPR testing

The ratio of true positive forecasts to the total number of real positive occurrences is known as the true positive rate, or TPR. It shows the percentage of positive cases that the model accurately detects. A high TPR suggests that the model maximizes the detection of genuine positives while decreasing false negatives and successfully capturing positive cases.

S.no	Models	Accuracy(in%)
1.	SVM	87.3
2.	Kernel Naïve Bayes	85.2
3.	Neural Network	84.3
4.	Gaussian Naïve Bayes	82.7

5.	Ensemble Boosted Trees	90.4
----	------------------------	------

V. CONCLUSION

The SEER dataset and machine learning methods hold great promise for breast cancer early diagnosis. The accuracy achieved by using ensemble boosted trees models is amazing, with an impressive 90.4%. This represents a major breakthrough in the early detection of breast cancer and presents an opportunity for prompt treatment and better patient outcomes. When cutting-edge computational methods are combined with extensive cancer registry data, strong predictive models that can recognize minute patterns suggestive of early-stage breast cancer can be created. These remarkably high accuracy rates highlight how machine learning has the ability to completely transform early detection techniques, which will eventually increase survival rates and lower mortality.

REFERENCES

- [1]. Sonam Maurya, Sushil Tiwari, Monika Chowdary Mothukuri, Chandra Mallika Tangeda, Rohitha Naga Sri Nandigam, Durga Chandana Addagiri. "A review on recent developments in cancer detection using Machine Learning and Deep Learning models." Biomedical Signal Processing and Control, 2023.
- [2]. BREASTCANCER.ORG. "Male breast cancer." https://www.breastcancer.org/symptoms/types/male_bc.
- [3]. American Cancer Society. "Key statistics for breast cancer in men." <https://www.cancer.org/cancer/breastcancerinmen/about.html>. (Accessed: 13-02-2022).
- W.H. Organization. "Breast cancer: Introduction and treatment." 2021.
- [5]. S. Mojarad, S. Dlay, W. Woo, G. Sherbet. "Breast cancer prediction and cross validation using multilayer perceptron neural networks." In: 2010 7th International Symposium on Communication Systems, Networks Digital Signal Processing (CSNDSP 2010), pp. 760–764, 2010. <http://dx.doi.org/10.1109/CSNDSP16145.2010.5580318>.
- [6]. S. Guan, M. Loew. "Breast cancer detection using transfer learning in convolutional neural networks." In: 2017 IEEE Applied Imagery Pattern Recognition Workshop, AIPR, pp. 18,017.
- [7]. S. Khan, N. Islam, Z. Jan, I. Ud Din, J.J.P.C. Rodrigues. "A novel deep learning based framework for the detection and classification of breast cancer using transfer learning." Pattern Recognition Letters, 125, pp. 1–6, 2019.
- [8]. M. Kashif, K.R. Malik, S. Jabbar, J. Chaudhry. "Application of machine learning and image processing for detection of breast cancer." 2020, pp. 145162. <http://dx.doi.org/10.1016/B978-0-12-819043-2.00006-X>.
- [9]. Sharifian A, Pourhoseingholi MA, Emadedin M, et al. "Burden of Breast Cancer in Iranian Women is Increasing." Asian Pac J Cancer Prev, 16(12):5049-52, 2015.
- [10].Tüba Kiyand Tülay Yildirim. "Breast cancer diagnosis using statistical neural networks." Journal of Electrical & Electronics Engineering, vol. 4, pp. 1149-1153, 2004.
- [11].Dheiver Santos, "Breast Cancer Survival Prediction using Machine Learning and Gene

ExpressionProfiles”2022 <https://doi.org/10.1101/2022.01.22.22269470>

[12].Alvaro Zorrilla-Martinez, Olga Diaz-Rodriguez, Javier Pereira, Carmen Cadarso-Suarez, "Evaluation of Machine Learning Algorithms for the Prognosis of Breast Cancer from the Surveillance, Epidemiology, and End Results Database" Journal of Medical Systems, 2020

BIOGRAPHY



Dr. S. Xavier Arokiaraj,
Associate Professor,
Electronics and Communication Engineering
Department,
Adhiyamaan college of Engineering,
Hosur



Abarna Lakshmi. R,
Electronics and Communication
Engineering Department,
Adhiyamaan college of
Engineering,
Hosur



Elfalin Shruthi.B,
Electronics and Communication
Engineering Department,
Adhiyamaan college of
Engineering,
Hosur



Jencisusan.A,
Electronics and Communication
Engineering Department,
Adhiyamaan college of
Engineering,
Hosur



Uma Devi.S,
Electronics and Communication
Engineering Department,
Adhiyamaan college of
Engineering,
Hosur



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details