

Original Article

Feature Based Ensemble Learning Model for Breast Cancer Reoccurrence Retrieval

Mohan Kumar¹, Sunil Kumar Khatri², Masoud Mohammadian³

¹ Amity Institute of Information Technology, Amity University, Noida (UP), India

² Director of Campus, Amity University Tashkent, Tashkent City, Uzbekistan

³ Faculty of Science and Technology, University of Canberra, Canberra. ACT. Australia

¹Corresponding Author : mohan.kumar@student.amity.edu

Received: 20 June 2022

Revised: 12 August 2022

Accepted: 13 September 2022

Published: 20 October 2022

Abstract - In recent years, 30% of women have been diagnosed with cancer yearly. Improvement of medical treatments made patients would be in remission but with challenges. It's estimated that there will be 14.8% of new cases in the last year. Breast cancer reoccurrence regenerates new challenges causing severe effects and even causing life loss. So, if it is detected early, it can cure it. Various latest techniques like machine learning are very much required for solving and predicting the reoccurrence for reducing mortality to some extent. The research paper proposed the ensemble approach, which used the Voting method for combining techniques used ensemble methods for detecting the two classes of tumors benign and malignant. Ensemble used Meta to implement more than one classifier. Experiments conducted on Wisconsin Diagnostic Breast Cancer (WDBC) dataset and voting techniques are used to get better results for model evaluation. Logistics regression, support vector machine, RBFF and Linear classifier, decision tree classifier and Random Forest are used to get classification accuracy, precision, recall, and F1 measure. The results obtained show that the ensemble models showed significant achievement in terms of performance and 98 % in terms of accuracy. With AdaBoost and cost-sensitivity in a model, a reasonable accuracy has been achieved. The proposed model in this research supports setting and evaluating various follow-up visit interventions and very advanced treatment recommendations, so there should be very low cancer mortality.

Keywords – Machine learning, Classification, AdaBoost, Ensemble learning.

1. Introduction

The most frequent breast cancer is invasive malignancies. Cancer in the breast is very common among women, according to a WHO report released in 2018 [1]. Breast cancer kills an estimated 2.1 million women annually, making it the most common cancer among women. There is a claim by clinical experts that breast cancer claimed the lives of around 627,000 women in 2018 (or 15% of all female cancer fatalities) [1]. The incidence of invasive breast cancer in women increased by only 0.3% per year between 2007 and 2016. The statistic of the American Cancer Society, breast cancer in females and casualty of death rates reported much in 1989 as thirty-two percent. 100,000 people then fell by forty percent to nineteen percent 100,000 cancers in female people in 2017, representing a significant mortality rate [2]. Those who participate in routine screenings every 1–2 years can lower their death from breast cancer by forty percent [5, 6].

Breast cancer patients, on the other hand, would be in remission because of earlier diagnosis and better therapy. 90% survival is typically 5 years for breast cancer, and 10 years is the average survival rate among 83% of women. It has been seen in most patients that breast cancer recurrence is very common, even though early detection and better

medical treatment can put the disease into remission. Breast cancer recurrence is the leading cause of breast cancer-related death [8] and a major clinical symptom. Breast cancer recurrence has recently become a major focus for researchers [9]. For example, ER, PR, HER2, and TNBCs can determine whether breast cancer subtypes have a higher recurrence risk than others in various years or situations [10–12]. According to research, breast cancer recurrence is associated with lymph node metastasis in the upper arm. By treating the metastases as soon as they appear, the risk of a recurrence of breast cancer can be lowered. On the other hand, these patterns necessitate a significant investment of time and money [13]. Therefore, there is a need to present a very non-invasive model supporting computations for predicting the probability of repetitive occurrence of cancer in the breast in women who have already been diagnosed with this disease.

In the same way that [14,15] did, the patients' data and treating medical information from the breast cancer registry to develop a prediction model and test alternative techniques to accomplish the research objective. Compared to the patterns discussed previously, the proposed model can be applied in a clinical context following the treatment of primary breast cancer in a low-cost and time-saving manner. Adjuvant treatment advancements and earlier cancer



identification have helped lower mortality and recurrence rates [16]. In cases where recurrence is due to micrometastatic illness, adjuvant therapy may be appropriate. Recurrence risk is determined by examining a variety of criteria related to the patient and the tumor, including the patient's age at diagnosis, tumor size, axillary ganglia condition, differentiation stage, and the presence or absence of vascular or lymphatic invasion [17].

2. Literature Review

The active participation of physicians in research investigations is not viable. As a result, a computer program is required to convert unstructured data to a structured format [28]. Natural Language Processing (NLP) algorithms use a set of theories and computational approaches to analyze, analyze, and interpret documents written in natural language [16]. On EHR data, NLP has previously been used in many pieces of research. For example, an automated technique for extracting medical concerns from EHR has been developed [17]. Hepatocellular carcinoma was identified from EHR data using a hybrid technique combining natural language processing (NLP) and (ICD-9) codes [18]. As a result, NLP looks to be a viable option for extracting info from unstructured and diverse medical data, such as medical data and radiology reports [16].

However, with other types of cancer, the disease can progress quickly from one stage to the next, making treatment more difficult [23]. As a result, it's critical to find the best treatment strategy as soon as possible. Machine learning may be able to produce the required result.

Topography [28], energy management [26, 42], text document classification [2, 4, 3], and preventative maintenance [24] are only a few of the disciplines where machine learning algorithms have been used. The outcomes of these machine learning applications assist us in making the best decision possible.

Breast cancer recurrence has been predicted and predicted using ML approaches in numerous studies. A decision tree (DT), artificial neural networks (ANNs), and a support vector machine (SVM) were all utilized by the authors in [5]. SVM was the most accurate and least error-prone method for assessing breast cancer. Random forest, SVM, logistic regression, and the Bayesian classification algorithm were used by the authors of [38] to create a prediction model based on three biomarkers. Slightly earlier, the software could predict the return of the disease in patients [33]. Researchers in [12] used SVM and ANN to develop a strategy for predicting breast cancer recurrence. ANN was able to predict the recurrence of the disease accurately. Numerous risk factors and predictors are discussed in the literature for breast cancer recurrence. For example, the XGBoost ensemble learning algorithm was used by authors in [14] to predict breast cancer recurrence using 23 predictors. Doctors can employ machine learning algorithms to predict breast cancer recurrence. Some risk factors include cancer therapy [18], molecular data and vulnerable genes [41], BMI [20], and molecular subtype [35,36]. Although some studies have investigated recurrence risk factors and predictors, more research is required. For this reason, no single machine learning technique is suitable for all data [19].

Table 1. Exploratory analysis of studies

Author	Study Description	Technique
Wang H. et al. [2017]	A support vector machine (SVM)-based ensemble learning system for breast cancer diagnosis is investigated in this study.	This research focuses on breast cancer diagnosis using an SVM-based ensemble learning method to reduce the variation in diagnosis and boost the accuracy of diagnosis. Hybridization of twelve SVMs Ensemble technique is done.
Farhad K. et al. [2021]	The most challenging part is detecting cancer and distinguishing between a diagnostic to determine if a patient has a malignant or benign condition	The K-Nearest Neighbors classifier, for example, is a machine learning technique that helps tackle this challenge by giving excellent accuracy performance. One of the machine learning methods used to improve the diagnosis accuracy of mammograms is K-Nearest Neighbors. This study examines several recent studies demonstrating the accuracy of K-Nearest Neighbors as a machine learning method for diagnosing breast cancer.
Kaustubh C. et al. [2019]	Machine learning techniques are used to detect and forecast breast cancer recurrence, and various metrics, like accuracy, precision, etc., will be used to compare all the models. Breast cancer recurrence can be accurately predicted using the models developed.	Wisconsin Prognostic Breast Cancer Dataset is used to create all models (WPBC). Multiple Linear Regression, Support Vector Machine and Decision Tree were all created utilizing metrics such as Gini Index, Entropy, and Information Gain, respectively. Regarding predicting recurrence and non-recurrence, Support Vector Machine and K-fold Cross-Validation performed the best.

Gupta M. et al. [2018]	Based on breast cancer Wisconsin data, the research offered an overview of machine learning approaches in cancer disease evolution.	The results suggest that the Multilayer Perceptron is superior to other approaches in terms of performance by applying learning algorithms such as Linear regression, MP and Random Forest to the data (DT).
Bustan M. et al. [2021]	This study aims to establish the correlation between pathology diagnosis and the likelihood of breast cancer spreading to other organs. Tumor location, topography, behavior, grade, and status are all part of the pathology diagnosis process. and grade.	The researchers used binary regression analysis modelling to examine if there was a link between the pathological diagnostic of breast cancer (laterality, topography, behavior, grade, and tumor status) and the probability of metastasis to other organs. The study's findings revealed a link between the tumor's location and status, as well as the occurrence of metastasis.
Gupta N. et al. [2021]	This paper aims to analyse different classifiers on the Breast Cancer dataset and suggest a new Machine Learning Classification ensemble training approach.	An Ensemble Learning model for Prediction is presented to classify the outcomes among several classifiers. Finally, the Voting Ensemble determines the best classifier for breast cancer prediction.
Abdullah D. et al. [2018]	This study is about the analysis of the recurrence of breast cancer.	An ensemble-based approach is used. This study aims to use an ensemble learning technique to enhance breast cancer recurrence prediction and create a website that allows clinicians to enter features about a breast cancer patient and acquire the probability of recurrence.
Sultana J. et al. [2019]	To aid and oversee patients, the primary identification and prediction of cancer types should become a compulsion in cancer research.	The Logistic Regression technique and multi-classifiers have been developed to predict breast cancer. Make deep predictions on breast cancer data in a new setting. This research investigates the many data mining methodologies used to construct deep predictions utilizing the classification of Breast Cancer data.
Iqbal A. et al. [2016]	This research aims to use various data mining approaches to determine the likelihood of breast cancer recurrence to improve the accuracy of such models.	The authors calculated the accuracy of the Naive Bayes, C4.5 Decision Tree, and Support Vector Machine (SVM) classification algorithms. By removing some lower-ranked attributes, they could increase each model's accuracy using an efficient feature selection approach.
Shibahara T. et al. [2021]	Deep learning-based model for prediction of breast cancer.	The authors created an explainable deep learning model termed a point-wise linear model, which leverages meta-learning to construct a custom-made logistic regression for each sample to expose the mechanisms inherent in the PAM50 subtypes.

3. Methodology

3.1 Design of study and Data Collection

It is a quantitative data set from a real-world multivariate Wisconsin breast cancer database (WBCD) made public in 1995. This data has been used in many research experiments in life science and healthcare domains [42,43]. Most of the publications use traditional machine learning algorithms, but in this work, we are comparing

those traditional machine learning algorithms' performance with the ensemble machine learning model [39,40]. The data set used for this work has 32 attributes and 569 instances of adult female participants with breast cancer. Out of these 32 attributes, one attribute diagnosis has the name of the class, and there are only two classes one is malignant, and another one is benign. Quantitative features are taken from images from the Breast Cancer Dataset of

569 patients on these 32 attributes. Out of these 32 attributes, one attribute with a name class is used as the dependent variable, while the rest of the variables are used as independent variables. The data set has been categorized into two classes malignant and benign. It has been observed that 357 instances are of the malignant type and the rest 212 are benign, as shown in Fig 1. The complete methodology adopted for research work is shown in Fig 2.

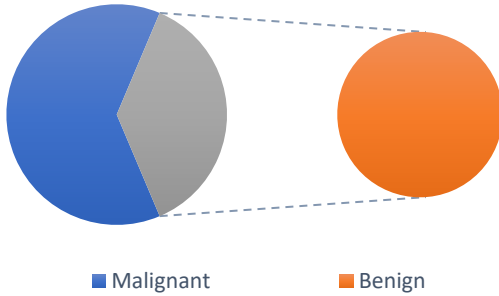


Fig. 1 Malignant and Benign Cancer

The data set has variations to bring out strong patterns in a dataset among 32 attributes and consider the most important feature using Principal Component analysis. Data exploration is performed to understand the dataset and take forward the most important data characteristics [41]. The most important features are shown in Table 2.

3.2. Feature Selection and Training Data

The Machine Learning model is built around on most important Features after Data exploration. The proposed research work comprised ten features selected as feature variables and the class as the dependent variable. This data is split into different ratios for evaluating the model. Data splitting is performed at 70:30, 60:40 and 50:50[25].

3.3. Multi-class Logistic Regression

A supervised machine learning algorithm, along with statistics, is used in solving multiple class and binary class classification problems. For Binary classification, the range of dependent variables may lie outside $[0, 1]$.

$$y \in [0, 1]$$

$$X \in R$$

Let Probability($y=1|X$)= $p(X)$,

Sigmoid Function: $p(X) = 1/(1+e^{-\beta x})$

The logistic regression uses maximum likelihood for parameter estimation.

For example, Consider N samples with labels either 0 or 1. For Samples labelled “1”: Estimate β^{\wedge} such that $p(X)^{\wedge}$ is as close to 1 as possible.

For Samples labelled “0”: Estimate β^{\wedge} such that $1-p(X)^{\wedge}$ is as close to 1 as possible.

3.4. Decision Tree

A supervised learning method can be utilized as a decision tree. It can be used for both characterizing and issues related to regression. It’s a classifier based on tree data structure, where inside hubs address the highlights of a dataset, branches address the Decision guidelines, and each leaf hub addresses the result [46]. A decision tree has two nodes: the decision node and the leaf Node. Decision nodes are very much utilised for settling any decision having various branches, while Leaf nodes are the result of various decisions and do not contain any branches. There are various highlights on the decisions and testing on a performance basis on highlighting the given dataset. The same has been applied to the breast cancer dataset for finding leaf hubs to determine decisions about retrieving cancer reoccurrence [27].

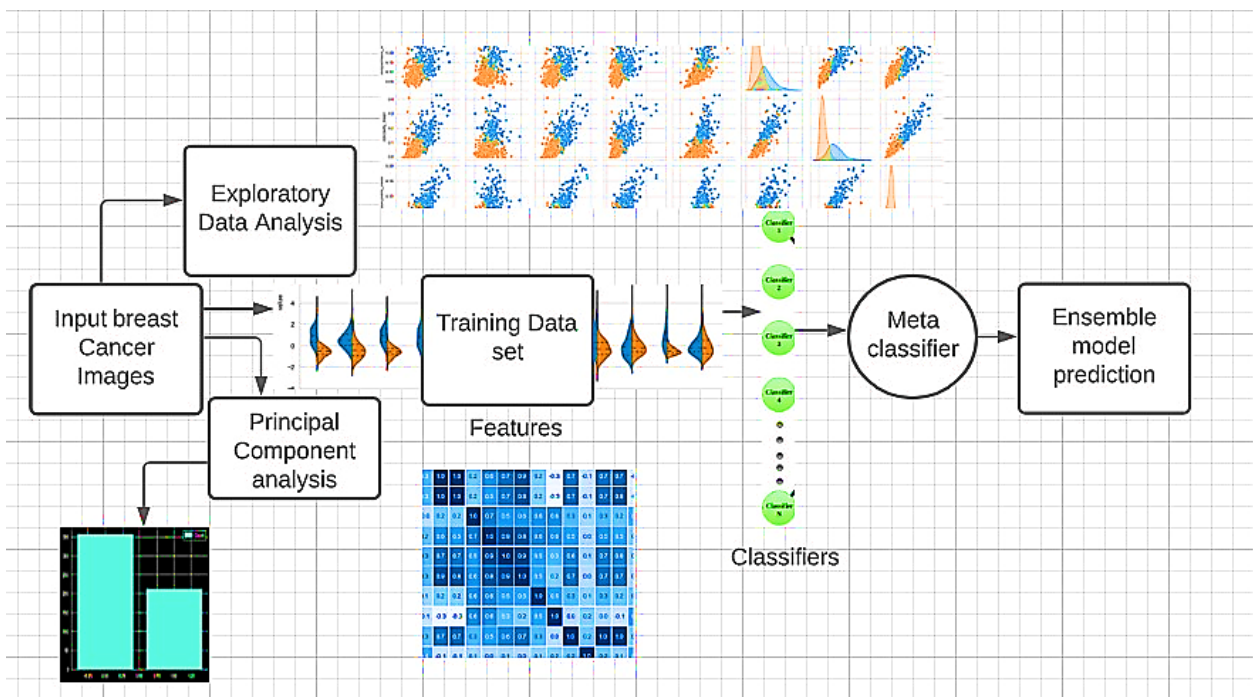


Fig. 2 Methodology

3.5. Support Vector Machine (SVM)

The various grouping and relapsing difficulties Support Vector Machine is an AI calculation that can be utilized for both grouping and relapse difficulties. In any case, it is generally utilized in arrangement issues. In the SVM calculation, a plot for every piece of information is taken as a point in n-layered space (n is highlighted), with each element being the worth of a specific direction. Support Vector Machine is applied to breast cancer datasets for binary linear classification to have the boundary of decisions, and it's very helpful for building a generalization error. The model performed well in finding outliers by observing the hyper-plane that separates the two classes of Malignant and Benign Cancer [26,29].

3.6. Voting Ensemble Model

A Voting group works by joining the forecasts from numerous models. It tends to be utilized for grouping or relapse. We have used this work's three most popular supervised machine learning models to vote for predictions [32]. Those three machine learning models are Logistic Regression, Decision Trees, and Support Vector Machine. We have observed that when using these standalone algorithms, the Logistic Regression and Support Vector Machine algorithms show around 98.4% and 98.7% accuracy, respectively. In comparison, the Voting Ensemble machine learning model shows an accuracy of 99%.

4. Results and Discussion

This analysis of Ensemble learning methods with breast cancer survival prediction on WBCD (the Wisconsin Breast Cancer Database) as multivariate data. In this feature, data are computed through breast mass. Data has been classified into two classes as 37.3% malignant and 62.7% Benign, as shown in Fig 3.

Data exploration is performed to understand data properly. Data is visually explored below the most important features. These features are further taken as training Data set. Results of Data exploration are shown in Fig 4. The most important features used for Training are eleven, as shown in Table 2 and Fig 4. A voting algorithm in Ensemble learning is implemented for performance evaluation of machine learning algorithms.

In Machine Learning, data exploration has been used to transform existing variables and create new ones. Transformation of variables in the data set can help in scaling 32 variables for better visualization, while the creation of variable help in highlighting new relationships between those 32 variables. The data is further investigated for illuminating specific patterns and characteristics for better understanding and insights about the reoccurrence of cancer possibility so that the same can be loaded to ensemble methods [31]. Combined Data exploration has been generated for the Benign and Malignant breast cancer classes from the data set, as shown in Fig 5.

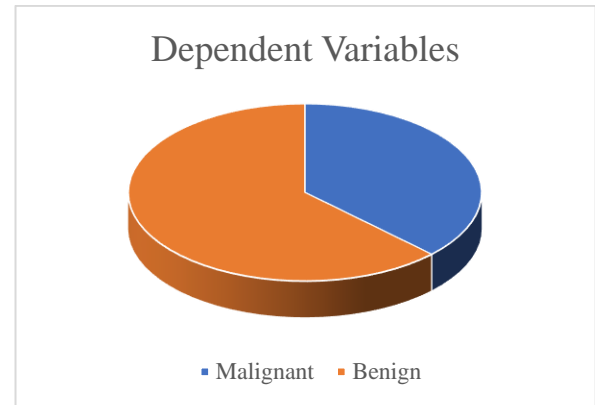


Fig. 3 Dependent variables of the Data set

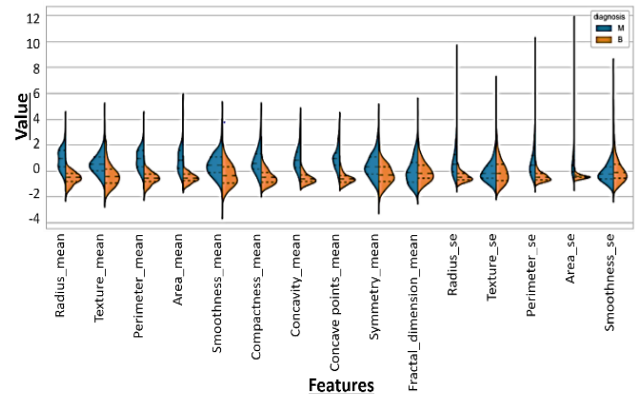
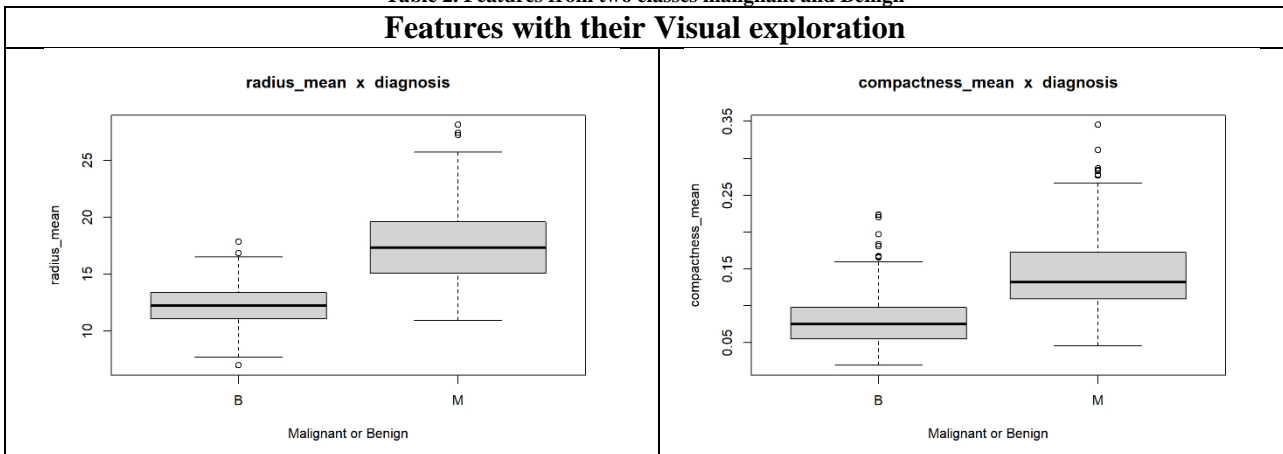
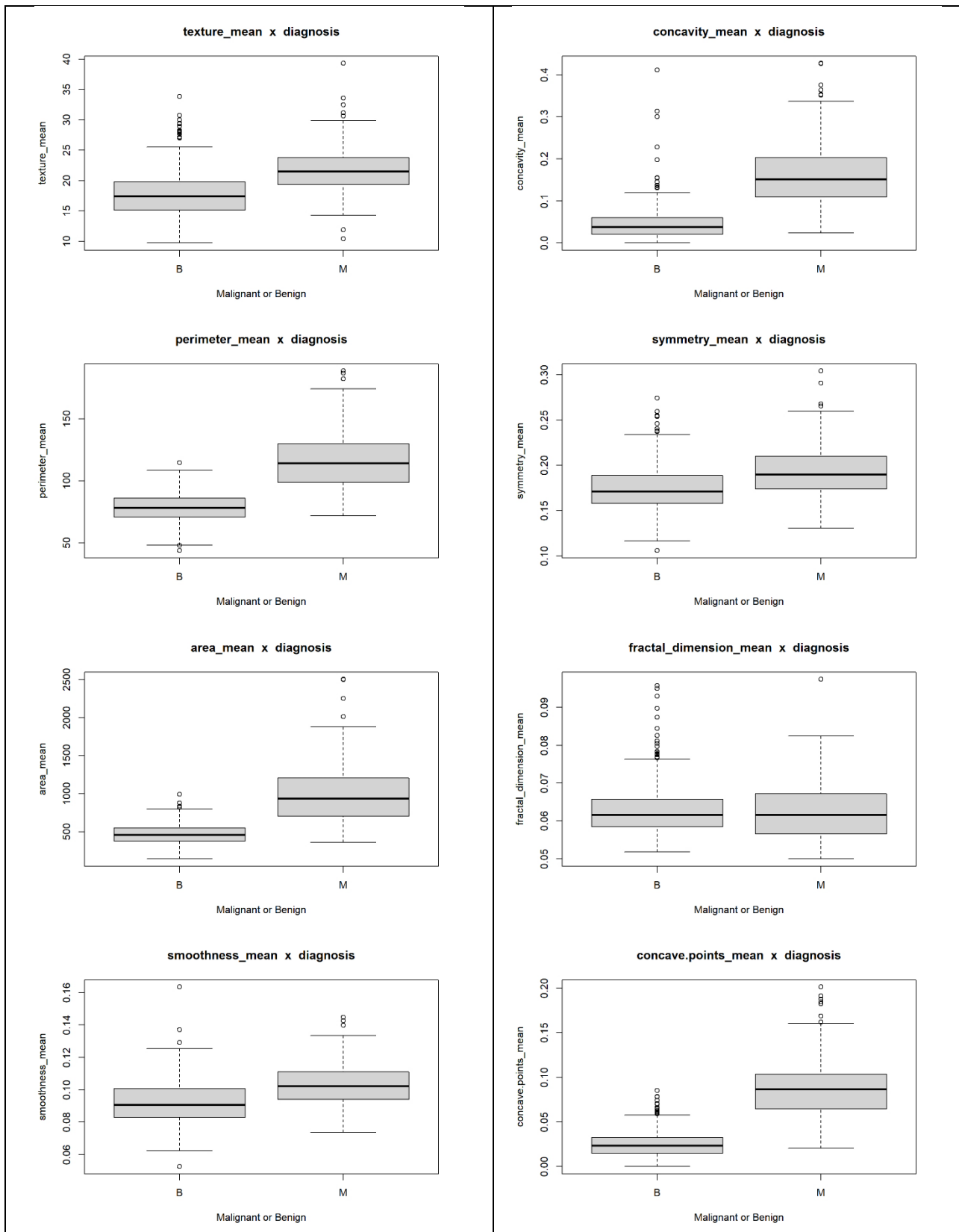


Fig. 4 Features used for Training

Table 2. Features from two classes malignant and Benign

Features with their Visual exploration





Feature Selection is required to implement machine learning. Variables are selected to act as an input variable. The output is then predicted. A correlation matrix has been utilised to display the correlation coefficients for 32

different variables [30]. The correlation between all possible features [airs is obtained using a Heat map. The patterns are visualized easily from the Breast cancer data set about the most useful features.

The linear relationship between feature variables from quantitative data shows a statistical measure of dependency between these variables. In this proposed work, the correlation has been shown with the help of a heatmap in the form of a correlation matrix representing the correlation between different variables from the breast cancer dataset, as shown in Fig 6.

Data set has classes as dependent variable like ID number, Diagnosis (M = malignant, B = benign). So, diagnosis has been taken as dependent variable has been utilized on Ten real-valued features are computed for each cell nucleus. The results from the same are shown in Fig 7.

4.1 Configurations of the Proposed Models

Ensemble Model is being implemented after taking training data set. Machine Learning algorithms like Logistic regression, Support vector machine as Linear classifier, Support vector machine as RBF classifier, Decision tree and

Random forest methods are been implemented as meta classifiers for proposed ensemble model. The results obtained are shown in Table 3 and Fig 8.

Accuracy of all models has been retrieved in terms of precision, recall f1-score and *Model 0* support as shown in below tables in terms of classes Malignant (M) Benign(B).

Table 3. Training accuracy and Testing accuracy

Models	Training accuracy	Testing accuracy
Multiclass Logistic regression	0.98	0.98
Support vector machine (linear classifier)	0.98	0.96
Support vector machine (RBF classifier)	0.98	0.98
Decision tree	1	0.94
Random forest	0.99	0.95

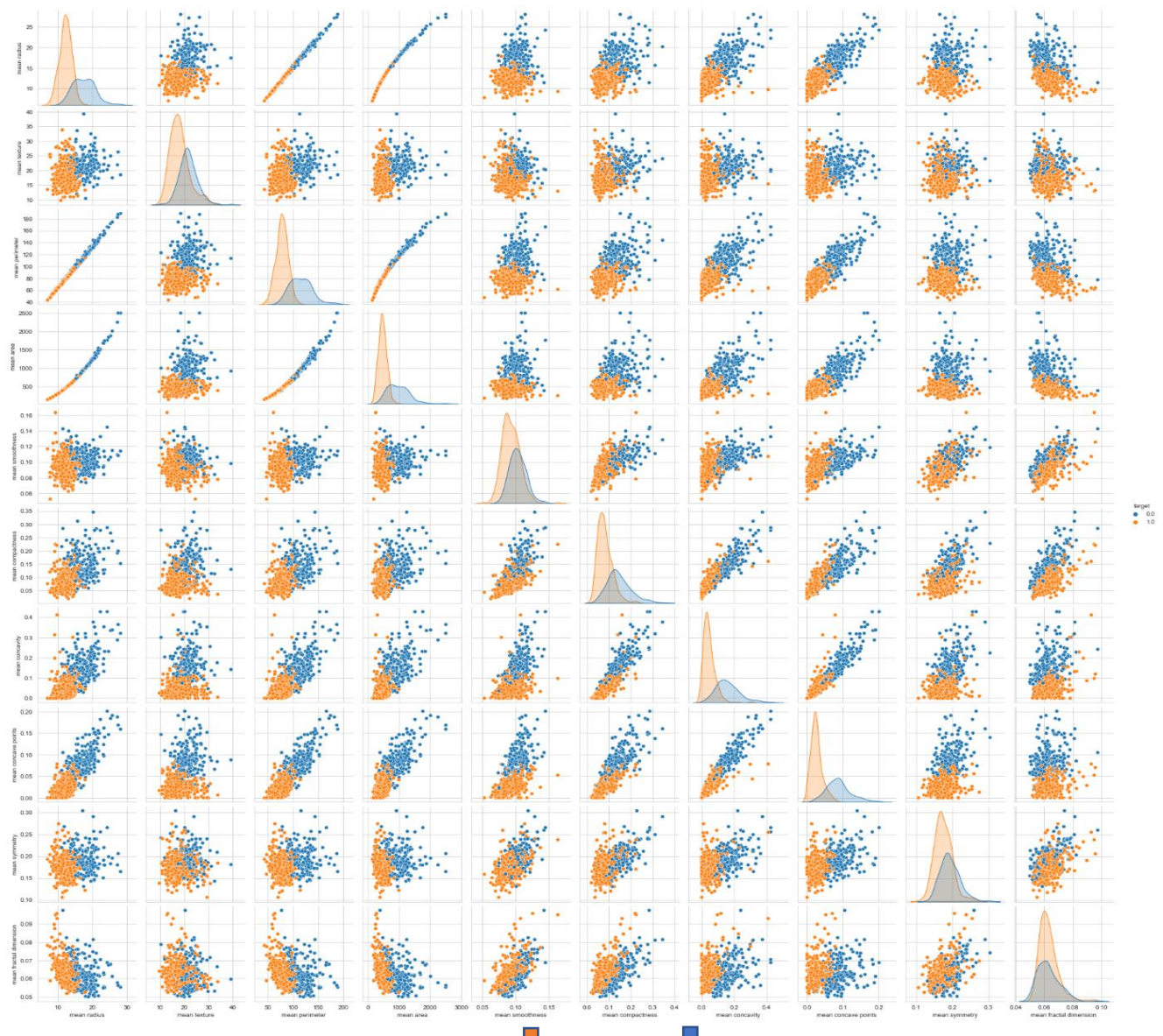


Fig. 5 Data exploration of Benign and Malignant cancer

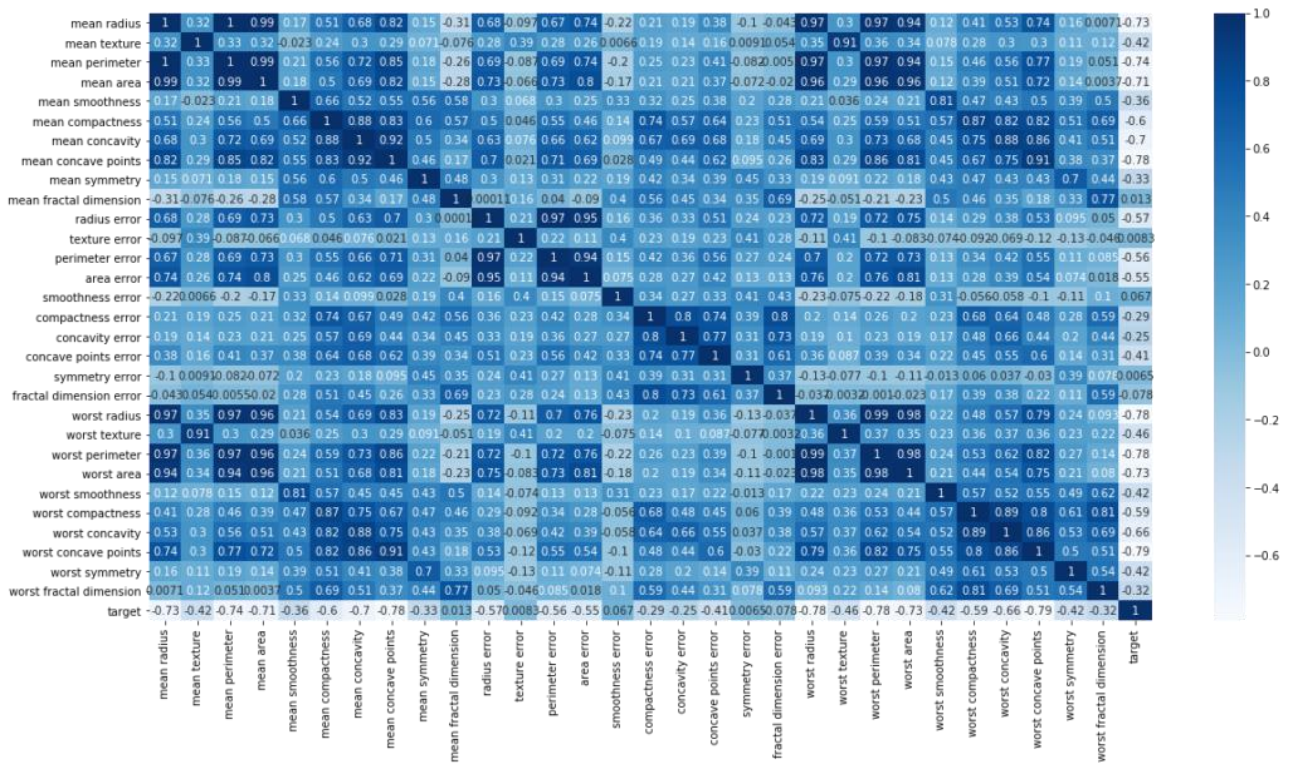


Fig. 6 correlation heatmap for correlation matrix between variables

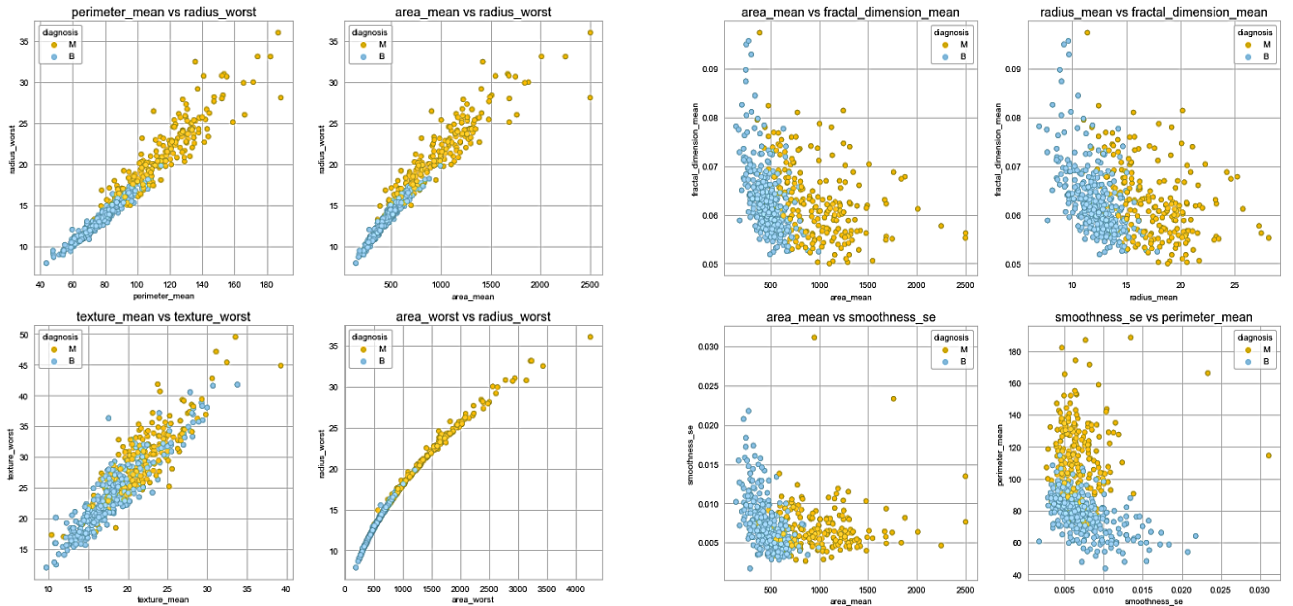


Fig. 7 Data exploration on dependent variable Diagnosis for both variables

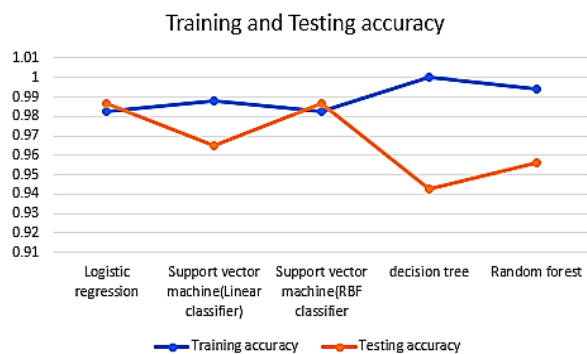


Fig. 8 Training and Testing accuracy

Model 1				
Model 1	Precision	Recall	f1-score	Support
B	0.99	0.99	0.99	143
M	0.99	0.98	0.98	85
Accuracy			0.99	228
Macro avg	0.99	0.98	0.99	228
Weighted avg	0.99	0.98	0.99	228

Model 2

Model 2	Precision	Recall	f1-score	Support
B	0.99	0.99	0.99	143
M	0.99	0.98	0.98	85
Accuracy			0.99	228
Macro avg	0.99	0.98	0.99	228
Weighted avg	0.99	0.98	0.99	228

Model 3

Model 3	Precision	Recall	f1-score	Support
B	0.97	0.97	0.97	143
M	0.94	0.94	0.94	85
Accuracy			0.96	228
Macro avg	0.95	0.95	0.95	228
Weighted avg	0.96	0.96	0.96	228

Model 4

Model 4	Precision	Recall	f1-score	Support
B	0.94	0.97	0.96	143
M	0.95	0.89	0.92	85
Accuracy			0.94	228
Macro avg	0.94	0.93	0.94	228
Weighted avg	0.94	0.94	0.94	228

5. Conclusion

This paper proposes an ensemble approach to build a computational model for predicting the risk of breast cancer occurrence on structured data. The prediction model is helpful in various clinical applications of early diagnosis of cancer symptoms to assess and predict the risk of recurrence after the proper treatment has been provided to patients detected with original breast cancer. There is a need for the early prediction that can help with early diagnosis and prevention of cancer recurrence.

Based on the proposed ensemble model, a clinical expert can take the prediction results for deciding treatment methods. Ensembles provide extra support for much and fast better decision-making. We constructed five models. Model 1, Model, Model 3, Model 4, and Model 5. All the models shown Logistic regression, Support vector machine and decision tree showed more than 98% training accuracy. The model will be helpful for health experts to utilize for further clinical decisions.

References

- [1] World Health Organization., "WHO Position Paper on Mammography Screening [Internet]? Switzerland: World Health Organization, " Available From: <https://apps.who.int/iris/handle/10665/137339>, 2014.
- [2] American Cancer Society, Cancer Facts & Figures 2020 [Internet]. Atlanta: American Cancer Society, 2020. Available From: <https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/annual-cancer-facts-and-figures/2020/cancer-facts-and-figures-2020.pdf>
- [3] Kim J, Shin H, "Breast Cancer Survivability Prediction Using Labeled, Unlabeled, and Pseudo-Labeled Patient Data," *Journal of the American Medical Informatics Association*, vol.20, no.4, pp.613–8. Doi: 10.1136/amiajnl-2012-001570. Pubmed PMID: 23467471; Pubmed Central PMCID: PMC3721173.
- [4] Hsu JL, Hung PC, Lin HY, Hsieh CH, "Applying Under-Sampling Techniques and Cost-Sensitive Learning Methods on Risk Assessment of Breast Cancer," *Journal of Medical Systems*, vol.39, no.4, pp.1–3, 2015. Doi: 10.1007/S10916-015-0210-X. Pubmed PMID: 25712814.
- [5] Seely JM, Alhassan T, "Screening for Breast Cancer In 2018-What Should We Be Doing Today?" *Current Oncology*, pp. S115–24, 2018. Doi: 10.374/Co.25.3770. Pubmed PMID:29910654; Pubmed Central PMCID: PMC6001765.
- [6] Uhry Z, Hédélec G, Colonna M, Asselain B, Arveux P, Rogel A, Et Al, "Multi-State Markov Models In Cancer Screening Evaluation: A Brief Review and Case Study," *Statistical Methods in Medical Research*, vol.19, no.5, pp.463–86, 2010. Doi: 10.1177/0962280209359848. Pubmed PMID:20231370.
- [7] Immadi Murali Krishna, Pendem Durga Bhavani, Tiriveedhi M S Madhuvani, Vajja Poojitha, "An Effective Segmentation and Modified Ada Boost CNN Based Classification Model for Fabric Fault Detection System," *SSRG International Journal of Computer Science and Engineering*, vol. 7, no. 7, pp. 34-40, 2020. Crossref, <https://doi.org/10.14445/23488387/IJCSE-V7I7P106>.
- [8] Moody SE, Perez D, Pan TC, Sarkisian CJ, Portocarrero CP, Sterner CJ, Et Al, " the Transcriptional Repressor Snail Promotes Mammary Tumor Recurrence," *Cancer Cell*, vol.8, no.3, pp.197–209, 2005. Doi: 10.1016/J.Ccr.2005.07.009. Pubmed PMID: 16169465.
- [9] Ahmad A, "Pathways to Breast Cancer Recurrence," *ISRN Oncology*, pp.290568, 2013.Doi: 10.1155/2013/290568. Pubmed PMID: 23533807; Pubmed Central PMCID: PMC3603357.
- [10] Saphner T, Tormey DC, Gray R, " Annual Hazard Rates of Recurrence for Breast Cancer After Primary Therapy," *The Journal of Clinical Oncology*, vol.14, no.10, pp.2738–46, 1996. Doi: 10.1200/JCO.1996.14.10.2738. Pubmed PMID: 8874335.
- [11] Wang SY, Shamliyan T, Virnig BA, Kane R, "Tumor Characteristics as Predictors of Local Recurrence After Treatment of Ductal Carcinoma In Situ: A Meta-Analysis," *Breast Cancer Research and Treatment*, vol.127, no.1, pp.1–14, 2011. Doi: 10.1007/S10549-011-1387-4. Pubmed PMID: 21327465.
- [12] Chacón RD, Costanzo MV, "Triple-Negative Breast Cancer," *Breast Cancer Research and Treatment*, vol.12, no.S3, 2010. Doi: 10.1186/Bcr2574. Pubmed PMID: 21050424; Pubmed Central PMCID: PMC2972557.
- [13] Cavalli LR, "Molecular Markers of Breast Axillary Lymph Node Metastasis," *Expert Review of Molecular Diagnostic*, vol.9, no.5, pp.441–54, 2009. Doi: 10.1586/ErM.09.30. Pubmed PMID: 19580429.

- [14] Kim W, Kim KS, Lee JE, Noh DY, Kim SW, Jung YS, Et Al, "Development of Novel Breast Cancer Recurrence Prediction Model Using Support Vector Machine," *Journal of Breast Cancer*, vol.15, no.2, pp.230–8, 2010. Doi: 10.4048/Jbc.2012.15.2.230. Pubmed PMID: 22807942; Pubmed Central PMCID: PMC3395748.
- [15] Ahmad LG, Eshlaghy AT, Pourebrahimi A, Ebrahimi M, Razavi A, "Using Three Machine Learning Techniques for Predicting Breast Cancer Recurrence," *Journal of Health & Medical Informatics*, vol.4, no.2, pp.124–30, 2013. Doi: 10.4172/2157-7420.1000124.
- [16] Young I, Luz S, Lone N, "A Systematic Review of Natural Language Processing for Classification Tasks In the Field of Incident Reporting and Adverse Event Analysis," *International Journal of Medical Informatics*, vol.132, pp.103971, 2013. <https://doi.org/10.1016/J.Ijmedinf.2019.103971>.
- [17] Meystre S, Haug P, "Natural Language Processing to Extract Medical Problems From Electronic Clinical Documents: Performance Evaluation," *The Journal of Biomedical Informatics*, vol. 39, no.6, pp.589–599, 2006.
- [18] Sada Y Et Al, "Validation of Case Finding Algorithms for Hepatocellular Cancer From Administrative Data and Electronic Health Records Using Natural Language Processing," *Med Care*, vol.54, no.2, pp. E9-14, 2016.
- [19] Alzu'bi A, Zhou L, Watzlaf V, "Personal Genomic Information Management and Personalized Medicine: Challenges, Current Solutions, and Roles of HIM Professionals," *Perspectives in Health Information Management*, 11(Spring):1c, 2014.
- [20] Hardavella J Et Al, "Top Tips to Deal With Challenging Situations: Doctor-Patient Interactions," *Breathe*, vol.13, no.2, pp.129–135
- [21] Nayana Banjan, Prajka Dalvi and Neha Athavale, "Melanoma Skin Cancer Detection By Segmentation and Feature Extraction Using Combination of OTSU and STOLZ Algorithm Technique," *SSRG International Journal of Electronics and Communication Engineering*, vol. 4, no. 4, pp. 21-25, 2017. Crossref, <https://doi.org/10.14445/23488549/IJECE-V4I4P105>.
- [22] Kundra H, Sadawarti H, "Hybrid Algorithm of Cuckoo Search and Particle Swarm Optimization for Natural Terrain Feature Extraction," *Research Journal of Information Technology*, vol.7, no.1, pp.58–69, 2015.
- [23] Hong W Et Al, "SVR With Hybrid Chaotic Immune Algorithm for Seasonal Load Demand Forecasting," *Energies*, vol. 4, pp.960–977, 2011.
- [24] Zhang Z, Hong W, Li J, "Electric Load Forecasting By Hybrid Self-Recurrent Support Vector Regression Model with Variational Mode Decomposition and Improved Cuckoo Search Algorithm," *IEEE Access*, vol.8, pp.14642– 14658.
- [25] Abualigah L, "Feature Selection and Enhanced Krill Herd Algorithm for Text Document Clustering. Studies In Computational Intelligence," *Springer International Publishing, Berlin*, 2019.
- [26] Abualigah L, Khader A, "Unsupervised Text Feature Selection Technique Based on Hybrid Particle Swarm Optimization Algorithm With Genetic Operators for the Text Clustering," *Journal of Supercomputer*, vol.73, pp.4773–4795, 2017. <https://doi.org/10.1007/S11227-017-2046-2>.
- [27] Abualigah L, "Multi-Verse Optimizer Algorithm: A Comprehensive Survey of Its Results, Variants, and Applications," *Neural Computer Application*, vol. 32, pp.12381–12401, 2020.
- [28] Guo J, Sun Z, Tang H, Jia X, Wang S, Yan X, Ye G, Wu G, "Hybrid Optimization Algorithm of Particle Swarm Optimization and Cuckoo Search for Preventive Maintenance Period Optimization," *Discrete Dynamics in Nature and Society*, 2016. <https://doi.org/10.1155/2016/1516271>.
- [29] Vinitha S, Hao Y, Hwang K, Wang Lu, Wang Li, "Disease Prediction By Machine Learning Over Big Data From Healthcare Communities," *Computing in Science & Engineering*, vol.8, no.1, 2019. <https://doi.org/10.1109/ACCESS.2017.2694446>.
- [30] Dahiwade D, Patle G, Meshram E, "Designing Disease Prediction Model Using Machine Learning Approach," *In 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC)*. IEEE, New York, 2019.
- [31] Chae S, Kwon S, Lee D, "Predicting Infectious Disease Using Deep Learning and Big Data," *International Journal of Environmental Research and Public Health*, vol.15, no.8, pp.1596, 2018.
- [32] Battineni G Et Al, "Applications of Machine Learning Predictive Models in the Chronic Disease Diagnosis," *Journal of Perinatal Medicine*, vol.10, no.2, pp.21.
- [33] Dawes T Et Al, "Machine Learning of Three-Dimensional Right Ventricular Motion Enables Outcome Prediction in Pulmonary Hypertension: A Cardiac MR Imaging Stud," *Radiology*, vol.283, no.2, pp.381–390, 2017.
- [34] Ford E, Carroll JA, Smith HE, Scott D, Cassell JA, "Extracting Information From the Text of Electronic Medical Records to Improve Case Detection: A Systematic Review," *Journal of the American Medical Informatics Association*, vol.23, no.5, pp.1007–1015, 2016.
- [35] Sharma H, Rizvi M, "Prediction of Heart Disease Using Machine Learning Algorithms: A Survey," *International Journal on Recent and Innovation Trends in Computing and Communication*, vol. 5, no.8, pp.99–104, 2017.
- [36] Ahmad L, Eshlaghy A, Poorebrahimi A, Ebrahimi M, Razavi A, "Using Three Machine Learning Techniques for Predicting Breast Cancer Recurrence," *Journal of Health & Medical Informatics*, vol. 4, no.2, 2017. <https://doi.org/10.4172/2157-7420.1000124>
- [37] Tseng Y Et Al, "Predicting Breast Cancer Metastasis by Using Serum Biomarkers and Clinicopathological Data With Machine Learning Technologies," *International Journal of Medical Informatics*, vol.128, pp.79–86, 2019.
- [38] Boeri C Et Al, "Machine Learning Techniques in Breast Cancer Prognosis Prediction: A Primary Evaluation," *Cancer Medicine*, vol.9, no.9, pp.3234–3243, 2009.
- [39] Chang C, Chen S, "Developing A Novel Machine Learning-Based Classification Scheme for Predicting SPCS in Breast Cancer Survivors," *Frontiers in Genetics*, vol.10, Bo.848, 2019. <https://doi.org/10.3389/Fgene.2019.00848>.

- [40] Eidemiller M Et Al, “ Long-Term Health Risk After Breast-Cancer Radiotherapy: Overview of Passes Methodology and Software,” *Radiation Protection Dosimetry*, vol.183, pp.259–263.
- [41] Yousefi M Et Al, “ Organ-Specific Metastasis of Breast Cancer: Molecular and Cellular Mechanisms Underlying Lung Metastasis,” *Cell Oncology* , vol.41, no.2, pp.123–140, 2018.
- [42] Feliciano E Et Al, “ Body Mass Index, Pam50 Subtype, Recurrence, and Survival Among Patients with Nonmetastatic Breast Cancer,” *Cancer*, vol. 123, no.13, pp.2535–2542, 2017.
- [43] Shim H Et Al., “Breast Cancer Recurrence According to Molecular Subtype,” *The Asian Pacific Journal of Cancer Prevention*, vol.15, no.14, pp.5539–44, 2014.
- [44] Nidhi Mongoriya, Vinod Patel, "Review the Breast Cancer Detection Technique Using Hybrid Machine Learning," *SSRG International Journal of Computer Science and Engineering*, vol. 8, no. 6, pp. 5-8, 2021.
Crossref, <https://doi.org/10.14445/23488387/IJCSE-V8I6P102>
- [45] Cancer.Net. [Internet] Doctor-approved patient information from ASCO. Breast cancer: statistics, 2021 Jan. Available from: <https://www.cancer.net/cancer-types/breast-cancer/statistics>.
- [46] Gerhard W, “The Diagnosis, Pathology, and Treatment of the Diseases of the Chest,” *Philadelphia: E. Barrington And G.D. Haswell*, 1850. <http://resource.nlm.nih.gov/101505669>.