



Data Splitting Techniques to Reduce False-Positive and False-Negative Cases in Breast Cancer Prediction

Vijay Birchha¹, Bhawna Nigam²

¹ Computer Engineering, Department of Computer Engineering, IET, DAVV, Indore, 452001, India

² Department of Information Technology, IET, DAVV, Indore, 452001, India.

9410

Abstract-

The massive worldwide number of women affected with breast cancer; is the most common and severe cause of women's high mortality rate. The false diagnosis can be considered the most significant cause of the late discovery of breast cancer. The chances of curing breast cancer increase if the number of false-positive and false-negative predictions is reduced. The research objectives are; can dataset splitting techniques used to train the machine learning classifiers affect the classifier performance?; do they help to minimize false-positive and false-negative predictions of breast cancer? In this work, artificial neural network (NN), support vector machine (SVM), logistic regression (LR) and decision forest (DF) machine learning (ML) classifiers were used with The breast cancer Wisconsin (original) dataset (WBC). The classifier's false-positive and false-negative predictions were compared with different dataset splitting techniques train-test (TT), train-test-validation (TTV) and k-fold cross-validation. The neural network classifier scored zero FP predictions with the train-test-validation dataset splitting method. The support vector machine recorded zero FN predictions with the k-fold cross-validation dataset splitting method. The results proved that the selection of dataset splitting techniques significantly impacts machine learning classifier performance. The result will help implement a computer-aided system to diagnose breast cancer more accurately.

Keywords- Breast cancer, Wisconsin dataset, machine learning, false-positive, false-negative, support vector machine, decision forest, neural network, logistic regression, dataset split,

DOI Number: 10.14704/nq.2022.20.8.NQ44961

NeuroQuantology 2022; 20(8): 9410-9424

Introduction

Breast cancer is a highly crucial cause of mortality in women due to cancer [1]. The false-positive recall is a severe problem in breast cancer screening; due to this, patient anxiety and healthcare expenses can be increased. Optimized screening techniques can decrease false-positive rates; thus, breast radiologists apply more research efforts to reduce false-positive recall rates. However, false-negative predictions are also hazardous to patient life. Most of the research was conducted on medical image analysis and cancer tissue biopsy; still, there is a gap in applying efforts to standardize machine

learning methods to reduce false-positive and false-negative predictions.

In the present time, machine learning (ML) has substantially impacted healthcare systems. Breast cancer can be detected by using an ML-enabled computer-aided diagnostics (CAD) system more accurately in the early stage. Using ML-enabled CAD systems in healthcare can overcome the deficiency of medical experts. The four types of predictions made by machine learning-based computer-aided diagnostic (CAD) systems are "true-positive (TP), true-negative (TN), false-positive (FP), and false-negative (FN)" [2]. The CAD system's forecast using the confusion matrix is shown in Figure 1. A confusion matrix is a



table that summarizes how well a classifier performed when faced with a classification challenge. The confusion matrix is used to evaluate classifiers by comparing the predicted and actual outcomes of the models. A confusion matrix shows how many instances are classified correctly or incorrectly. A false-positive case is predicted as positive (cancerous), but actually, it is negative (non-cancerous), and a false-negative case is indicated as non-cancerous

but cancerous. The false-positive recall is one of the breast cancer screening primary shortcomings, and it is frequently related to increased patient anxiety, healthcare expenses, and the need for invasive treatments [3]. A decrease in false-positive rates can help standardize screening techniques; thus, breast radiologists must maintain the importance of reducing false-positive recall rates. However, false-negative predictions are also hazardous to patient life.

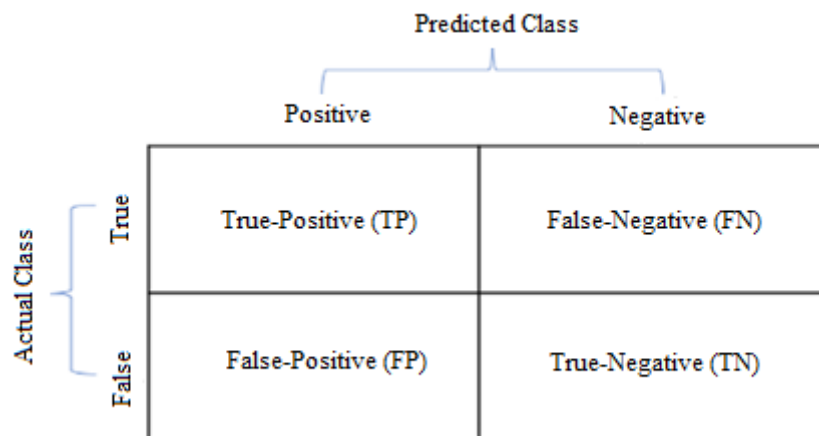


Figure 1 Confusion matrix

2. Related work

The following studies have worked and proposed various solutions to reduce false-positive and false-negative predictions of breast cancer detection.

In [4], The author experimented with and compared various magnetic resonance imaging (MRI)/CAD systems and found that systems can diagnose correctly but showed breast cancer varying FP rates; The breast MRI FP rates can be minimalized by ignoring lesions with a small threshold to reduce unnecessary biopsies of freshly diagnosed breast cancer [4]. Similarly, In [5] author worked with MRI images of women with dense breasts and suggested considering BI-RAD 3 and BI-RAD 4 as benign to reduce false-positive diagnoses.

In [6] and [7], both authors have worked on ultrasound (US) images of the breast and proposed solutions to reduce false-positive predictions of breast cancer. In [6] author has submitted to combine axillary lymph nodes and primary tumour features to minimize the false-positive prediction; this will help to evade needless sentinel lymph node biopsy (SLNB). In [7], the author found breast ultrasonography (US) screening has a high rate of false-positive biopsies; the author developed a CAD prediction system based on deep learning to reduce false-positive diagnoses.

In [8], the author explored the deep learning method with annotation to improve the accuracy with reduced FP and FN breast cancer predictions on mammography images. In [9], the author explored that finding breast abnormalities from X-ray, MRI and



mammography images is a critical task for radiologists. The author proposed a CAD system using a Deep Convolutional Neural Network (CNN) to segregate and categorize the many kinds of breast abnormalities. In [10], the author started the research with the objective that the application of artificial intelligence (AI) can minimize the interval cancer rate in mammography screening. The author found that Using AI in screen reading can reduce the interval cancer rate without supplementary screening modalities. In [11], the author presented a CAD system to detect lesions or anomalies in breast MRI data using a regional CNN model. In [12], The author introduced a lightweight breast tumour detection and classification system based on enhanced YOLOv5. To reduce the impact of uneven staining on pathological image categorization, the author presented a CycleGAN-based colour normalizing technique for pathological image slices [13].

While studying the related research work, it is absorbed that some authors tried to reduce false-positive and false-positive predictions by ignoring small lesions [4], and some said to consider BIRAD3 and BI-RAD4 as benign [5]. One author made ultrasound responsible for the higher rate of false FP diagnosis [7]. Various authors used the latest deep convolution neural network (CNN) to improve the accuracy and reduce the FP diagnosis [8]–[12]. As a research gap, it is observed that the effect of data splitting techniques on machine learning classifiers for breast cancer detection is unexplored.

The research started with the question; can dataset splitting techniques used to train the machine learning classifiers affect the classifier performance; do they help to minimize false-positive and false-negative predictions of breast cancer? Dataset splitting techniques are required to design a subset of datasets to train and test the ML classifiers; for research, three dataset splitting techniques were considered: test and train,

test-train-validation, and k-fold cross-validation. The ML classifiers identified for this work are artificial neural network (NN), two-class support vector machine (SVM), two-class logistic regression (LR) and two-class decision forest (DF).

The paper has organized into various sections; section 1 briefly introduces breast cancer, possible causes and the application of an MI-enabled CAD system; section 2 includes related research work done by other authors in the same problem domain. Section 3 discussed the dataset, Dataset splitting techniques, ML methods and model evaluation parameters considered for work and the evaluation parameters considered for research. Section 4 has the research process that includes dataset pre-processing, data visualization, and research execution plan, and section 5 explains the experiment's results and limitations. At last, section 6 contains the research findings and the future recommendation for extension of work.

3. Material and methods

This section discusses dataset, dataset analysis and machine learning methods.

3.1 Breast cancer Wisconsin (original) dataset (WBC)

At the University of Wisconsin Hospitals at Madison, Dr William H. The WBC dataset was collected from Wolberg [14]. The dataset consists of features that are dependent on various parameters. There are a total of 699 samples. The dataset has been divided into two categories: benign and malignant samples. The dataset contains 458 benign and 241 malignant samples; all the given samples were used in this research. The original dataset has 11 columns; one of them, sample ID, is less valuable. The class feature works as a target discriminator for a cancer diagnosis; Breast cancer has two possible values in the class variable: benign and malignant. The remaining nine features, clump thickness,



uniformity of cell size, uniformity of cell shape, marginal adhesion, single epithelial cell size, bare nuclei, bland chromatin, normal nucleoli and mitoses correspond to nine

characteristics of breast fine-needle aspirates (FNA s).

Table 1 WBC Dataset summary

Features	Count	Count	Count	Min	Max	Mean	Deviation	Median	Mode	Range	Variance	SD	Skewness	Kurtosis
Sample code number	69	64	0	-	-	-	-	-	-	-	-	-	-	-
Clump Thickness	69	10	0	1	10	4.42	2.3	4	1	9	7.93	2.82	0.59	-0.62
Uniformity of Cell Size	69	10	0	1	10	3.13	2.51	1	1	9	9.31	3.05	1.23	0.1
Uniformity of Cell Shape	69	10	0	1	10	3.21	2.47	1	1	9	8.83	2.97	1.16	0.01
Marginal Adhesion	69	10	0	1	10	2.81	2.24	1	1	9	8.15	2.86	1.52	0.99
Single Epithelial Cell Size	69	10	0	1	10	3.22	1.69	2	2	9	4.9	2.21	1.71	2.17
Bare Nuclei	69	11	0	0	10	3.46	3.15	1	1	10	13.25	3.64	1.02	-0.73
Bland Chromatin	69	10	0	1	10	3.44	1.95	3	2	9	5.95	2.44	1.1	0.18
Normal Nucleoli	69	10	0	1	10	2.87	2.46	1	1	9	9.32	3.05	1.42	0.47
Mitoses	69	9	0	1	10	1.59	0.98	1	1	9	2.94	1.72	3.56	12.66
Class	69	2	0	2	4	-	-	-	-	-	-	-	-	-

3.2 Dataset split methods

The dataset splitting techniques considered for the research are Train and test (TT) method, Train-test and validation (TTV) method and K-fold cross-validation (k-fold) method.

Train and test set split

The dataset splits into two subsets, and records are selected randomly; one set is the training set which consists of 70% of records of the original set, and the second set is the test set containing the remaining records of

the dataset. The data in the training set is utilized for training the model with selected ML classifier algorithms. The test set was used to assess the trained model’s performance.

Train, test and validation set split

Dataset was split into three sets; the first set was the train set containing 80%, the second set was the test set including 10% records, and validation had the remaining 10% records. The records were selected randomly. The train and test sets were utilized the same way as mentioned above for the TT dataset



split. The validation set was used for tuning the hyperparameters,

K-fold cross-validation split

The dataset is separated into the test set and training set; Then, the training set is further divided into k distinct sets. In this research,

the training data is divided into ten different groups. Each iteration uses nine random sets to train the model and one set as a validation[15]; Finally, the test set was utilized to assess the classifiers' performance.

3.3 Machine learning algorithms

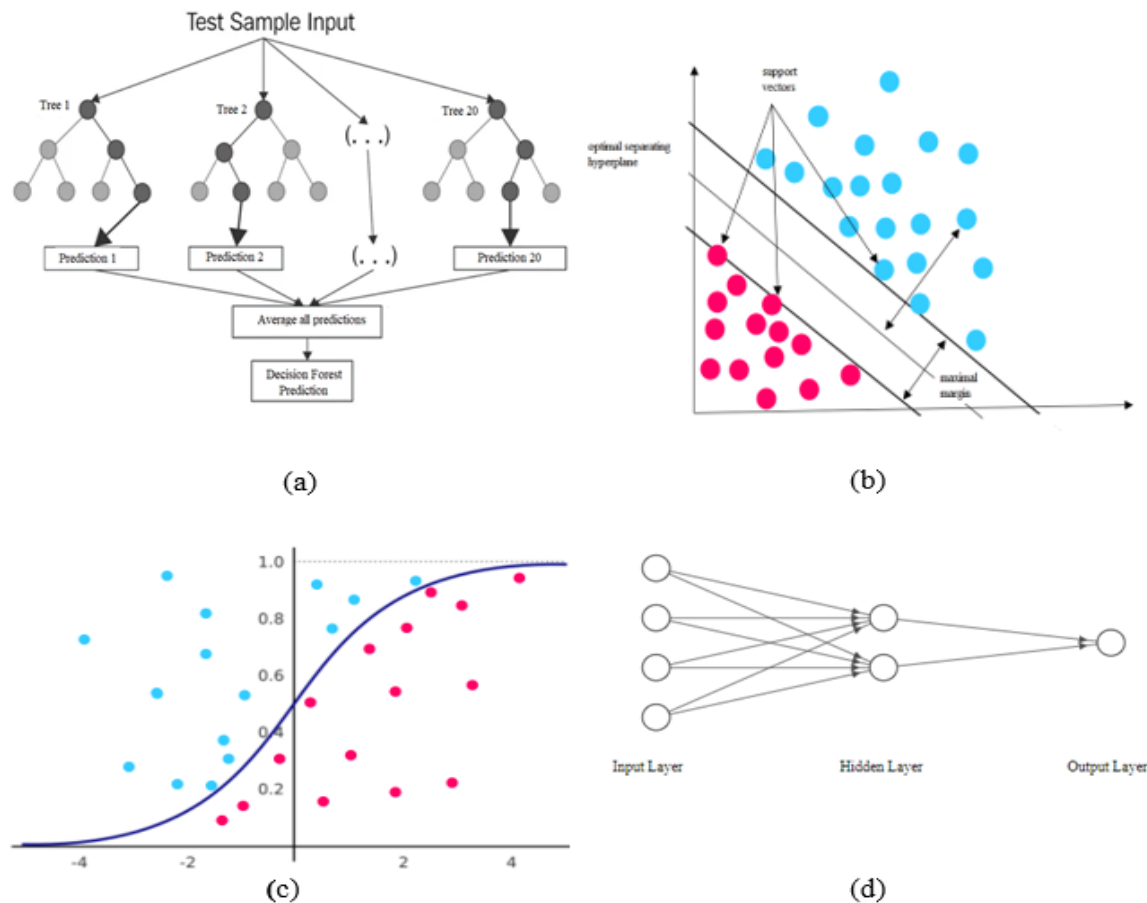


Figure 2 Machine learning algorithms (a) Decision Forest, (b) Support vector machine, (c) Logistic regression, (d) Artificial neural network

Decision forest- A decision forest comprises several decision trees, the predictions of which merge to provide a final forecast. A decision forest increases the accuracy of a single decision tree by aggregating the predictions of multiple trained trees. The concept of decision forests was developed by fusing the ensemble and decision tree learning techniques. Decision forest is utilized in various real-world applications, such as medicine, engineering, and information retrieval [16].

Support vector machine- A support vector machine (SVM) is a supervised learning algorithm for classification and regression. SVMs are capable of handling both continuous and categorical inputs. SVM attempts to discover a hyperplane that separates training examples into binary classes; A SVM maximizes the margin. SVM makes a slight trade-off between high margins and classification accuracy. If the accurate classification is employed without compromising any sample, the margin of error



may be relatively small, resulting in a lesser level of precision. However, by widening the margin in the middle of classes, support vectors closer to the hyperplane could be evaluated with members of other classes[17].

Logistic regression-“Logistic regression is a mathematical modelling approach that can describe the relationship of several independent variables to the dependent variable of the binary class”[18]. The logistic model is widely accepted in healthcare because it represents the combined effect of several disease risk factors in an S-shaped graph, and its estimates must lie between zero and one [18]. In addition, logistic regression works well with classification and regression problems[19].

Artificial neural network-The neural network is a replica of biological neurons. The majority of Artificial Neural Networks consist of

learning algorithms that adjust the weight of connecting nodes to the input patterns[20]. Each processing unit in a neural network is a node. A node is divided into two parts: the summing part and the output part; the summing part accepts N input values and generates a weighted sum. The activation function is applied to the weighted sum to generate the output part; several nodes are interconnected to form a neural network[21].

3.4 EVALUATION PARAMETERS

Accuracy, F1 score, precision, recall, and specificity compare the ML algorithm’s performance. The evaluation parameters given below are derived from the confusion matrix.

Accuracy-Classification accuracy is a statistic that expresses the performance of the classification model as the ratio of correct predictions to total predictions.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (5)$$

Precision-Precision is the measure of a model to classify actual positive cases out of all the positive classifications.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (6)$$

Sensitivity-Sensitivity or recall is a metric that indicates how well the ML model correctly identifies True Positives samples[16].

$$\text{Sensitivity(Recall)} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (7)$$

F1 Score- The F1 score combines precision and recall metrics into a single value.

$$\text{F1 Score} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (8)$$

Specificity- Specificity is a metric that indicates how well the ML model correctly identifies true negatives samples.

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \quad (9)$$

4. Methods

4.1 Data analysis and visualization

Data visualization is the graphical display of data. These tools help users understand data by using visual features such as charts, graphs, and maps.

In the heatmap, correlated features are represented with reddish color blocks, and

less correlated features show with blueish color blocks. Uniformity of cell size is closely connected with uniformity of cell shape, the size of a single epithelial cell, and the absence of chromatin structure. The clump thickness, bare nuclei and bland chromatin are more helpful in creating a more accurate predictive model using machine learning.



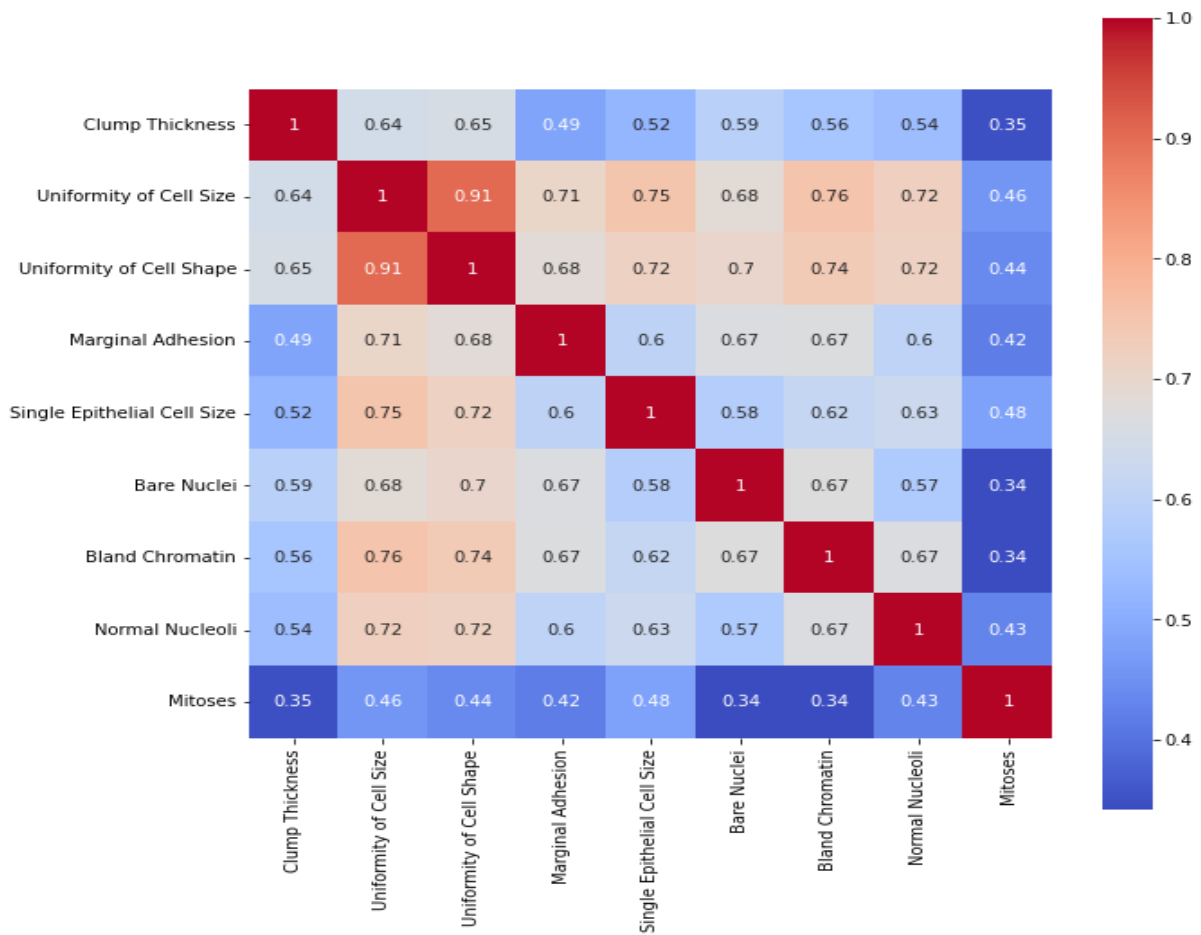


Figure 3 Heatmap of WBC dataset

4.2 Dataset pre-processing

All the records of the WBC dataset were used in the investigation. When the dataset was visualized, the sample id first column was less helpful for further studies. In the second step, remove missing values and duplicate records. Such cleaning operations aimed to prevent problems caused by missing and duplicate

data when training the model. The mean value of that column was used to fill missing values. After cleaning the missing value, the WBC dataset was normalized using the MaxMin method. The class column was excluded from the normalization process. After the normalization, the dataset was divided.



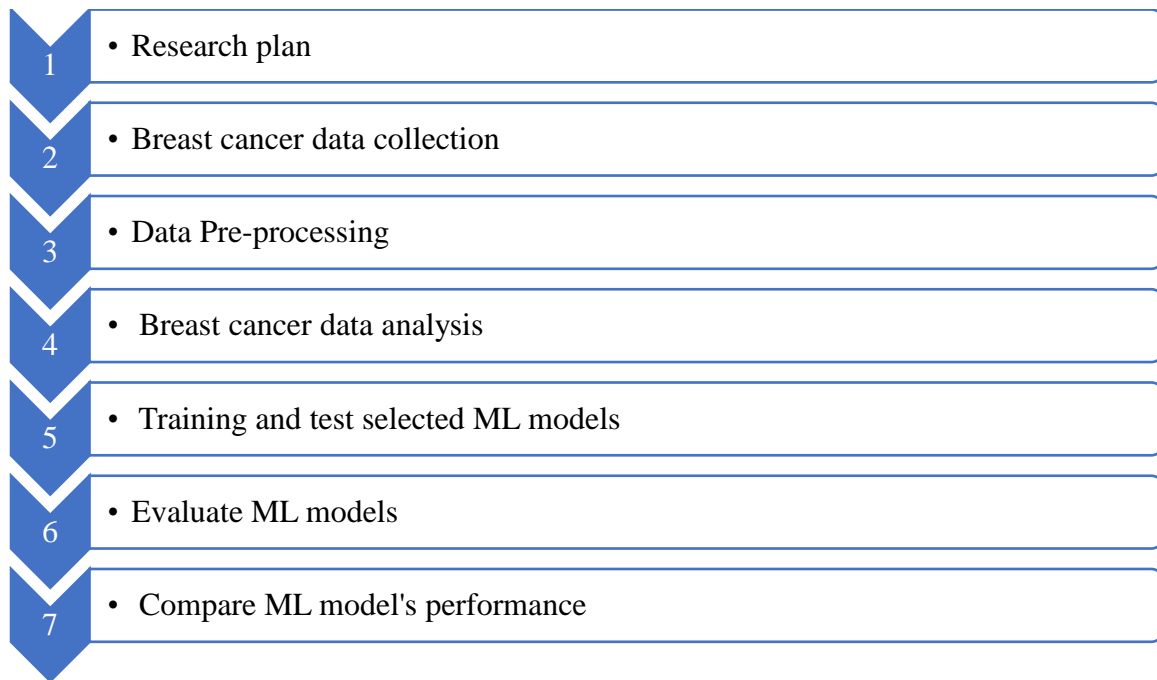


Figure 4 Research process

4.3 Research plan and implementation

The research experiment was conducted using a WBC dataset. The WBC dataset analysis and data visualization was done using the language R on RStudio stated in section 4.1; The WBC dataset was pre-processed using the method said in section 4.2; the WBC dataset features correlation visualized in heatmap figure 3. The selected ML classifiers were implemented using the algorithms given in section 3.3. The selected ML model’s performance was evaluated using the equations mentioned in section 3.4. The research process was implemented using Microsoft Azure Machine Learning Studio. The research process was executed successfully by utilizing all the steps mentioned in figure 4.

5. Result and discussion

The WBC dataset contains 699 records; all the unique 645 records of the WBC dataset were utilized for the experiment; all the features were employed in experiments except sample id. WBC dataset last column denote class or category of diagnosis. The experiments were conducted using SVM, LR, DF and NN classification algorithms; one by one, different dataset split techniques TT, TTV and k-fold were employed. Initially, the first set of experiments was conducted with a TT dataset split; the training set contains 70% of unique records, and the test includes the remaining 30%. Before initiating the experiment, all four classification models’ hyperparameter

Table 2 ML classifiers performance evaluation with Train-Test (TT) dataset split

ML Model	Accuracy	Precision	Recall	F1	Specificity	TN	FP	FN	TP
Support Vector Machine	0.9793	0.9733	0.9733	0.9733	0.9831	116	2	2	73
Logistic Regression	0.9896	0.9867	0.9867	0.9867	0.9915	117	1	1	74
Decision Forest	0.9896	0.9867	0.9867	0.9867	0.9915	117	1	1	74
Neural Network	0.9896	0.9867	0.9867	0.9867	0.9915	117	1	1	74



tuning was done using a TT dataset split method on the WBC dataset. The performance of all four classification models with TT, TTV, and k-fold dataset splits methods were recorded in Tables 2, 3 and 4, respectively. Table 2 consists of the experiment results; LR, DF and NN recorded an accuracy of 0.99 and an F1 score of 0.987. LR, DF and NN have predicted minimum values of one FP and FN. SVM classifier scored minimum in all parameters; its performance was more than 95% in all parameters but smaller than other classifiers. In Table 2 decision to choose the best classifier was a tie between LR, DF and NN.

The second set of experiments was done using TTV dataset splits, test and validation set consisting of 10% and 10% of records and the training set includes 80% of the WBC dataset unique records. The experiment results are stored in table 3. SVM, LR, DF and NN classifiers were trained using a train set. Models performance tuning was done using a validation set. After the completion of training, models were evaluated using a test set; SVM and LR classifiers have gained higher accuracy of 0.968. DF and NN classifiers scored an accuracy of score 0.95. LR and

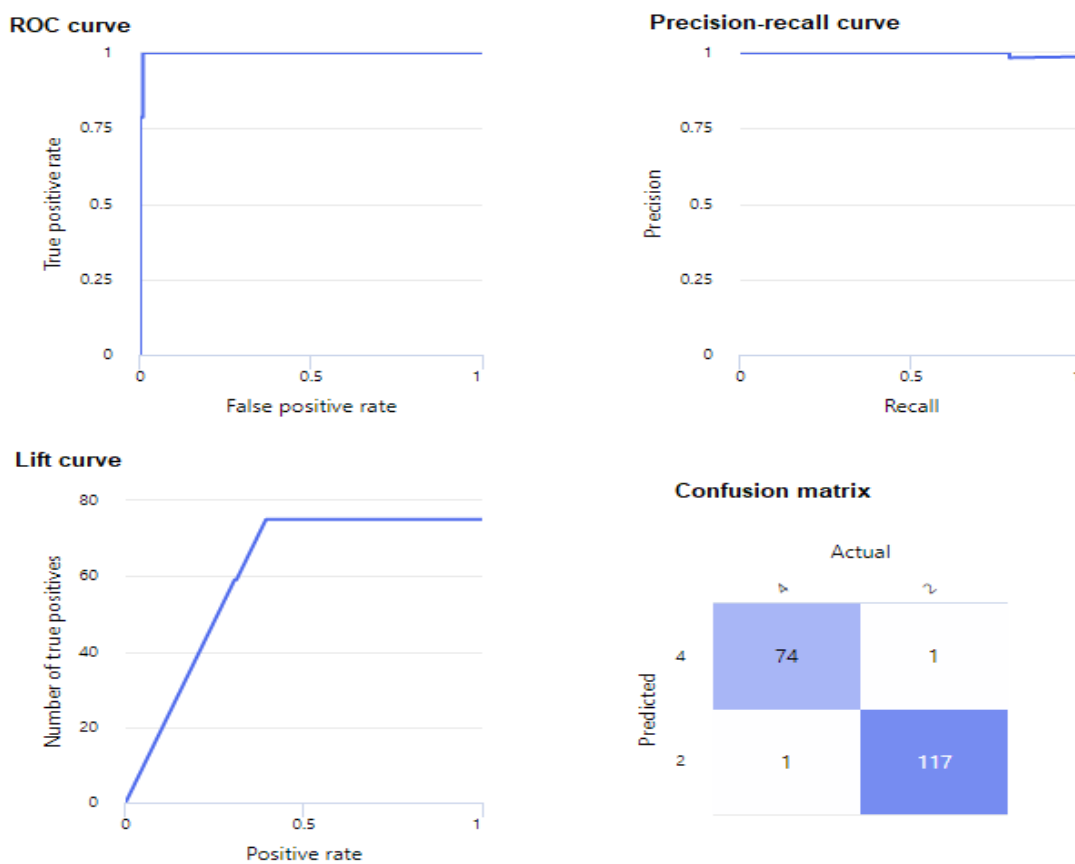


Figure 5 LR model performance with TT dataset splitting method

SVM classifiers achieved the highest F1 score of 0.96. NN model has classified zero false-positive samples. SVM and LR scored the minimum number of false-negative predictions, that is, one. The third set of experiments was done with k-fold cross-validation, the value of k was set to 10, and obtained results are recorded in table 4; the SVM classifier scored the highest accuracy score of 0.98, and the SVM classifier predicted zero false-positive samples and scored recall value 1. SVM, LR, DF, and NN classifiers predicted a false-positive score of one and a specificity score of 0.97.



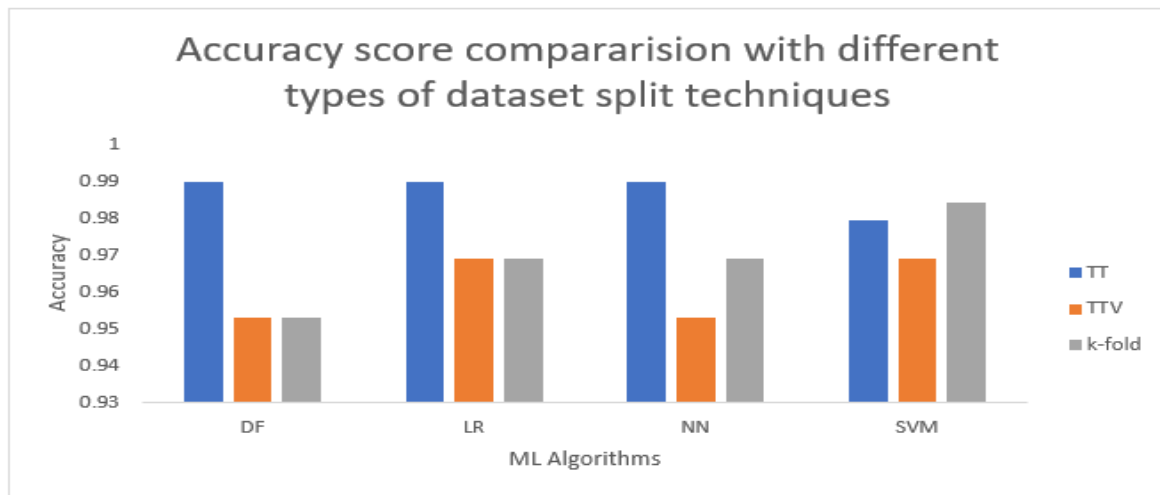


Figure 6 Comparison of classifier’s accuracy

The research work started with the question; can dataset splitting techniques used to train the machine learning classifiers affect the false-positive and false-negative predictions of breast cancer? DF, LR, NN and SVM classifiers were trained using the WBC dataset; the classifier’s hyperparameters were already tuned to improve performance. However, the result mentioned in table 2,3 and 4 indicates that the performances of various classifiers were altered with the dataset split methods.

Table 3 ML classifiers performance evaluation with Train-Validation-Test splits.

ML Model	Accuracy	Precision	Recall	F1	Specificity	TN	FP	FN	TP
Support Vector Machine	0.9688	0.9583	0.9583	0.9583	0.975	39	1	1	23
Logistic Regression	0.9688	0.9583	0.9583	0.9583	0.975	39	1	1	23
Decision Forest	0.9531	0.9565	0.9167	0.9362	0.975	39	1	2	22
Neural Network	0.953	1	0.875	0.933	1	40	0	3	21

The instances from the WBC dataset were distributed into two sets; samples were randomly selected; random selection is the best option if you want to give equal weightage to all instances. It is the preferred choice when creating training and test datasets. When increased the ratio of records in the train set, the models scored higher performance. The research tried with 50% to 90% split options and found that 70% was given a balanced performance. In the TTV method, the WBC dataset was split into three parts. The validation set was highly beneficial to tuning hyperparameters and helped to avoid the model’s overfitting.

Based on tables 2, 3 and 4, DF, LR, and NN, classifiers scored the highest number of correct predictions of breast cancer from the total predictions. DF, LR, and NN classifiers achieved an accuracy score of 0.99 with the TT dataset split method, and this was the highest score compared to all other classifiers. DF and NN classifiers scored the minimum accuracy score of 0.95 with the TTV dataset split method, which is the minimum score compared to all others; DF, LR and NN classifiers scored the maximum and minimum accuracy score with different dataset split methods.



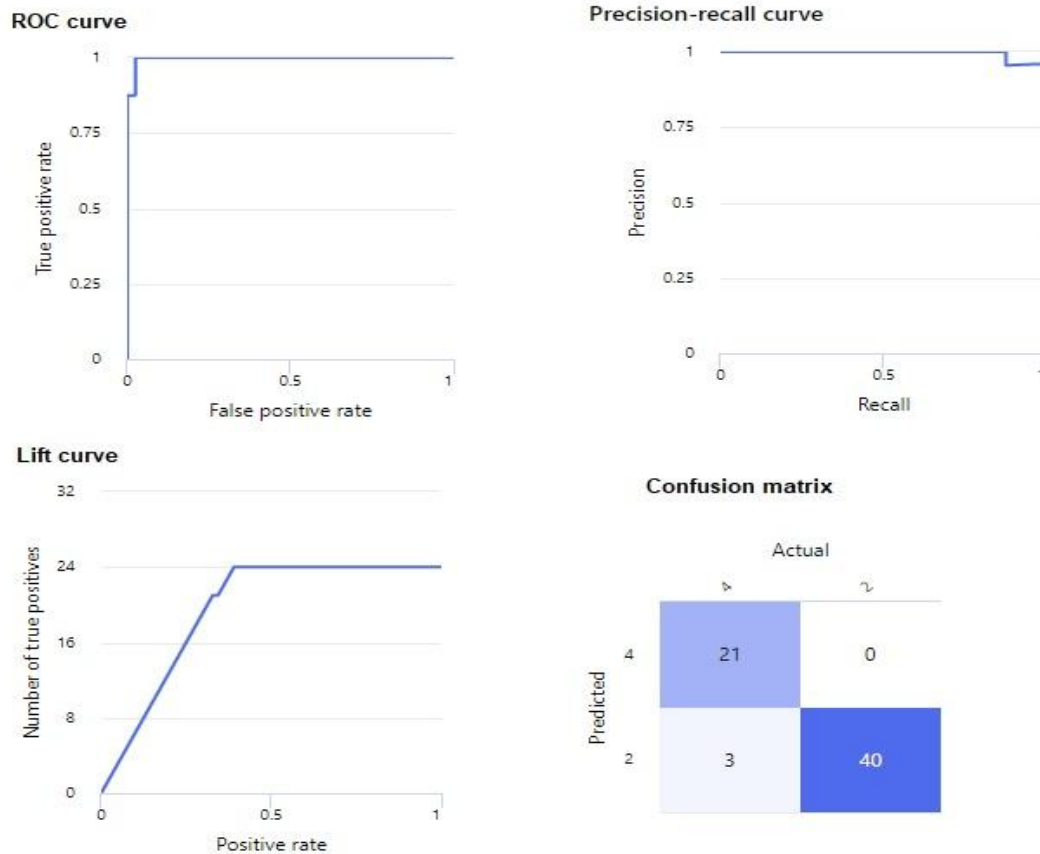


Figure 7 SVM model performance with TTV dataset splitting method

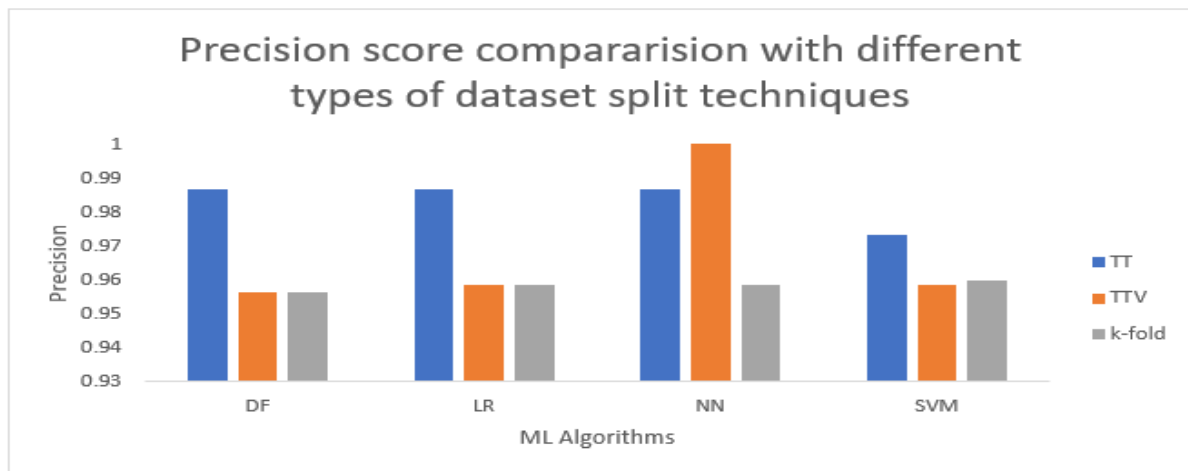


Figure 8 Comparison of classifier's precision

Based on figure 6 and Tables 2,3 and 4, the TP and FP considered positive predictions, and the SVM classifier correctly classified the largest number of positive cases from all predictions. DF classifiers scored a minimum precision score of 0.95 with the TTV and k-fold dataset split methods, and NN classifiers achieved a higher precision score of 1 with the TTV dataset method; in the case of precision TTV and k-fold created the same effect on DF and LR classifier and the other classifiers were performed differently. NN classifier recorded a precision score of 0.95 to 1 with different dataset split methods.



Table 4 10-fold Cross Validation splits ML models performance results

ML Model	Accuracy	Precision	Recall	F1	Specificity	TN	FP	FN	TP
Support Vector Machine	0.984	0.96	1	0.98	0.975	39	1	0	24
Logistic Regression	0.9688	0.9583	0.9583	0.9583	0.975	39	1	1	23
Decision Forest	0.9531	0.9565	0.9167	0.9362	0.975	39	1	2	22
Neural Network	0.9688	0.9583	0.9583	0.9583	0.975	39	1	1	23

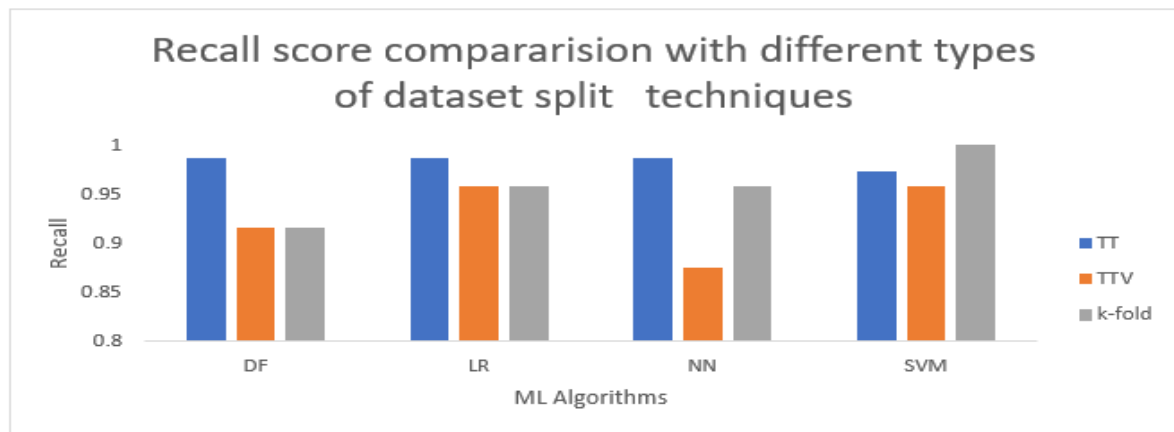


Figure 9 Comparison of classifier’s recall score

The false-negative predictions are cancerous cases but predicated as non-cancerous; the sensitivity score helps evaluate any classifier’s performance to classify true-positives

10-fold Cross-validation is very productive when the dataset is small in size. For example, the result mentioned in Tables 2,3 and 4 helped to find that the SVM classifier predicted the highest number of true-positive cases out of total true-positive and false-negative predictions of breast cancer with the k-fold dataset split method.

When selecting any classifier based on precision and sensitivity score is complex, the F1 score compares classifiers’ performance. Furthermore, based on the F1-score, the DF, LR and NN are the best classifiers with the TT dataset split method, results accessed from tables 2,3 and 4.

From tables 2,3, and 4, The NN classifier was predicted, zero false-positive with TTV dataset split methods, thus scoring maximum specificity. On the other hand, DF, SVM, and LR classifiers recorded the unexpected poorest performance with TTV and k-fold dataset split methods. DF scored minimum accuracy, precision, recall, F1-score and specificity with TTV and k-fold dataset split method. Figures 5, 6 and 7 denote that the DF and LR classifier performed similarly with TTV and k-fold dataset split method. The NN and SVM showed the highest variation with the dataset set split method.



The research shows that the dataset split methods TT, TTV, and k-fold significantly impact classifier performance in detecting breast cancer. The study was conducted with the Wisconsin original dataset (WBC).

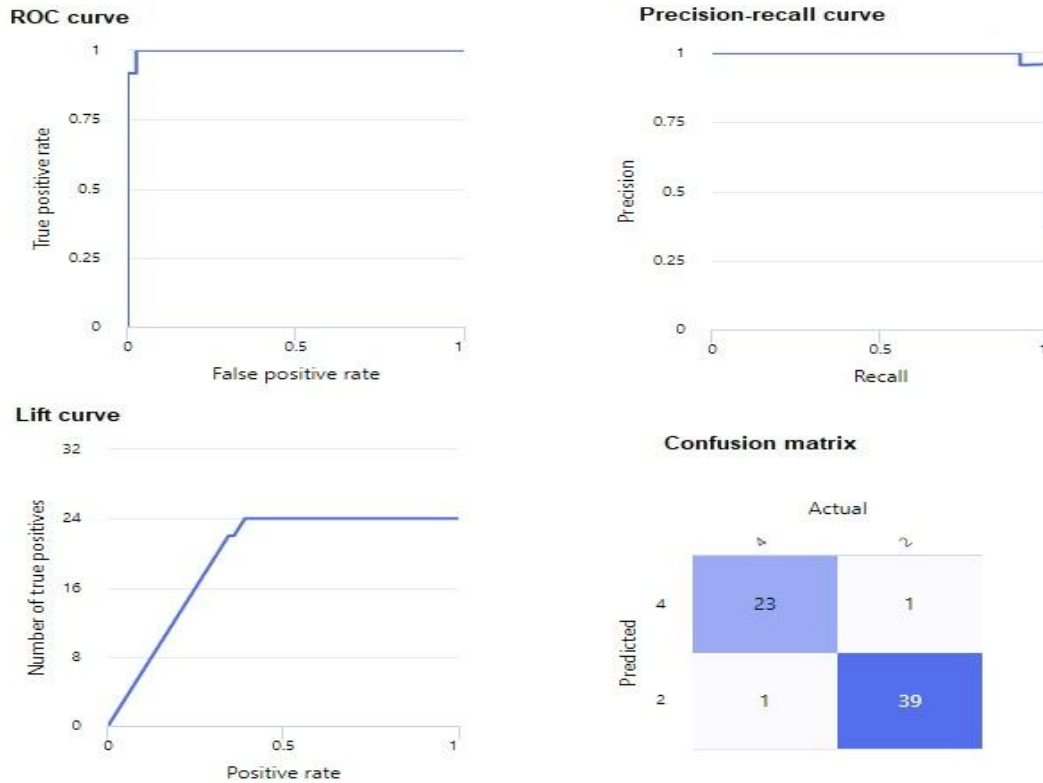


Figure 10 NN model performance with k-fold cross validation dataset splitting method

6. CONCLUSION

This research aimed to investigate the dataset splitting technique's effect on machine learning classifier's false-positive and false-negative predictions of breast cancer; This work has considered three dataset splitting techniques: test and train, test-train-validation, and k-fold cross-validation; the ML classifiers identified for this work are neural network (NN), support vector machine (SVM), logistic regression (LR) and decision forest (DF). The research found that the dataset splitting techniques affected the performance of classifiers; The classifiers were performed differently with the different splitting methods; the neural network scored zero false-positive with the test-train and

validation dataset splitting method. The support vector machine classifier achieved zero false-negative with the k-fold cross-validation dataset splitting method. The research work was conducted with the breast cancer BBC dataset. The neural network classifier achieved a specificity of 1.0, an accuracy of 0.95 and a sensitivity of 0.87 with the test-train and validation dataset splitting method. The support vector machine scored zero false-negative with an accuracy score of 0.98, sensitivity score of 1.0 and specificity score of 0.975 with the k-fold cross-validation dataset splitting technique. In future research, work will further extend with feature selection techniques impact with best dataset splitting methods.

REFERENCES

- [1] "Cancer Facts & Figures 2022," p. 80, 1930.
- [2] D. Geng, D. Yang, M. Cai, and L. Zheng, "A Novel Microwave Treatment for Sleep Disorders and Classification of Sleep Stages Using Multi-Scale Entropy," *Entropy*, vol. 22, no. 3, p. 347, Mar. 2020, DOI: 10.3390/e22030347.
- [3] A. G. V. Bitencourt, C. R. Saccarelli, and E. A. Morris, "How to Reduce False Positive Recall Rates in Screening Mammography?," *Acad. Radiol.*, vol. 26, no. 11, pp. 1513–1514, Nov. 2019, DOI: 10.1016/j.acra.2019.06.003.
- [4] T. Ha, D. K. Kang, and T. H. Kim, "Percentage volume of delayed kinetics in computer-aided diagnosis of MRI of the breast to reduce false-positive results and unnecessary biopsies," *Clin. Radiol.*, vol. 75, no. 12, p. 962.e1-962.e8, Dec. 2020, DOI: 10.1016/j.crad.2020.08.005.
- [5] E. Verburg *et al.*, "Computer-Aided Diagnosis in Multiparametric Magnetic Resonance Imaging Screening of Women With Extremely Dense Breasts to Reduce False-Positive Diagnoses," *Invest. Radiol.*, vol. Publish Ahead of Print, Mar. 2020, DOI: 10.1097/RLI.0000000000000656.
- [6] C.-B. Yi *et al.*, "Combining the Ultrasound Features of Primary Tumor and Axillary Lymph Nodes Can Reduce False-Negative Rate during the Prediction of High Axillary Node Burden in BI-RADS Category 4 or 5 Breast Cancer Lesions," *Ultrasound Med. Biol.*, vol. 46, no. 8, pp. 1941–1948, Aug. 2020, doi: 10.1016/j.ultrasmedbio.2020.04.003.
- [7] S.-Y. Kim *et al.*, "Deep learning-based computer-aided diagnosis in screening breast ultrasound to reduce false-positive diagnoses," *Sci. Rep.*, vol. 11, no. 1, p. 395, Dec. 2021, doi: 10.1038/s41598-020-79880-0.
- [8] W. Lotter *et al.*, "Robust breast cancer detection in mammography and digital breast tomosynthesis using an annotation-efficient deep learning approach," *Nat. Med.*, vol. 27, no. 2, pp. 244–249, Feb. 2021, DOI: 10.1038/s41591-020-01174-9.
- [9] M. Heenaye-Mamode Khan *et al.*, "Multi-class classification of breast cancer abnormalities using Deep Convolutional Neural Network (CNN)," *PLOS ONE*, vol. 16, no. 8, p. e0256500, Aug. 2021, DOI: 10.1371/journal.pone.0256500.
- [10] K. Lång, S. Hofvind, A. Rodríguez-Ruiz, and I. Andersson, "Can artificial intelligence reduce the interval cancer rate in mammography screening?," *Eur. Radiol.*, vol. 31, no. 8, pp. 5940–5947, Aug. 2021, DOI: 10.1007/s00330-021-07686-3.
- [11] Y. Zhang *et al.*, "Automatic Detection and Segmentation of Breast Cancer on MRI Using Mask R-CNN Trained on Non-Fat-Sat Images and Tested on Fat-Sat Images," *Acad. Radiol.*, vol. 29, pp. S135–S144, Jan. 2022, DOI: 10.1016/j.acra.2020.12.001.
- [12] A. Mohiyuddin *et al.*, "Breast Tumor Detection and Classification in Mammogram Images Using Modified YOLOv5 Network," *Comput. Math. Methods Med.*, vol. 2022, pp. 1–16, Jan. 2022, DOI: 10.1155/2022/1359019.
- [13] P. Chopra, N. Junath, S. K. Singh, S. Khan, R. Sugumar, and M. Bhowmick, "Cyclic GAN Model to Classify Breast Cancer Data for Pathological Healthcare Task," *BioMed Res. Int.*, vol. 2022, pp. 1–12, Jul. 2022, DOI: 10.1155/2022/6336700.
- [14] O. L. Mangasarian, W. N. Street, and W. H. Wolberg, "Breast Cancer Diagnosis and Prognosis Via Linear Programming," *Oper. Res.*, vol. 43, no. 4, pp. 570–577, 1995.
- [15] D. Asanka, "Building Ensemble Classifiers in Azure Machine Learning," *SQL Shack - articles about database auditing, server performance, data recovery, and more*, Sep. 21, 2021. <https://www.sqlshack.com/building-ensemble-classifiers-in-azure-machine-learning/> (accessed Nov. 15, 2021).



[16] L. Rokach, "Decision forest: Twenty years of research," *Inf. Fusion*, vol. 27, pp. 111–125, Jan. 2016, doi: 10.1016/j.inffus.2015.06.005.

[17] A. F. M. Agarap, "On breast cancer detection: an application of machine learning algorithms on the Wisconsin diagnostic dataset," in *Proceedings of the 2nd International Conference on Machine Learning and Soft Computing - ICMLSC '18*, Phu Quoc Island, Viet Nam, 2018, pp. 5–9. DOI: 10.1145/3184066.3184080.

[18] S. Nanga and A. Lotsi, "A Comparison Between Logistic Regression And Neural Networks In Modeling Mortality Amongst Children Under Five Years In Ghana".

[19] D. G. Kleinbaum and M. Klein, "Introduction to Logistic Regression," in *Logistic Regression: A Self-Learning Text*, D. G. Kleinbaum and M. Klein, Eds. New York, NY: Springer New York, 2010, pp. 1–39. DOI: 10.1007/978-1-4419-1742-3_1.

[20] N. Ramprasad, "Understanding Neural Networks for Machine Learning using Microsoft Neural Network Algorithm," *Int. J. Comput. Appl.*, vol. 150, no. 3, pp. 32–38, Sep. 2016, doi: 10.5120/ijca2016911481.

[21] B. Yegnanarayana, *Artificial Neural Network*. Prentice-Hall of India Dob & Mnm, 2005.

