

A comparative performance of breast cancer classification using hyper-parameterized machine learning models

Kristoffersen Edward Mayce R. Lomboy^{1,2} and Rowell M. Hernandez^{2, 3*}

Occidental Mindoro State College, Mindoro Philippines¹

Batangas State University-Alangilan, Batangas, Philippines²

National Research Council of the Philippines³

Received: 07-July-2021; Revised: 22-September-2021; Accepted: 24-September-2021

©2021 Kristoffersen Edward Mayce R. Lomboy and Rowell M. Hernandez. This is an open access article distributed under the Creative Commons Attribution (CC BY) License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Breast cancer is the second most common cancer and has the second-highest mortality rate in women among all cancer types. Accurate cancer diagnosis plays a great part in breast cancer treatment. The application of machine learning methods in cancer classification has grown popular and has provided an accurate classification of malignant (cancerous) and benign (non-cancerous) breast cancer. This paper presents the application of three machine learning methods to classify malignant and benign breast cancer. The three machine learning methods used in this study are Support Vector Machine (SVM), Logistic Regression (LR), and Neural Network (NN) for breast cancer classification. For each machine learning method, multiple models had been tested with every model having a unique set of parameter values. This study used the breast cancer Wisconsin diagnostic (BCWD) dataset. The performance of the models is evaluated using the k-fold cross-validation technique and confusion matrix. The result shows that SVM outperformed both LR and NN in terms of classification accuracy, precision, recall, and specificity with k-fold cross validation technique. On the other hand, when the train-test split was used to validate the proposed model, the NN outperformed both SVM and LR achieving accuracy of 99.4%.

Keywords

Breast cancer, Breast cancer Wisconsin (diagnostic) data set, Support vector machines, Logistic regression, Neural network.

1. Introduction

Cancer is the result of the outgrowth of a clonal population of cells from bodily tissue [1]. It has been acknowledged that early oncogenesis, also known as carcinogenesis, can be determined by cell-intrinsic features. A way to determine is by illustrating these features of both cancer cells and tumors, which are the ability to provide their growth signals, disregard to growth-inhibitory signals, eluding of apoptosis (programmed cell death), unlimited replication, sustained vascularization and malignancy, invasion of tissues through basement membranes and capillary walls. Cancer is considered to have three developmental phases: initiation, promotion, and progression. In the initiation phase, genomic changes such as point mutations, amplification and gene deletion, and rearrangements in the chromosomes within the cancer cell occur.

These initiated cells expand clonally and survive promoting tumor development in the promotion phase. In the progression phase, significant tumor growth and metastasis are encompassed. Accumulation of genetic lesions also plays a role in the development of cancer. Inarguably, this phenomenon is present in the Initiation phase, but it may also be present in the promotion and progression phase. In the accumulation of genetic lesions, inactivation of tumor suppressor genes also occurs which provides the cells certain properties that are needed for tumorigenesis. Many types of research have established that both cancer cells and non-cancer cells with a significant amount are the recipes for the heterocellular tumor are involved in tumor development and malignancy [2, 3].

In most cases, cancer derives its name after the body part in which it emanated therefore, breast cancer is the term for the unstable growth of cells that emanated in the breast tissue. The growth results in a

*Author for correspondence

tumor that can be malignant (cancerous) or benign (non-cancerous). Malignant tumors are detrimental as they corrupt healthy cells in the body [4]. Breast cancer is the second most common cancer and has the second-highest mortality rate among all cancer types right after Lung cancer [5]. Accounting for 570,000 deaths in 2015, breast cancer is one of the most common cancers among women worldwide. Breast cancer is the second leading cause of death from cancer in women. Across the globe, of all the women diagnosed with cancer, 25% of which or over one and a half million are diagnosed with breast cancer every year. What mainly accounts for breast cancer's incurability is its metastatic nature which means it can transfer even to distant organs. Early diagnosis of breast cancer may help in having a more positive prognosis and increased survival rate [6–8].

One of the existing problems when it comes to cancer prognosis is that cancer patients are unaware of the disease until it becomes dangerous, which is problematic and unfortunate because it can be cured when detected at an early stage. Aside from early detection, another important thing to consider is an accurate classification of benign tumors for it may make cancer easier to bear by having unnecessary treatments circumvented [9]. Because of the underlying issue, countless researches have been done throughout the decades and due to the importance of cancer classification, many researchers have been studying the problems that are faced when it comes to cancer classification thus making it one of the major topics of research for cancer treatment [10, 11].

In this regard, machine learning can be thought of as a process of a computer complementing itself through collecting data and experience. It is a rapidly growing technical field. It is on the borderline of computer science and statistics and the principal idea for data science and Artificial Intelligence. Machine learning methods have a wide array of applications in different domains [12]. Machine learning methods are quite popular nowadays specifically in cancer classification because of their success in performing complex multidimensional classification tasks [13].

Despite the mentioned facts, this present undertaking aims to provide a comparative machine learning modelling that can result in a better classification of breast cancer diagnosis. This study implemented three machine learning classification methods to determine whether a patient's tumor is benign or malignant based on the available features tested from

Breast Cancer Wisconsin Diagnostic (BCWD) dataset. The main purpose of this study is to directly examine the models applied to the existing dataset from the University of California, Irvine, BCWD and determine the best model using k-fold rule validation and test-on-test data procedure. To achieve the optimal performance of the identified models derived from previous datasets, it was subjected to hyper parameterization.

This study was conducted with the following objectives:

1. To train the different hyper parameterized machine learning models and compare their performances using the BCWD dataset in terms of confusion matrix; accuracy, precision, recall, and specificity.
2. To analyse the influence of different attributes of the dataset on different machine learning models to their performances in the BCWD dataset.
3. To analyse the impact of each attribute in the BCWD dataset.
4. To propose the configuration of a fine-tuned model that will provide the optimum performance.
5. To contribute in addressing the issues in cancer classification by further exploring the use of machine learning methods.

The machine learning methods employed in this study are Support Vector Machine (SVM), Logistic Regression (LR), and Neural Network (NN). These are used as an approach to classifying benign and malignant breast cancers. The approaches were evaluated based on their accuracy, precision, recall, and specificity. As a limitation, this undertaking utilized an existing dataset from the University of California, Irvine BCWD containing diagnostic remarks only (i.e., cancer size, symmetry, location, metastatic properties

2.Related literature

2.1Literature based on BCWD dataset

In the same study of Lavanya and Rani [14] decision tree algorithms were used as a classification method and feature selection to increase the accuracy of the model. Using BCWD and other datasets, this study compared the accuracy of Classification and Regression Tree (CART) with and without feature selection and the performance of various feature selection methods. CART with feature selection proved to have increased accuracy which is at best, 94.72% compared to 92.97% without feature selection.

On the other hand, Salama et al. [15] compared the accuracy of naïve Bayes, multilayer perceptron, J48 decision tree, Sequential Minimal Optimization (SMO), and instance-based for k-Nearest Neighbor (KNN) classifiers on BCWD and two other datasets. For the BCWD experiment, SMO has the highest accuracy (97.72%).

Utomo et al. [16] showed the difference in the performance of Back Propagation Neural Networks (BPNN) and Extreme Learning Machine Neural Networks (ELMNN). The latter achieved more accuracy at 96.4%, while BPNN has only 92.1% accuracy despite having better specificity.

Obaid et al. [17] also made comparisons of the accuracy between classifiers using the BCWD dataset. In this study, three machine learning methods were used which are SVM, KNN, and decision tree. SVM proved to have the highest accuracy (98.1%) and Receiver Operating Characteristic (ROC) curve for both classifications.

Dhanya et al. [18] showed the differences in the performance of LR, naïve Bayes and random forest. This study also employed various methods of feature selection, such as sequential feature Selection, Recursive Feature Elimination (RFE), f-test and correlation. The experiment was conducted on two breast cancer dataset one of which is BCWD. Random forest with RFE algorithm yielded the best accuracy rate of 98.2% using 16 features of BCWD.

In Omondiagbe et al. [19] SVM, Artificial Neural Network (ANN) and naïve Bayes algorithms were compared and both feature extraction and feature selection were performed on BCWD. Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) were used for feature extraction while for feature selection, Correlation Based Feature Selection (CFS) and RFE were the methods of choice. SVM and NN with LDA feature extraction achieved the highest accuracy of 98.82%.

In the recent study conducted by Gupta and Garg [20] where a hyper-parameterization on six machine learning classification models were performed which are KNN, LR, decision tree, random forest, SVM with Radial Basis Function (RBF) kernel and deep learning ANN. The study showed the difference in accuracy of every unique set of hyperparameters of every machine learning classification model. Deep learning ANN achieved the highest level of accuracy with a notable score of 98.%.

In Balaraman [21] BCWD was normalized using Standard Scaler function before being utilized in comparing the performances of LR, KNN, SVM, naïve Bayes, decision tree and random forest. To evaluate the performances of each model, BCWD was split into a train and test set, of which 70% was sampled as the training set and the other 30% was sampled as the testing set. Aside from the train-test split method, the models were also evaluated using 10 fold Cross Validation technique. The study achieved 98.5% accuracy from the LR model.

In the recent research conducted by Laghmati et al. [22] which performed Neighborhood Component Analysis (NCA) to reduce the dimension of BCWD. NCA accomplished the dimension reduction by applying gradient-based optimizer ensuring the objective function to be optimal. The method of feature selection suggested that among the 32 different attributes of the dataset, only 12 were relevant. After identifying the irrelevant features, four machine learning, prediction models were hyper parameterized, which are KNN, decision tree, Binary SVM and Adaboost. The study divided the BCWD into training data (80%) and test data (20%). KNN proved to have the best accuracy (99%) and sensitivity.

2.2 Literature based on support vector machines

Durgesh and Lekha [23] showed the different results of the linear, polynomial, sigmoid, and RBF kernels which are promising. It also showed that the selection of kernel functions and the value of parameters for a specific kernel is crucial for a given amount of data.

According to Matsumoo et al. [24] advantages of SVM by comparing it with random forest using raw data for a case study of predicting Radiation Protection and toxicity in the field of drug discovery. SVM provided better performance in comparison with Random Forest where the target is a uniquely decided protein for the prediction of radiation protection function.

Chai et al. [25] demonstrated a net SVM model with L1/2 regularization and other regularization models such as Least Absolute Shrinkage and Selection Operator (LASSO) and Smoothly Clipped Absolute Deviation (SCAD). In their experiment, 12,600 genes from 102 examples from a prostate cancer dataset and 22284 genes from 107 samples from a lung cancer dataset were used. Both datasets are high dimensional and low sample sizes. The results from the experiments were promising as the Test errors are

quite low, 2.95/25 for the prostate tumor dataset and 3.88/27 for the lung cancer dataset.

2.3 Literature based on logistic regression

The study of Tirzite et al. [26] employed LR to classify lung cancer patients and non-lung cancer patients. This study used breath samples of 475 individuals, 252 of which are from patients with lung cancer and 223 are from individuals with no lung cancer. LR classified the breath samples of lung cancer patients and without lung cancer, individuals with above 90% precision for all the models used.

Moreover, the research made by Alarabeyyat and Alhanahnah. [27] showed that the feature utilization of LR is greater than back propagation NN in terms of quantity. In this study, the two classification models are utilized to detect breast cancer. The dataset used is 209 mammography images 50 of which are from patients with breast cancer. The images were then preprocessed.

In the LR, random forest, and SVM classification models are compared using age, Body Mass Index (BMI), glucose, insulin, Homeostatic Model Assessment (HOMA), leptin, adiponectin, Resistin, and Monocyte Chemoattractant Protein (MCP-1) as attributes of 166 data. This study performed analysis where multiple variations in the number of attributes are used in the prediction and showed the area under the curve, specificity, and sensitivity of the different results from the three classification models. LR provided consistent results for all variations [28].

2.4 Neural networks

Zhang et al. [29] aimed at a new method for improved cancer classification. They developed a Bimodal Distribution Removal (BDR)-Convolutional Neural Network (CNN)-graph convolutional network which is a combination of graph convolutional network and convolutional NN. This new method achieved 96.20±2.90% sensitivity, 96.00±2.31% specificity, and 96.10±1.60% accuracy. This was achieved by integrating two batch normalization and dropout and instead of using the conventional max pooling, they used rank-based stochastic pooling for an eight-layer convolutional NN then hybridized with a two-layer graph convolutional network. They used the mini-MIAS (The mini-MIAS database of mammograms 2018) database for mammograms 2018 dataset in which there are 322 single breast mammogram slices.

In case of Mohammed et al. [30] multi-fractal dimension ultrasound images and backpropagation NN used in a particular dataset consisted of 184 images of breast ultrasound images of which 112 are from normal cases and 72 tumor cases. This dataset was pre-processed by employing a median filter and adaptive weighted median filter to eliminate the speckle noise from the images. The classifier achieved 82% precision, 79% sensitivity and 85% specificity. Higa [31] applied decision tree and NN in cancer classification using the BCWD dataset with the help of weka3 software. After pre-processing the data, the experiment ended up using only 10 attributes. NN with 9,2,2 layers proved to have better classification accuracy which is 95.9% while the decision has 92.86%.

Similarly, Vijayakumar [32] showed the analysis of feed forward, recurrent, and convolutional NNs using the BCWD. As a result of the study, CNN with maximum pooling operation yielded the best performance with a precision of 95% and recall of 98.56%.

2.5 Literature analysis

BCWD has been the chosen dataset of countless researches. For most of the researches that used BCWD that is cited in this study. The SVM has the most success, such in the studies of Obaid et al. [17], Omondiagbe et al. [19], Gupta and Garg [20], Balaraman et al. [21] and the most recent study, Laghmati et al. [22]. Although not always having the best accuracy, the overall performance of SVM models regardless of the configuration, are modest at the very least. SVM performed well, even with low sample size and high-dimensional dataset like in Chai et al. [25] and Matsumoto et al. [24]. Similarly, in Durgesh and Lekha [23], SVM can have different kernel functions for various characteristics of a dataset. LR, in Tirzite et al. [26], Alarabeyyat and Alhanahnah [27], Patricio et al. [28], provided modest performance with datasets similar to BCWD in dimensionality or sample size. NN's versatility due to its multiple variations such as in Zhang et al. [29], Mohammed et al. [30], and Higa [31], is a great choice for hyper-parameterization. Similar to SVM, LR and NN models showed robustness with BCWD in the studies of Utomo et al. [16], Dhanya et al. [18], Omondiagbe et al. [19], Gupta and Garg [20] and Balaraman [21], making SVM, LR and NN, with rigorous examination, as the preferred classification model of the authors.

With the abundance of studies that aim to develop a classification model using machine learning for breast cancer classification, there have been great successes in this collective undertaking. However, even with the most recent and successful studies that tackle this issue, there is still undeniable room for improvement, rocks to be turned and pages to be filled to reach the quintessential knowledge in cancer classification. SVMs, LR, NNs and many other machines learning classification models have been proven to deliver results better by the year because of innovation and the emergence of new methods. Concepts in machine learning from data acquisition, feature selection, model hyper-parameterization to model validation, presents innumerable amalgamation and possibilities thus finding the foremost solution to a complicated issue such as breast cancer classification seems elusive. There is a need to come up with new and unique concepts not just in cancer classification, but also in other fields of studies that deal with structured data.

3.Methods

3.1Method

The *Figure 1* shows the comprehensive process adopted in the study. As an initial stage of the model development, the authors utilized the dataset from the University of California, Irvine which was subjected to a data distribution process to provide the amount of data for training and testing as well. Data distribution plays a significant role as one factor that could affect the performance of each model, whether the model utilized the standard setting and even on

their parameterization state. SVMs, LR and NNs are the machine learning models used in this study. K (10)-fold rule was used to evaluate the initial training of the model after the initial training. The statistical evaluation conducted during the initial training period yielded relatively higher results than most of the previous studies, but shows lower results compared to the most recent works. However, when the authors applied the cross-validation technique using the principle of test-on-test data, the evaluation metrics resulted in a superior rating, which made the phase the most notable contribution and important process in this work, as it is the first to calibrate such testing procedure to BCWD dataset with respect to the hyper-parameterization done in the models. The result shows a noticeable improvement in terms of accuracy performance of the model as compared to the previous studies mentioned in this paper, placing this study at the forefront in terms of classifying and predicting the benign and malignant diagnosis in the concerned dataset.

Identifying the impact of every attribute was also tested and went to a series of model training. The authors performed a multiple case analysis. The result shows how attributes in the data set affected the evaluation metrics of testing on the test data. Finally, the results generated from the series of experiments in this study can be compared to the most recent paper related to the utilization of different breast cancer datasets and BCWD dataset.

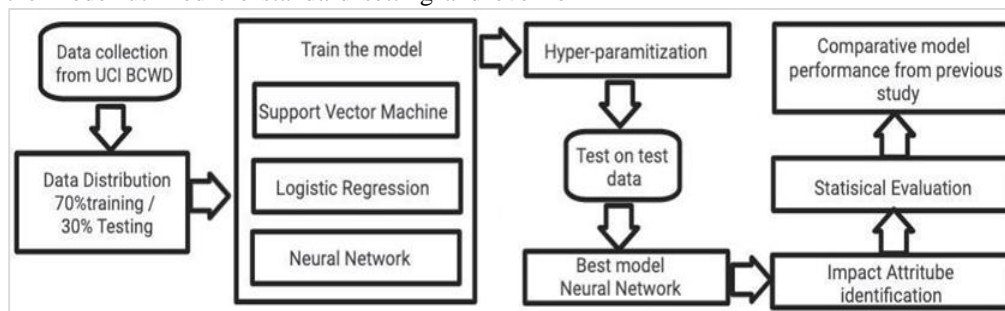


Figure 1 Process flow of the proposed model

3.2Dataset

This study used the BCWD. This particular dataset is obtained from the University of California, Irvine (UCI) machine learning repository [33] which is available to the public. It has 569 instances and 32 attributes including the ID number and the class label (diagnosis). The other 30 attributes are from ten real-valued features which are: radius, texture, perimeter, area, smoothness, compactness, concavity, concave

points, symmetry, and fractal dimension having the mean, Standard Error (SE), and the largest (worst) value computed for each of them. For a better description, the radius of a tumor will be represented by three attributes in the dataset which are: radius_mean, radius_se, and radius_worst. The ten real-valued features are computed from a digitized image of a Fine Needle Aspirate (FNA) of a breast mass describing the characteristics of the cell nuclei

present in the image. The range of the ten real-valued features and their mean, SE, and worst is shown in

Table 1.

Table 1 Range of attributes

Attribute	Mean	SE	Worst
Radius	6.98 - 28.11	0.11 - 2.87	7.93 - 36.04
Texture	9.71 - 39.28	0.36 - 4.89	12.02 - 49.54Z
Perimeter	43.79 - 188.50	0.76 - 21.98	50.41 - 251.20
Area	143.50 - 2501.0	6.80 - 542.20	185.20 - 4254.0
Smoothness	0.05 - 0.16	0.0 - 0.03	0.07 - 0.22
Compactness	0.02 - 0.35	0.0 - 0.14	0.03 - 1.06
Concavity	0.0 - 0.43	0.0 - 0.40	0.0 - 1.25
Concave Points	0.0 - 0.20	0.0 - 0.05	0.0 - 0.29
Symmetry	0.11 - 0.30	0.01 - 0.08	0.16 - 0.66
Fractal Dimension	0.05 - 0.10	0.0 - 0.03	0.06 - 0.21

Of the 569 instances, 357 (62.74%) are benign tumors and 212 (37.26%) are malignant tumors, which are indicated as B for benign and M for malignant in the diagnosis attribute. Figure 2 shows the distribution of the diagnosis attribute.

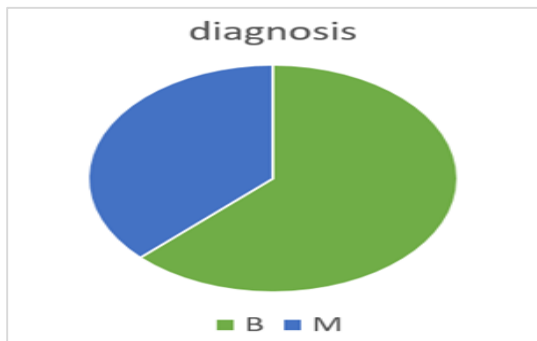


Figure 2 Distribution of the classes

3.3 Classification and prediction

To classify malignant and benign tumors, this study used three machine learning methods which are SVM, NN, and LR. These methods have provided great classification accuracy in research on cancer classification [34–36].

3.3.1 Support vector machine

Since the '90s, SVM has been popular especially when it comes to tasks like classification and regression. As a still-growing field, SVM's applications have been explored such as in image processing, speech processing, time-series prediction, data mining, power systems, process control, automotive, and more. All of the mentioned applications are due to SVM's advantages [37].

SVM has an advantage when it comes to modelling nonlinear class boundaries, avoiding overfitting, quadratic optimization problems, robust control of the complexity of decision rule, and produces lower

prediction error compared to artificial NNs. But it also presents some disadvantages such as the rate of training, the non-linear distinction of training data, and complex algorithms [38, 39].

SVM works by classifying different classes from one another through a multi-dimensional hyperplane. These classes are separated based on hidden patterns among the different features. The hyperplane is then used to identify with the most probability, the respective classes the data belongs to. SVM balances two complementary objectives, (1) optimizing its accuracy and (2) ensuring the maximum reproducibility [40]. As one of the most popular classification methods in machine learning, SVM has also been useful in the diagnosis and prognosis of cancer [41]. This method has been used to classify the tumor as benign or malignant based on the characteristics of a tumor [42].

When it comes to performance, SVMs are greatly dependent on the kernel functions used. Since SVMs classify data that are distinguished to be separated by a line or hyperplane [43] and the vectors of the features much like in the BCWD are on the other hand, not linearly separable, the use of kernel functions is necessary to solve this. This study employs four different kernel functions to identify the one that will provide the best performance. The kernel functions employed are linear, polynomial, RBF, and sigmoid. The kernel type and their functions used in this study are shown in Table 2.

Table 2 Kernel type and their Functions

Kernel type	Functions
Linear	x·y
Polynomial	(0.02 x·y + 1.0) ³
RBF	exp(-0.03 x-y ²)
Sigmoid	tanh(0.01 x·y + 1.0)

Other than the choice for a kernel, other parameters that SVMs performance rely on are Cost (C) and Regression Loss Epsilon (ϵ) parameters. C draws the line between the complexity of the model and the tolerance for deviation higher than the regression loss [44]. The value of C leaning towards either can have a significant effect on the performance. If the C parameter is set to a high or too high value, the accuracy is very high in the training stage, but very low in the testing stage, and on the other hand, if it is set too low the model will be rendered useless because of the unsatisfactory accuracy rate which greatly defeats the purpose of itself as a classifier [45].

In this study, different values for parameter C have been tested across all the kernel functions to identify which configuration of parameters would provide the optimum result, while ϵ is set to 0.10 as it does not significantly affect the performance of SVM in this study. For the optimization parameters, numerical tolerance is set to 0.001, and Iteration Limit is set to 1000 as this configuration provides the best output for this experiment.

3.3.2 Logistic regression

LR estimates the probability of an observation to identify its respective class by applying a linear model. The probability is identified by using a linear relation and the maximum likelihood estimation method.

A final model is created by eliminating non-essential variables and the essential variables are added. The final model will exclusively have the relevant variables. This selection of variables is achieved by employing an iterative algorithm to compare a model to any of its sub-model and top models [46]. LR is a regression model that predicts the probability that an entry belongs to the given class using a regression model [47]. It suggests that a linear function is accompanied by the data [48].

LR has been proven to be one of the most effective machine learning methods for cancer classification specifically when using BCWD [49]. This is due to its advantage when it comes to model regularization and feature correlation constraints [50].

LR, if not regularized, is an unbound and ungoverned convex optimization problem having a ceaselessly distinct function. This is why regularization is an essential task for this model, to avoid overfitting more so if there are a limited number of data with plenty of features. There are different ways to

regularize this model. One particularly common way is the LASSO or L1 regularization has proven to provide good generalization performance when data present irrelevant features [51–53]. L1 rallies the sum of the absolute values of the features to be small with the use of the penalty term while another regularization method, ridge or L2 regularization deals with the sum of the squares of the features by rallying it to be small [54, 55]. The L2 regularization method is known to outperform L1 when it comes to data with spatially global features [56] and has been used in research dealing with breast cancer classification [57]. Both LASSO and ridge regularization has been used in this study because of the nature of BCWD. As previously discussed, parameter C dictates the complexity of the model. This parameter, in LR, is the inverse of the regularization strength. The value of parameter C will have a certain impact on the performance of the model [58]. In the pursuit of the optimal performance of the model, this study explores the impact of parameter C on both LASSO and ridge regularizations.

3.3.3 Neural network

NNs are developed for a wide array of applications, especially in computational problems in cognition, pattern recognition, and decision making [59]. It has been popular in cancer classification with promising results [36].

Figure 3 shows the NNs are simply pre-established and interlinked neurons that compute information. NNs are generally known to be modelled after biological neurons of the human brain [30]. These links have weights or strengths where the network stores information that is assimilated by the network [59]. NNs use an activation function to determine the output of each neuron. Just like the human brain, NNs learn from experiences. If fed with an unknown input, it could derive from past experiences and form a conclusion [60]. Since NNs are made of neurons, it could commit to memory the information about the data making it adaptive. They are also “naturally massively parallel” though sometimes they are embedded in software on ordinary computers, parallel implementations suit NNs [31].

A Single-Layer Perceptron (SLP) NNs has only two input and output layers and are considered the simplest form of NNs. The problem with SLPs is that they could not efficiently deal with nonlinearly distinguishable data. This problem has been solved by Multi-Layer Perceptron Neural Network (MLPNN) which are great when it comes to non-

linearly separable data. MLPNNs employ one or more hidden layers in their network as shown in *Figure 3* thus solving the problem with SLPs [61].

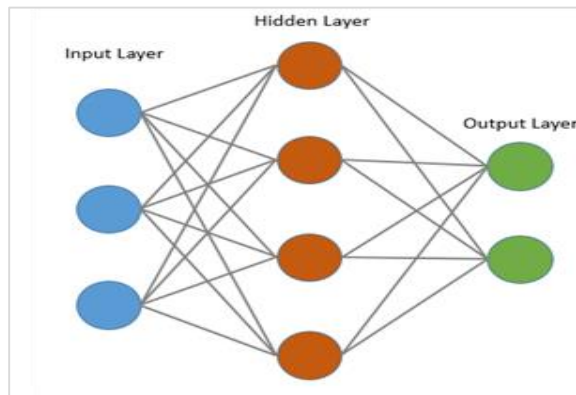


Figure 3 Multi-layer perceptron common structure of NNs

The number of hidden neurons depends on the training data set while the quantity of input layer nodes is quantified according to the selected factor in the dataset. The output layer serves as the configuration of the model and evidently, the hidden layers are for computation. Every node in the input layer is connected to every node of the hidden layer/s which are then connected to every node in the output layer. The training, implementation of Multi-Layer Perceptron (MLP) is in two stages, namely the forward and backward propagations using the back-propagation algorithm [62, 63].

A NN must assimilate any data and any arbitrary complex function that is crucial in mapping the inputs to the outputs and needs to piece together complex and arbitrary data to gain knowledge otherwise, it would only be a LR model. Although linear regression has its merit, it has limited complex mapping recognition. That is why a NN needs an activation function. With an activation function, NNs can recognize non-linear mappings from inputs to outputs [64, 65].

MLPNNs can be conceived as an optimization problem that every solution can be viewed as an architecture. Training error and the size of the network may be viewed as a function of the cost measure. To search for the optimal architecture, an optimization technique can be applied. Thus, to generate an efficient network, the optimization of the architectures and weights can be applied [66, 67].

This study used the MLPNN model with the back-propagation algorithm for training. For activation function, four have been tested for the hidden layer and the results of each are compared. These activation functions are identity, logistic, tanh and ReLu. For the weight optimization, three different solvers have been tested which are Limited-Memory Broyden Fletcher Goldfarb Shanno with Box constraint (L-BFGS-B), Stochastic Gradient Descent (SGD), and Adaptive Moment Estimation (ADAM) algorithm. The following hyper-parameters are constant throughout the experiment. The L2 penalty (regularization term) parameter (Alpha) is set to 0.01. The maximal number of iterations is set to 100 and the number of neurons in hidden layers is set to 100.

3.4 Model validation

Statistical validation is an essential aspect of model evaluation. After training, the three models' performance is then evaluated using stratified 10-fold cross-validation. Cross-Validation is a statistical technique commonly used for the evaluation of classification models [68]. This is done by splitting the dataset into k number of folds and then the k-1 fold is set aside to be the test data and the remaining folds are fed to the model/s to train with. This process is repeated until all of the folds have been used as a training set and then the average of all the performances' output is calculated [31] [69]. Cross-Validation is a great way to achieve higher accurate results when dealing with a limited number of data [70]. A stratified cross-validation is a type of cross-validation where the size of the folds is nearly proportional to the size of the output data [71], this process was applied to achieve the optimum performance of the model in the conduct of this study.

In addition, another method of validating the model is the train-test split, a simple and well-founded approach. Cross-Validation technique is often the method of choice to validate machine learning models in clinical studies most likely in studies that deal with small-sized datasets. Although cross-validation techniques are a great way to achieve higher accurate results when dealing with a limited amount of data. This technique has the tendency to generate fairly unclear results in clinical studies. This is due to its core process of having test data already incorporated into the training process, enabling the technique to produce biased performance estimates [72–75].

Train-test split partitions the dataset into training set and test set. Only the training set is learned from the model ensuring the model to be unbiased when predicting the test set. Previous studies that aimed to predict benign and malignant tumors had found better performance of a prediction model using train-test split compared to the cross-validation method [41, 76, 49].

In this study, a stratified 10-fold cross-validation method was used after hyper parameterization of each model as the goal is to develop a versatile and inclusive model providing the best overall performance. Since cross-validation evaluates the model on multiple test sets, cross-validation conveys a much more precise model performance [21] suitable to the purpose of hyper parameterization. Once the optimum configuration of parameters of each model has been identified and achieved, the proposed model is validated using the train-test split method. The dataset is partitioned deterministically, 70% of which serves as the train set and the remaining 30% serves as the test set. Train-test split validation method was applied to the proposed models for testing the models blinded from the test data provides an unbiased performance simulating a real-world scenario [74].

3.5 Confusion matrix

Confusion matrix is a tool to show the outcomes and predicted classes and their relationship [15]. It simply contains the information of the actual classification of the training dataset and the predicted classification of the classification model [4]. The performance of each model is calculated using the elements present in confusion matrix. These elements are the True Positives (TP) which are the number of correctly predicted positive classification, False Negatives (FN) which are the number of incorrectly predicted positive classification, False Positives (FP) which are the number of incorrectly predicted negative classification and the True Negatives (TN) which are the number of correctly predicted negative classification [71] as shown in *Table 3*.

Table 3 Elements in the confusion matrix

		Predicted	
		Positive	Negative
Actual	Positive	TP	FN
	Negative	FP	TN

With the elements of the confusion matrix, we can then evaluate the performance of every model. We can calculate the classification accuracy, precision,

recall, and specificity of the classification models [4, 71, 76] as shown in *Table 4*.

Classification accuracy refers to how often the classification model is correct. Precision refers to how often it is correct when it predicts positive. Recall refers to how often it predicts a positive if the actual classification is positive. While specificity refers to how often it predicts a negative when the actual classification is negative.

Table 4 Performance measure

Measures	Derivations
Classification accuracy	$TP + TN / TP + TN + FN + FP$
Precision	$TP / FP + TP$
Recall	$TP / FN + TP$
Specificity	$TN / TN + FP$

4. Results

4.1 Model training and hyper-parameterization

This section will discuss the performance of the different classification models with 10-fold cross validation technique as discussed in the methods section

4.1.1 SVM classifier performance

The SVM model had been tested with different hyper-parameter configurations and then validated using 10-fold cross validation. There are two hyper-parameters of the SVM model that the authors had rigorously tuned until the optimum accuracy rate had been achieved. First is the kernel function in which there are four: linear, polynomial, RBF and sigmoid kernels. Second is the cost value (C) which ranges from 0.1, the strongest to 10, the weakest regularization. The range of C had been tested with every kernel function. *Figure 4* shows the significant difference in the performance of the SVM model with a linear kernel against an increasing value of C from 0.1 to 1.

There is a significant difference in the performance of the SVM model with a polynomial kernel against an increasing value of C from 0.2 to 1.1 as shown in *Figure 5*. The significant difference in the performance of the SVM model with an RBF kernel against an increasing value of C from 1.8 to 2.7 as shown in *Figure 6*.

Figure 7 shows the significant difference in the performance of the SVM model with a sigmoid kernel against an increasing value of C from 0.1 to 1. After the ideal cost values for every kernel has been identified, the authors evaluated the performance of the model having each kernel type with their

respective ideal cost values as parameters in terms of accuracy, precision. Recall and specificity are shown in *Table 5*.

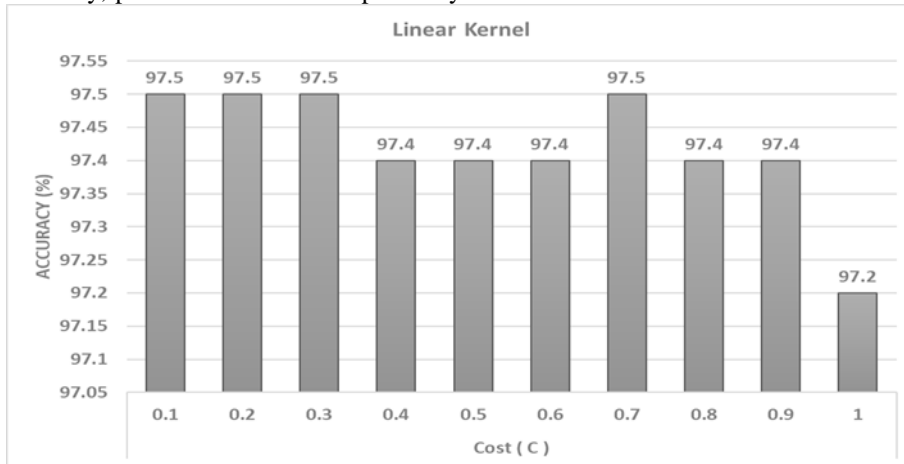


Figure 4 Accuracy of SVM with linear kernel and different cost value

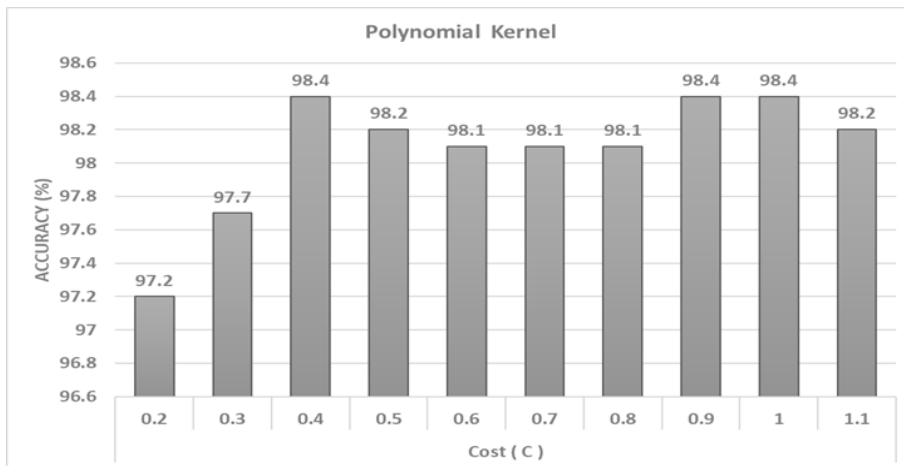


Figure 5 Accuracy of SVM with polynomial kernel and different cost value

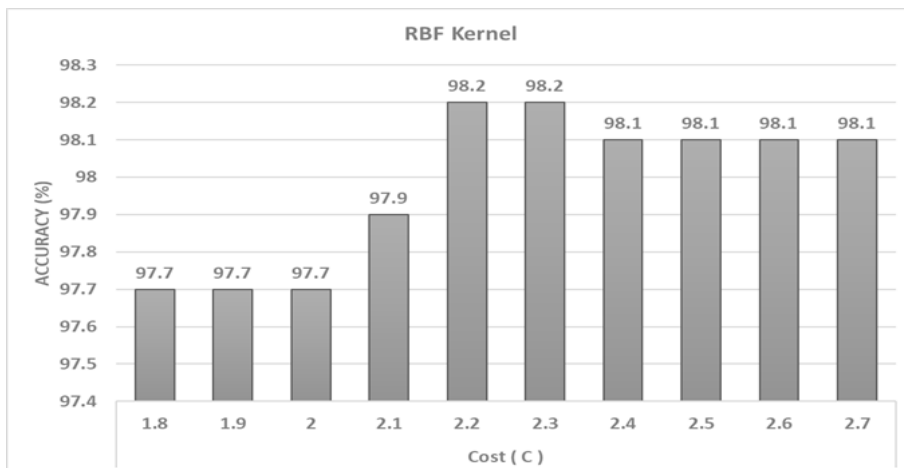


Figure 6 Accuracy of SVM with RBF kernel and different cost value

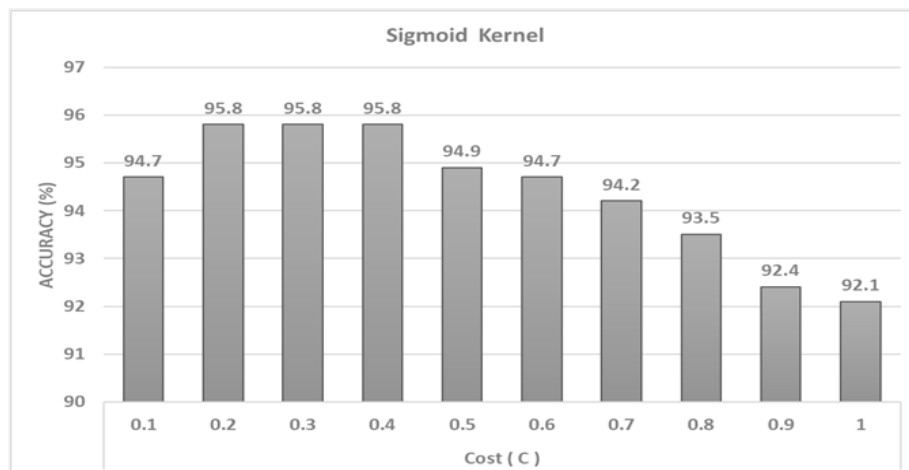


Figure 7 Accuracy of SVM with sigmoid kernel and different cost values

Table 5 Performances of different SVM models

Kernel	C	Accuracy (%)	Precision (%)	Recall (%)	Specificity (%)
Linear	0.1	97.5	97	99.2	94.8
Linear	0.2	97.5	97	99.2	94.8
Linear	0.3	97.5	97	99.2	94.8
Linear	0.7	97.5	97.5	98.6	95.8
Polynomial	0.4	98.4	97.8	99.7	96.2
Polynomial	0.9	98.4	97.8	99.7	96.2
Polynomial	1	98.4	97.8	99.7	96.2
RBF	2.2	98.2	98.1	99.2	96.7
RBF	2.3	98.2	98.1	99.2	96.7
Sigmoid	0.2	95.8	94.9	98.6	91
Sigmoid	0.3	95.8	94.9	98.6	91

4.1.2 Logistic regression classifier performance

The LR model has been tested with different hyper-parameter configurations and then validated using 10-fold cross validation. There are two hyper-parameters of the LR model that had rigorously tuned until the optimum accuracy rate had been achieved. The first hyper-parameter that the authors have considered is the regularization type, of which there are two that have been considered: LASSO (L1) and ridge (L2). The other hyper-parameter considered is the regularization strength or cost value (C). Evidently, both L1 and L2 regularization types with cost values ranging from .001, the strongest to 1000, the weakest regularization strength.

Figure 8 shows the significant differences in the performance of the LR model with L1 regularization with different cost values from 50 to 500.

After the ideal regularization cost value for the LR model with L1 regularization type had been identified, the LR model was evaluated having L1 regularization and the respective ideal cost value as a parameter in terms of accuracy, precision. Recall and specificity are shown in Table 6.

Figure 9 shows the significant differences in the performance of the LR model with L2 regularization with different cost values from .001 to 1000.

After the ideal regularization cost value for the LR model with L2 regularization type had been identified, the LR model was evaluated having L2 regularization and the respective ideal cost value as a parameter in terms of accuracy, precision. Recall and specificity are shown in Table 7.

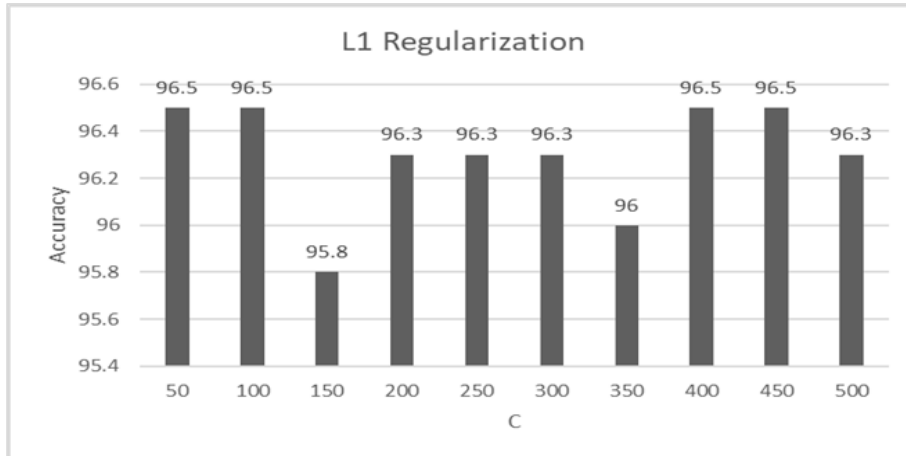


Figure 8 Accuracy of LR model with L1 regularization and different regularization strength

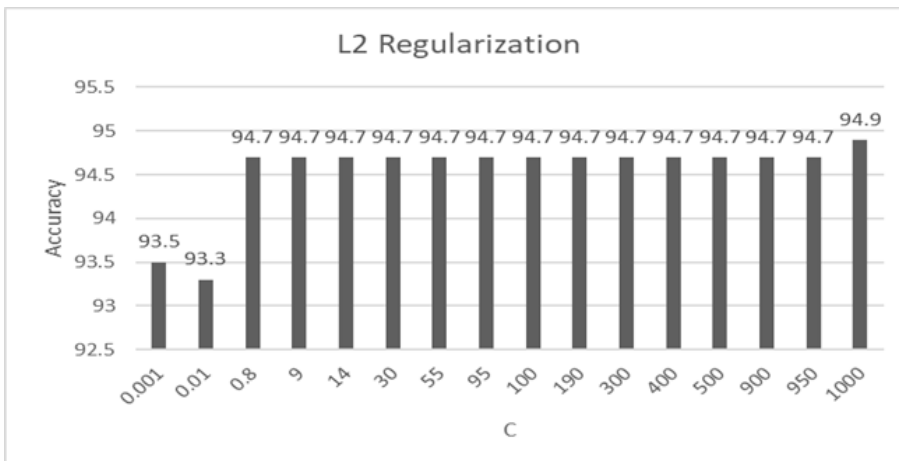


Figure 9 Accuracy of LR model with L2 regularization and different regularization strength

Table 6 Performances of LR models with L1 regularization

C	Accuracy (%)	Precision (%)	Recall (%)	Specificity (%)
50	96.5	96.9	97.5	94.8
100	96.5	96.7	97.8	94.3
150	95.8	96.4	96.9	93.9
200	96.3	96.7	97.5	94.3
250	96.3	96.9	97.2	94.8
300	96.3	96.9	97.2	94.8
350	96	96.9	96.6	94.8
400	96.5	97.2	97.2	95.3
450	96.5	97.2	97.2	95.3
500	96.3	97.2	96.9	95.3

Table 7 Performances of LR models with L2 regularization

C	Accuracy (%)	Precision (%)	Recall (%)	Specificity (%)
0.001	93.5	92.8	97.2	87.3
0.01	93.3	92.8	96.9	87.3

C	Accuracy (%)	Precision (%)	Recall (%)	Specificity (%)
0.8	94.7	95.3	96.4	92
9	94.7	95.5	96.1	92.5
14	94.7	95.3	96.4	92
30	94.7	95	96.6	91.5
55	94.7	95	96.6	91.5
95	94.7	95.3	96.4	92
100	94.7	94.8	96.9	91
190	94.7	95.5	96.1	92.5
300	94.7	95	96.6	91.5
400	94.7	95	96.6	91.5
500	94.7	95.5	96.1	92.5
900	94.7	95.5	96.1	92.5
950	94.7	95.3	96.4	92
1000	94.9	95.3	96.6	92

4.1.3 Neural Network classifier performance

The MLPNN with the back-propagation model was trained in the same manner as SVM and LR, 10-fold cross validation was performed and two hyper-parameters have been considered: activation function

and weight optimization. Identity, logistic, tanh and ReLu as activation functions were considered. L-BFGS-B, SGD and ADAM as weight optimization methods. The rest of the parameters for MLPNN is discussed in the method section. Every possible combination of the activation functions and weight optimization methods had been tested, resulting in 12 different hyper-parameter configurations which are treated as individual models of MLPNN. For each model, 10-fold cross validation was performed and is evaluated in terms of accuracy, precision, recall and specificity. *Figure 10* shows the accuracy of the models with identity, logistic, tanh and ReLu activation function and for each activation functions, L-BFGS-B, SGD and ADAM weight optimization was applied.

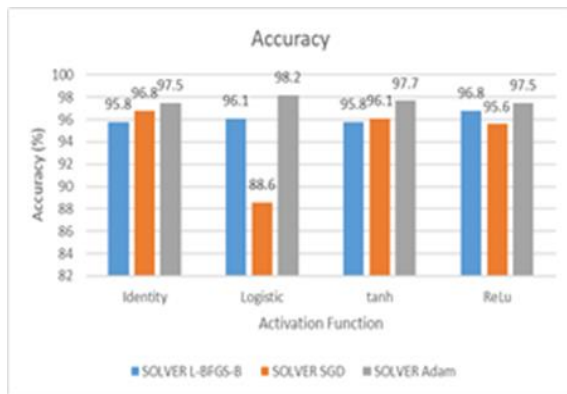


Figure 10 Accuracy of multi-layer perceptron models with weight optimization solver

Figure 11 shows the precision of the models with identity, logistic, tanh and ReLu activation function and for each activation functions, L-BFGS-B, SGD and ADAM weight optimization was applied.

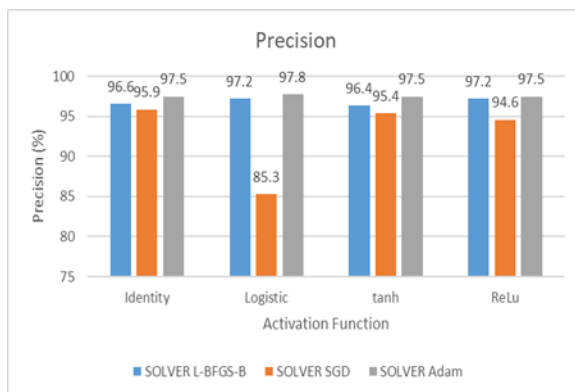


Figure 11 Precision of multi-layer perceptron models with different weight optimization solvers

Figure 12 shows the recall of the models with identity, logistic, tanh and ReLu activation function and for each activation functions, L-BFGS-B, SGD and ADAM weight optimization was applied.

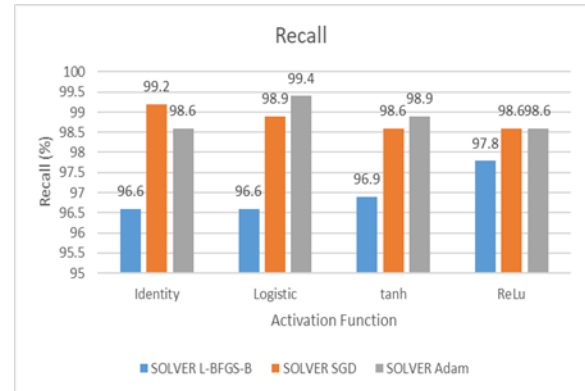


Figure 12 Recall of multi-layer perceptron models with different weight optimization solvers

Figure 13 shows the specificity of the models with identity, logistic, tanh and ReLu activation function and for each activation functions, L-BFGS-B, SGD and ADAM weight optimization was applied.

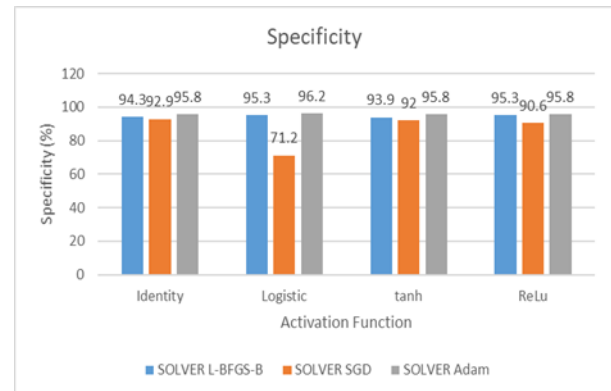


Figure 13 Specificity of multi-layer perceptron models with different weight optimization solvers

4.2 Comparative performance of models in the initial training

Results in *Table 8* generated from this experiment, it is notable that the result in terms of accuracy is lower than the most recent study who used the BCWD data set. With this regard, the authors perform another evaluation metric or procedure that employs the test-on-test data which. This process provided a notable result that surpasses the accuracy rating of all studies prior. Using hyper-parameterization, the authors have identified the best configuration of hyper-parameters for each model that would be implemented to develop a proposed model.

Table 8 shows the performances of the proposed model for SVM, LR and NN in terms of Area Under ROC curve (AUC), accuracy, precision, recall and specificity using 10-fold cross validation.

4.3Comparative performance of models in the train-test splits validations

Table 8 Result from 10-fold cross validation

Proposed model	AUC	Accuracy	Precision	Recall	Specificity
SVM	99.7%	98.4%	97.8%	99.7%	96.2%
LR	99.3%	96.5%	96.9%	97.5%	94.8%
NN	99.55	98.25	97.85	99.45	96.2%

Table 9 Result from the test-on-test data

Proposed model	AUC	Accuracy	Precision	Recall	Specificity
SVM	99.8%	98.8%	98.2%	100%	96.8%
LR	99.6%	97.7%	99.1%	97.2%	98.4%
NN	99.9%	99.4%	99.1%	100%	98.4%

After the training and the process of hyper-parameterization of three machine learning model using the BCWD and 10-fold cross-validation technique to propose an architecture of hyper-parameters for prediction models that would achieve optimum results, and training the proposed hyper-parameterized models with 70%-30% train-test split, we had come up with a proposed model: MLPNN with back propagation with a parameter architecture show in Table 10.

By validating the proposed models with 70%-30% train-test split method, it achieved significantly better overall performances as shown in Table 9. The SVM, LR, NN proposed models achieved better results in terms of all measures, including AUC with train-test split compared to with 10-fold cross-validation.

Table 10 Neural network hyper-parameterization

Architecture parameters	Value
Neurons in hidden layers	100
Activation	Logistic
Solver	ADAM
Regularization	0.01
Maximal number of iterations	100

By separating the training and test data the proposed NN model was able to deliver unbiased predictions and handled overfitting better as shown in Figure 14.

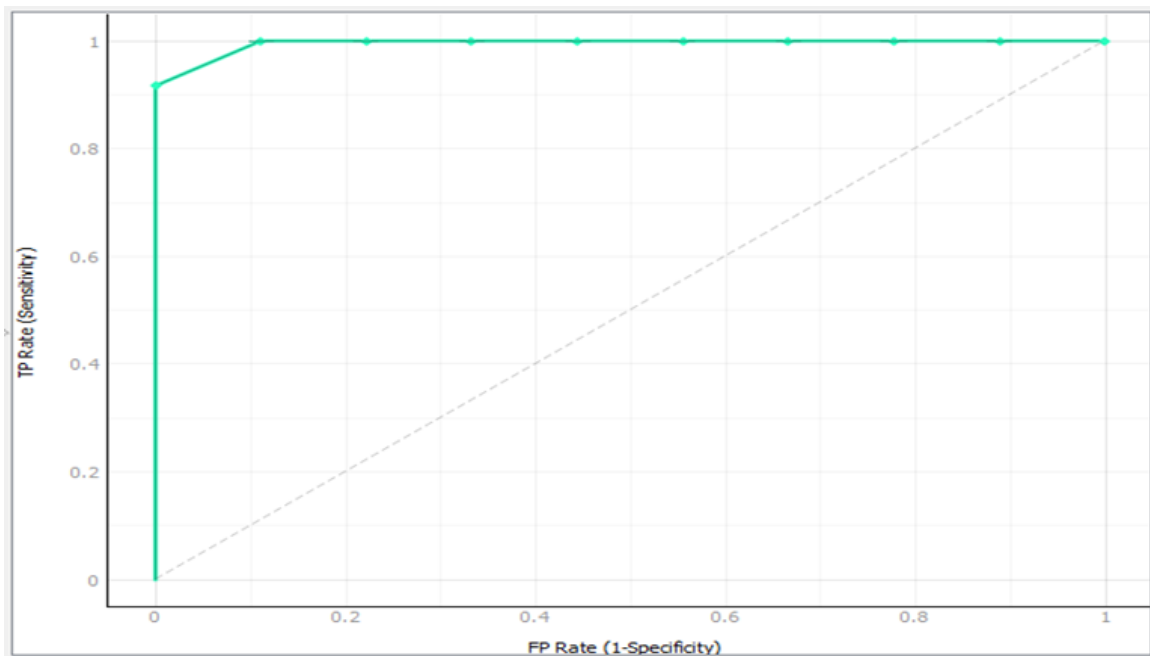


Figure 14 ROC for classification for proposed NN model

4.4 Impact of attributes in the model performance

The attributes of all the data sets, whether structured, semi-structured or unstructured, play a vital role in the training of the model. Hence, providing the right and appropriate significant attributes can make an important contribution and impact in every experiment. A series of experiments were provided

that involved cases of how attributes can affect the final output in terms of accuracy. *Figure 15* shows the cases which involve a diminishing factor of attribute that deduct 1 attribute at a time and re-train and test the model again. This process continues until only one attribute is left in the dataset.

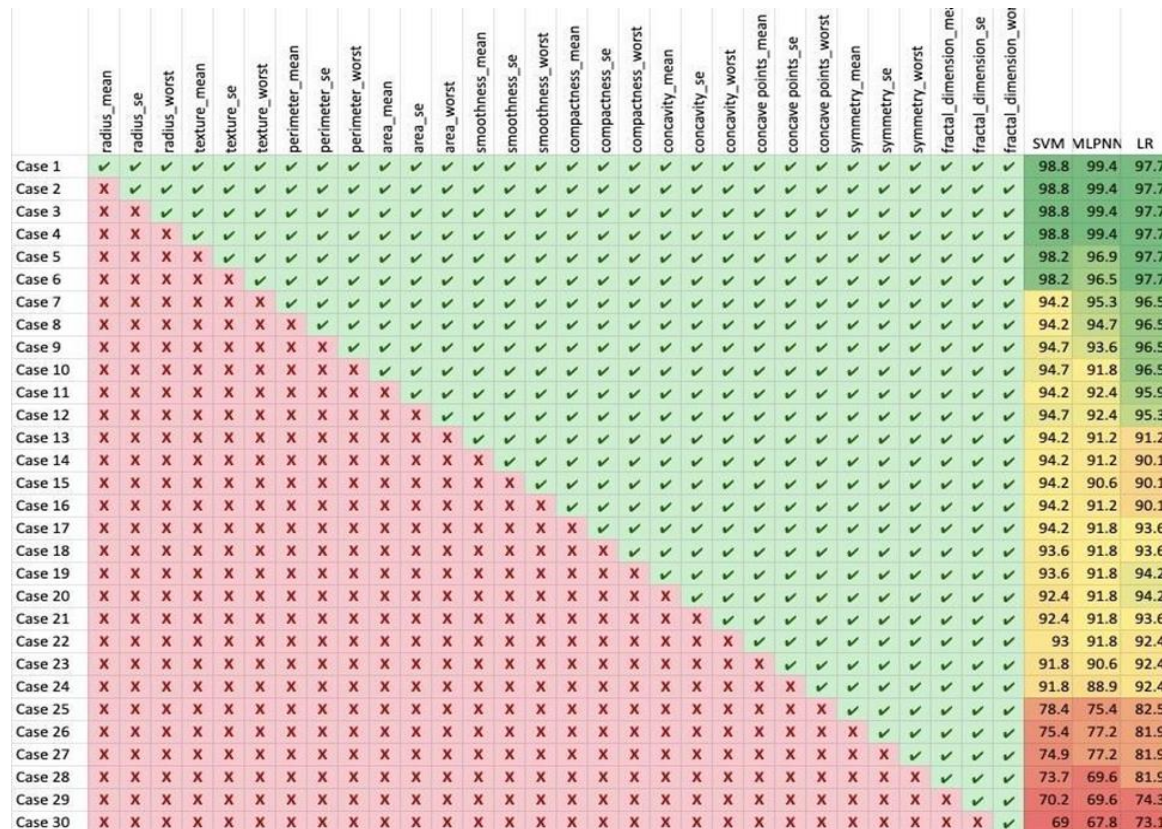


Figure 15 Impact of attributes in model performance per case after attribute deduction

4.5 Comparison from previous study

Table 11 shows the most notable studies involving the BCWD Dataset. Each paper presents a unique contribution to the field to breast cancer classification using the different model training. Every study has

provided a well explained and well discussed process of how they are able to generate superior findings at the time of writing their individual paper.

Table 11 Results from previous researches that used BCWD

Authors	Method	Dataset Used	Performance Metric		
			Accuracy	Sensitivity	Specificity
Lavanya et al. [14]	CART-with feature selection	BCWD	94.72	N/A	N/A
	CART-without feature selection	BCWD	92.9	N/A	N/A
Salama et al. [15]	SMO	BCWD	97.71	N/A	N/A
	Instance-based KNN	BCWD	95.95	N/A	N/A
Utomo et al. [16]	BPANN	BCWD	92.1	84.3	98
	ELMNN	BCWD	96.4	94.8	97.4
Obaid et al. [17]	SVM-linear kernel	BCWD	97.9	N/A	N/A
	SVM-quadratic kernel	BCWD	98.1	N/A	N/A
	SVM-cubic kernel	BCWD	97	N/A	N/A

Authors	Method	Dataset Used	Performance Metric		
			Accuracy	Sensitivity	Specificity
Dhanya et al. [18]	Random Forest with RFE	BCWD	98.24	N/A	N/A
	Naïve Bayes with SFS	BCWD	98.24	N/A	N/A
	LR with SFS	BCWD	96.49	N/A	N/A
Omondiagbe et al. [19]	SVM with LDA	BCWD	98.82	98.41	99.07
	NN with LDA	BCWD	98.82	98.41	99.07
Gupta et al. [20]	Naive Bayes with LDA	BCWD	98.24	96.83	99.07
	SVM	BCWD	97.2	N/A	N/A
	DL_ANN	BCWD	98.24	N/A	N/A
Balaraman et al. [21]	LR - train-test split	BCWD	97.66	N/A	N/A
	LR - 10-fold cross-validation	BCWD	98.5	N/A	N/A
	SVM - train-test split	BCWD	95.9	N/A	N/A
	SVM - 10-fold cross-validation	BCWD	97.49	N/A	N/A
Laghmati et al. [22]	KNN	BCWD	99.12	100	98.88
	SVM	BCWD	94.74	98.86	80.77
Present work	Multilayer Perceptron with Backpropagation NN	BCWD	99.4	100	98.4
	SVM	BCWD	98.8	100	96.8
	LR	BCWD	97.7	97.2	98.4

5. Discussion

In terms of the performance of different SVM models. Each model has a different kernel function and is tested with different values of Cost (C) to find out which configuration of parameters for the SVM model will produce the optimal performance. The accuracy of all the models has been compared then the models with the highest classification accuracy are compared based on their precision, recall, and specificity.

The model training and hyper-parameterization is essential in the initial stage of identifying the best model in this study. The SVM with linear kernel performed best with the lowest C values. With both C=0.1, C=0.2, C=0.3, and C=0.7 linear kernels achieve the best accuracy which is 97.5% shown in *Figure 4* while *figure 5* shows the SVM model with the polynomial kernel performing best with C=0.4, C=0.9, and C=1 having the accuracy at 98.4%.

In this experiment, for all the SVM models, models with a polynomial kernel achieved the highest accuracy which is 98.4%. Of all the models that achieved the highest accuracy from their respective kernel group, RBF achieved the highest precision at 98.1% and the highest specificity at 96.7%, while polynomials also achieved the highest recall which is 99.7% shown in *Table 5*. It also shows that while having the same linear kernel and accuracy of 97.5%, the model with C=0.7 deviates from its kernel group when it comes to precision, recall, and specificity while the rest of all the models do not. The accuracy

of the model with RBF increases with the value of C until C =2.2 and C=2.3 where the accuracy is at its peak which is 98.2% then gradually declines. Hence, the sigmoid kernel reached its peak at C=0.2 with the accuracy of 95.8% until C=0.4 then continuously declined its performance.

Likewise, in LR, the models with the L1 regularization type achieved better accuracy with weak regularization (high value of C). The best accuracy comes with C=50, C=100, C=400, and C=450 reflected in *figure 8*. The highest precision for the L1 models is 97.2%, which is achieved by C=400, C=450, and C=500. The highest recall is 97.8%, which is achieved by C=100 while the model with the highest specificity is also C=400, C=450, and C=500 with a specificity of 95.3% as shown in *Table 6*. Just like L1 models, the models with L2 regularization achieved better classification accuracy with a weak regularization. The best accuracy is 94.9%, which comes from the model with the weakest regularization, which is C=1000 as shown in *Figure 8*. It also shows that the second to the best accuracy rate, which is 94.7% is achieved by models in a wide range of C values from .8 to 950. The highest precision for the L2 models is 95.5%, which is achieved by C=9, C=190, C=500, and C=900.

The highest recall is 97.2%, which is achieved by the model with the strongest regularization, which is C=.001 while the best precision is 92.5%, which is achieved by C=9, C=190, C=500, and C= 900. As shown in *Table 7*. For LR models, the best accuracy

is achieved with L1 regularization and C value of 50, 100, 400, and 450. Of all the models with the best accuracies including the models with L2 regularization, the ones that achieved the best precision, recall and specificity are achieved with L1 regularization exclusively. In terms of NN, the best classification accuracy which is 98.2% is achieved by the model which uses a logistic activation and ADAM solver. Although having the best accuracy, logistic function in term of accuracy, have not been compatible with all of the solvers specifically with SGD as with it, provided the least accuracy of all the models which is 88.6% having 9.6% difference with ADAM and 7.5% with L-BFGS-B. The rest of the activation functions provided small differences in accuracy with all the solvers shown in *Figure 10*. In terms of precision, the same model with Logistic activation and ADAM solver yielded the best which is 97.8%. The same thing can be said about logistics and SGD solvers. It provided the least precision as shown in *Figure 11*. The optimum recall for the models is 99.4%, which is provided by still the same model (with logistic and ADAM). L-BFGS-B with all the activation functions provided the least recall shown in *Figure 12*.

Surprisingly, with logistic and ADAM, the model provided the best Specificity at 96.2%, while the model with logistic activation and SGD solver provided the list at 71.2% as shown in *Figure 13*. The model with the logistic activation function and ADAM weight optimization provided the best performance in terms of classification accuracy, precision, recall, and specificity.

With regards to the comparative performance of the model in the initial training, the proposed model for SVM achieved the best AUC, accuracy, precision, recall and specificity having 99.7% AUC, 98.4% accuracy, 97.8% precision, 99.7% recall and 96.2% specificity. The proposed model for NN is not far behind SVM achieving 99.55% AUC, 98.25% accuracy, 97.85% precision, 99.45 % recall and 96.2% specificity, matching the precision and specificity of the proposed SVM model. LR achieved 99.3% AUC, 96.5% accuracy, 96.9% precision, 97.5% recall and 94.8% specificity.

Although having fairly impressive performances, the proposed models, though comparable, are still not able to exceed or even match the performances of some of the most recent related studies. With the goal of pushing the boundaries of machine learning application in breast cancer classification, the authors

opted to try other various approaches to better the performances of the proposed models.

Likewise, in the comparative performance of models in the train-test split validation shown in *Table 9*, the most notable difference would be the overall performance of the proposed NN model generating the highest accuracy achieved in this study which is 99.4%. With 10-fold cross-validation, the proposed NN model falls second after the proposed SVM model in accuracy. The proposed SVM model achieved an AUC of 99.8%, accuracy of 98.8%, precision of 98.2%, 100% recall and specificity of 96.8% having 0.1% difference in AUC, 0.4% in accuracy, 0.4% in precision, 0.3% in the recall and 0.4% difference in specificity compared to its performance with 10-fold cross-validation. The proposed LR model achieved an AUC of 99.6%, accuracy of 97.7%, precision of 99.1%, the recall of 97.2% and specificity of 98.4% having 0.3% difference in AUC, 1.2% in accuracy, 2.2% in precision and 4% difference in specificity compared to its performance with 10-fold cross-validation. Finally, the proposed NN model achieved an AUC of 99.9%, accuracy of 99.4%, precision of 99.1, 100% recall and specificity of 98.4% having 0.4% difference in AUC, 1.2% in accuracy, 1.3% in precision, 0.6% in the recall and 2.2% difference in specificity compared to its performance with 10-fold cross-validation.

The most significant difference between train-test split and 10-fold cross-validation methods are the recall achieved from the proposed LR model, aside from NN, proposed LR model also achieved the highest precision and specificity while SVM also achieve 100% recall similar to the proposed NN model. The hyper parameterization performed in the NN can be seen in *Table 10*. This includes the number of hidden layers set to 100, with the Logistic set as the activation function of the model. In addition, the output optimizer used in the model is ADAM, the regularization is set to 10% with a maximum iteration set to 100.

Moreover, the noticeable achievement of this study is the identification of how each attribute makes an impact on the hyper-parameterized models can be seen in *Figure 15*. In contrast with the performance of each model which involves SVM, MLPNN or NN and LR. NN provided the highest accuracy, performance among the three models that registered a 99.4% rating, which is significantly higher than the 98.8% of SVM and to 97.7% of LR. More so, in

terms of attribute deduction, LR shows a less receding validation as per cases compared to NN and SVM. SVM received the performance drop between case 6 and case 7. Moreover, attribute deduction starting in case 7 and beyond shows that LR is better than NN and SVM. NN as the best performing model in this experiment shows unaffected when the first four attributes are removed from the dataset for re-training, surprisingly similar to SVM. LR seems not to be affected by the first six attributes in the BCWD dataset. Overall, as the number of attributes reduced by one every re-training the performance of the model in general decreases, which implies that all attributes in the BCWD dataset is contributory in achieving the best performance of the model.

Moreover, the present work achieved outstanding findings in terms of several statistical methods that can be seen in *Table 11* using the proposed model mentioned in the method section of this paper. NN was identified by the authors using several statistical validations which includes confusion matrix, ROC, AUC, specificity, and sensibility to be the best model which generated the most superior result as compared to all existing studies that utilized the same dataset. The present work is able to achieve the 99.4% accuracy rating, higher than all previous experiments and researchers that used the BCWD dataset. A complete list of abbreviations is shown in *Appendix I*.

6. Conclusion and future work

Accurate classification of cancer is critical for the development of the most effective treatment plan for the patients. This study provided further exploration in the application of machine learning models in cancer classification thus contributing largely to cancer studies that can be utilized by experts in the field of medicine. Since there are few research papers published related to this study, this paper provided an improvement in the classification of breast cancer using hyper-parameterized models.

The three machine learning methods used; SVM, LR, and NN, have been tested with BCWD dataset. Each machine learning method has been tested by having multiple models of each method having its own unique set of parameters. Then, the performance of each model was compared to one another to determine the best classifier in terms of classification accuracy, precision, recall, and specificity.

The most notable contribution of this study would be the overall performance of the proposed NN model which achieved the highest accuracy level of 99.4%.

With 10-fold cross-validation, the proposed NN model falls second after the proposed SVM model in accuracy. This was achieved by separating the training and test data the proposed models were able to deliver unbiased predictions and handled overfitting better. This process has directed the study to generate a result that bested all the current accuracy, performance from previous studies concerning BCWD.

Different feature selection techniques could be applied and additional parameter tuning can be introduced to enhance the performance of the classifiers in the future.

Acknowledgment

The authors would like to acknowledge their respective institutions, the College of Informatics and Computing Sciences Graduate Program of Batangas State University-Alangilan Campus, the College of Arts, Sciences and Technology of Occidental Mindoro State College. The dataset used for the experimentation has been obtained from the University of California, Irvine (UCI) machine learning repository which is available to the public.

Conflicts of interest

The authors have no conflicts of interest to declare.

References

- [1] Ghanbari A, Rahmatpour P, Hosseini N, Khalili M. Social determinants of breast cancer screening among married women: a cross-sectional study. *Journal of Research in Health Sciences*. 2020; 20(1):1-5.
- [2] Zitvogel L, Tesniere A, Kroemer G. Cancer despite immunosurveillance: immunoselection and immunosubversion. *Nature Reviews Immunology*. 2006; 6(10):715-27.
- [3] Rakoff-nahoum S. Cancer issue: why cancer and inflammation? *The Yale Journal of Biology and Medicine*. 2006; 79(3-4):123-30.
- [4] Akay MF. Support vector machines combined with feature selection for breast cancer diagnosis. *Expert Systems with Applications*. 2009; 36(2):3240-7.
- [5] Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*. 2018; 68(6):394-424.
- [6] Sun YS, Zhao Z, Yang ZN, Xu F, Lu HJ, Zhu ZY, et al. Risk factors and preventions of breast cancer. *International Journal of Biological Sciences*. 2017; 13(11):1387-97.
- [7] Drukteinis JS, Mooney BP, Flowers CI, Gatenby RA. Beyond mammography: new frontiers in breast cancer screening. *The American Journal of Medicine*. 2013; 126(6):472-9.
- [8] Mckinney SM, Sieniek M, Godbole V, Godwin J, Antropova N, Ashrafian H, et al. *International*

- evaluation of an AI system for breast cancer screening. *Nature*. 2020; 577(7788):89-94.
- [9] Sahu B, Mohanty S, Rout S. A hybrid approach for breast cancer classification and diagnosis. *EAI Endorsed Transactions on Scalable Information Systems*. 2019; 6(20):1-8.
- [10] Cho SB, Won HH. Machine learning in DNA microarray analysis for cancer classification. In *proceedings of the first Asia-pacific bioinformatics conference on bioinformatics 2003* (pp. 189-98).
- [11] Liu Y. Active learning with support vector machine applied to gene expression data for cancer classification. *Journal of Chemical Information and Computer Sciences*. 2004; 44(6):1936-41.
- [12] Jordan MI, Mitchell TM. Machine learning: trends, perspectives, and prospects. *Science*. 2015; 349(6245):255-60.
- [13] Jin X, Xu A, Bie R, Guo P. Machine learning techniques and chi-square feature selection for cancer classification using SAGE gene expression profiles. In *international workshop on data mining for biomedical applications 2006* (pp. 106-15). Springer, Berlin, Heidelberg.
- [14] Lavanya D, Rani DK. Analysis of feature selection with classification: breast cancer datasets. *Indian Journal of Computer Science and Engineering*. 2011; 2(5):756-63.
- [15] Salama GI, Abdelhalim M, Zeid MA. Breast cancer diagnosis on three different datasets using multi-classifiers. *International Journal of Computer and Information Technology*. 2012; 1(1):36-43.
- [16] Utomo CP, Kardiana A, Yuliwulandari R. Breast cancer diagnosis using artificial neural networks with extreme learning techniques. *International Journal of Advanced Research in Artificial Intelligence*. 2014; 3(7):10-4.
- [17] Obaid OI, Mohammed MA, Ghani MK, Mostafa A, Taha F. Evaluating the performance of machine learning techniques in the classification of Wisconsin breast cancer. *International Journal of Engineering & Technology*. 2018; 7(4.36):160-6.
- [18] Dhanya R, Paul IR, Akula SS, Sivakumar M, Nair JJ. A comparative study for breast cancer prediction using machine learning and feature selection. In *international conference on intelligent computing and control systems 2019* (pp. 1049-55). IEEE.
- [19] Omondiagbe DA, Veeramani S, Sidhu AS. Machine learning classification techniques for breast cancer diagnosis. In *IOP conference series: materials science and engineering 2019* (pp.1-16). IOP Publishing.
- [20] Gupta P, Garg S. Breast cancer prediction using varying parameters of machine learning models. *Procedia Computer Science*. 2020; 171:593-601.
- [21] Balaraman S. Comparison of classification models for breast cancer identification using google colab. Preprints 2020.
- [22] Laghmati S, Cherradi B, Tmiri A, Daanouni O, Hamida S. Classification of patients with breast cancer using neighbourhood component analysis and supervised machine learning techniques. In *3rd international conference on advanced communication technologies and networking 2020* (pp. 1-6). IEEE.
- [23] Durgesh KS, Lekha B. Data classification using support vector machine. *Journal of Theoretical and Applied Information Technology*. 2010; 12(1):1-7.
- [24] Matsumoto A, Aoki S, Ohwada H. Comparison of random forest and SVM for raw data in drug discovery: prediction of radiation protection and toxicity case study. *International Journal of Machine Learning and Computing*. 2016; 6(2):145-8.
- [25] Chai H, Huang HH, Jiang HK, Liang Y, Xia LY. Protein-protein interaction network construction for cancer using a new L1/2-penalized Net-SVM model. *Genetics and Molecular Research: GMR*. 2016; 15(3):1-14.
- [26] Tirzite M, Bukovskis M, Strazda G, Jurka N, Taivans I. Detection of lung cancer with electronic nose and logistic regression analysis. *Journal of Breath Research*. 2018; 13(1):1-9.
- [27] Alarabeyyat A, Alhanahnah M. Breast cancer detection using k-nearest neighbor machine learning algorithm. In *international conference on developments in eSystems engineering 2016* (pp. 35-9). IEEE.
- [28] Patrício M, Pereira J, Crisóstomo J, Matafome P, Gomes M, Seíça R, et al. Using resistin, glucose, age and BMI to predict the presence of breast cancer. *BMC Cancer*. 2018; 18(1):1-8.
- [29] Zhang YD, Satapathy SC, Guttery DS, Górriz JM, Wang SH. Improved breast cancer classification through combining graph convolutional network and convolutional neural network. *Information Processing & Management*. 2021; 58(2).
- [30] Mohammed MA, Al-khateeb B, Rashid AN, Ibrahim DA, Abd GMK, Mostafa SA. Neural network and multi-fractal dimension features for breast cancer classification from ultrasound images. *Computers & Electrical Engineering*. 2018; 70:871-82.
- [31] Higa A. Diagnosis of breast cancer using decision tree and artificial neural network algorithms. *International Journal of Computer Applications Technology and Research*. 2018;7(1): 23-7.
- [32] Vijayakumar T. Neural network analysis for tumor investigation and cancer prediction. *Journal of Electronics*. 2019; 1(2):89-98.
- [33] <http://archive.ics.uci.edu/ml>. Accessed 26 May 2021.
- [34] Hazra A, Mandal SK, Gupta A. Study and analysis of breast cancer cell detection using naïve bayes, SVM and ensemble algorithms. *International Journal of Computer Applications*. 2016; 145(2):39-45.
- [35] Seddik AF, Shawky DM. Logistic regression model for breast cancer automatic diagnosis. In *SAI intelligent systems conference 2015* (pp. 150-4). IEEE.
- [36] Thein HT, Tun KM. An approach for breast cancer diagnosis classification using neural network. *Advanced Computing*. 2015; 6(1):1-11.
- [37] Ukil A. Support vector machine. In *intelligent systems and signal processing in power engineering 2007* (pp. 161-226). Springer, Berlin, Heidelberg.

- [38] Cruz JA, Wishart DS. Applications of machine learning in cancer prediction and prognosis. *Cancer Informatics*. 2006; 2:59-78.
- [39] Byvatov E, Schneider G. Support vector machine applications in bioinformatics. *Applied Bioinformatics*. 2003; 2(2):67-77.
- [40] Pisner DA, Schnyer DM. Support vector machine. In *Machine Learning 2020* (pp. 101-21). Academic Press.
- [41] Bayrak EA, Kırıcı P, Ensari T. Comparison of machine learning methods for breast cancer diagnosis. In *scientific meeting on electrical-electronics & biomedical engineering and computer science 2019* (pp. 1-13). IEEE.
- [42] Kourou K, Exarchos TP, Exarchos KP, Karamouzis MV, Fotiadis DI. Machine learning applications in cancer prognosis and prediction. *Computational and Structural Biotechnology Journal*. 2015; 13:8-17.
- [43] Hussain M, Wajid SK, Elzaart A, Berbar M. A comparison of SVM kernel functions for breast cancer detection. In *eighth international conference computer graphics, imaging and visualization 2011* (pp. 145-50). IEEE.
- [44] Cherkassky V, Ma Y. Practical selection of SVM parameters and noise estimation for SVM regression. *Neural Networks*. 2004; 17(1):113-26.
- [45] Lin SW, Lee ZJ, Chen SC, Tseng TY. Parameter determination of support vector machine and feature selection using simulated annealing approach. *Applied Soft Computing*. 2008; 8(4):1505-12.
- [46] Ruiz A, Villa N. Storms prediction: logistic regression vs random forest for unbalanced data. *arXiv preprint arXiv:0804.0650*. 2008.
- [47] Yusuff H, Mohamad N, Ngah UK, Yahaya A. Breast cancer analysis using logistic regression. *International Journal of Research and Reviews in Applied Sciences*. 2012; 10(1):14-22.
- [48] Murtirawat R, Panchal S, Singh VK, Panchal Y. Breast cancer detection using k-nearest neighbors, logistic regression and ensemble learning. In *international conference on electronics and sustainable communication systems 2020* (pp. 534-40). IEEE.
- [49] Graja O, Azam M, Bouguila N. Breast cancer diagnosis using quality control charts and logistic regression. In *9th international symposium on signal, image, video and communications 2018* (pp. 215-20). IEEE.
- [50] Sharma A, Kulshrestha S, Daniel S. Machine learning approaches for breast cancer diagnosis and prognosis. In *international conference on soft computing and its engineering applications 2017* (pp. 1-5). IEEE.
- [51] Goodman J. Exponential priors for maximum entropy models. In *proceedings of the human language technology conference of the north american chapter of the association for computational linguistics: HLT-NAACL 2004* (pp. 305-12).
- [52] Lee SI, Lee H, Abbeel P, Ng AY. Efficient l-1 regularized logistic regression. In *AAAI 2006* (pp. 401-8).
- [53] Salehi F, Abbasi E, Hassibi B. The impact of regularization on high-dimensional logistic regression. *arXiv preprint arXiv:1906.03761*. 2019.
- [54] Ng AY. Feature selection, L1 vs. L2 regularization, and rotational invariance. In *proceedings of the twenty-first international conference on machine learning 2004*.
- [55] Demir-kavuk O, Kamada M, Akutsu T, Knapp EW. Prediction using step-wise L1, L2 regularization and feature selection for small data sets with large number of features. *BMC Bioinformatics*. 2011; 12(1):1-10.
- [56] Jaiswal S, Mehta A, Nandi GC. Investigation on the effect of L1 and L2 regularization on image features extracted using restricted boltzmann machine. In *second international conference on intelligent computing and control systems 2018* (pp. 1548-53). IEEE.
- [57] Li M, Nanda G, Chhajed S, Sundararajan R. Machine learning-based decision support system for early detection of breast cancer. *Indian Journal of Pharmaceutical Education and Research*. 2020; 54(3):S705-15.
- [58] Tian H, Cai H, Wen J, Li S, Li Y. A music recommendation system based on logistic regression and eXtreme gradient boosting. In *international joint conference on neural networks 2019* (pp. 1-6). IEEE.
- [59] Floyd JCE, Lo JY, Yun AJ, Sullivan DC, Kornguth PJ. Prediction of breast cancer malignancy using an artificial neural network. *Cancer: Interdisciplinary International Journal of the American Cancer Society*. 1994; 74(11):2944-8.
- [60] Karabatak M, Ince MC. An expert system for detection of breast cancer based on association rules and neural network. *Expert Systems with Applications*. 2009; 36(2):3465-9.
- [61] Heidari AA, Faris H, Aljarah I, Mirjalili S. An efficient hybrid multilayer perceptron neural network with grasshopper optimization. *Soft Computing*. 2019; 23(17):7941-58.
- [62] Bui DT, Tuan TA, Klempe H, Pradhan B, Revhaug I. Spatial prediction models for shallow landslide hazards: a comparative assessment of the efficacy of support vector machines, artificial neural networks, kernel logistic regression, and logistic model tree. *Landslides*. 2016; 13(2):361-78.
- [63] Pham BT, Nguyen MD, Bui KT, Prakash I, Chapi K, Bui DT. A novel artificial intelligence approach based on multi-layer perceptron neural network and biogeography-based optimization for predicting coefficient of consolidation of soil. *Catena*. 2019; 173:302-11.
- [64] Sharma S, Sharma S, Athaiya A. Activation functions in neural networks. *International Journal of Engineering Applied Sciences and Technology*. 2017; 4(12):310-6.
- [65] Karlik B, Olgac AV. Performance analysis of various activation functions in generalized MLP architectures of neural networks. *International Journal of Artificial Intelligence and Expert Systems*. 2011; 1(4):111-22.

[66] Liao TW, Chen LJ. A neural network approach for grinding processes: modelling and optimization. *International Journal of Machine Tools and Manufacture*. 1994; 34(7):919-37.

[67] Amrane M, Oukid S, Gagaoua I, Ensari T. Breast cancer classification using machine learning. In *electric electronics, computer science, biomedical engineering' meeting 2018* (pp. 1-4). IEEE.

[68] Faraggi D, Simon R. A simulation study of cross-validation for selecting an optimal cutpoint in univariate survival analysis. *Statistics in Medicine*. 1996; 15(20):2203-13.

[69] Nematzadeh Z, Ibrahim R, Selamat A. Comparative studies on breast cancer classifications with k-fold cross validations using machine learning techniques. In *Asian control conference 2015* (pp. 1-6). IEEE.

[70] Mojarad SA, Dlay SS, Woo WL, Sherbet GV. Breast cancer prediction and cross validation using multilayer perceptron neural networks. In *international symposium on communication systems, networks & digital signal processing 2010* (pp. 760-4). IEEE.

[71] Kumar GR, Ramachandra GA, Nagamani K. An efficient prediction of breast cancer data using data mining techniques. *International Journal of Innovations in Engineering and Technology*. 2013; 2(4):139-44.

[72] Alakus TB, Turkoglu I. Comparison of deep learning approaches to predict COVID-19 infection. *Chaos, Solitons & Fractals*. 2020.

[73] Vabalas A, Gowen E, Poliakoff E, Casson AJ. Machine learning algorithm validation with a limited sample size. *PloS one*. 2019; 14(11):1-20.

[74] De MBAF, Miraglia JL, Donato TH, Chiavegatto FAD. COVID-19 diagnosis prediction in emergency care patients: a machine learning approach. *medRxiv*. 2020.

[75] Chen HL, Yang B, Liu J, Liu DY. A support vector machine classifier with rough set-based feature selection for breast cancer diagnosis. *Expert Systems with Applications*. 2011; 38(7):9014-22.

[76] Hernandez RM, Hernandez AA. Classification of Nile Tilapia using convolutional neural network. In *9th international conference on system engineering and technology 2019* (pp. 126-31). IEEE.



Kristoffersen Edward Mayce R. Lomboy is a part-time Instructor in the College of Arts, Sciences, and Technology Department at Occidental Mindoro State College Mamburao campus. He holds a Bachelor's Degree in Information Technology from Systems Technology Institute (STI)

College Global City. He is currently pursuing his Master's Degree in Information Technology (MIT) at the Batangas State University, Alangilan. His current research interests are Machine Learning and Data Mining.

Email: klomboy_cast@omsc.ph.education



Rowell M. Hernandez received BSIT in AMA Computer College Batangas with recognition, Master's Degree in Information Technology at Batangas State University. A graduate of Technological Institute of the Philippines-Manila with a Doctorate Degree of Information Technology with research anchored in machine learning application to Agriculture and Bioinformatics. He is program chair of the Master of Science in Computer Science and Master of Science in Data Science. He is also a faculty researcher of Batangas State University under the Digital Transformation Center of Science, Technology, Engineering and Environment Research Hub (STEER Hub) that aims to provide research that can create an impact to the local and national level. He is also an Associate member of the National Research Council of the Philippines under the Engineering and Industrial Research Division.
Email: hernandezrowell@gmail.com

Appendix I

S. No.	Abbreviation	Description
1	ANN	Artificial Neural Network
2	AUC	Area under Curve
3	BCWD	Breast Cancer Wisconsin Diagnostic
4	BDR	Bimodal Distribution Removal
5	BMI	Body Mass Index
6	BPNN	Back Propagation Neural Network
7	C	Cost
8	CART	Classification And Regression Tree
9	CFS	Correlation based Feature Selection
10	CNN	Convolutional Neural Network
11	ELMNN	Extreme Learning Machine Neural Networks
12	FN	False Negatives
13	FNA	Fine Needle Aspirate
14	FP	False Positives
15	HOMA	Homeostatic Model Assessment
16	KNN	K-Nearest Neighbors
17	LASSO	Least Absolute Shrinkage and Selection Operator
18	L-BFGS-B	Limited-Memory Broyden Fletcher Goldfarb Shanno
19	LDA	Linear Discriminant Analysis
20	LR	Logistic Regression
21	MLP	Multi-Layer Perceptron
22	MLPNN	Multi-Layer Perceptron Neural Networks

23	NCA	Neighborhood Component Analysis
24	NN	Neural Network
25	PCA	Principal Component Analysis
26	RBF	Radial Basis Function
27	RFE	Recursive Feature Elimination
28	ROC	Receiver Operating Characteristic
29	SCAD	Smoothly Clipped Absolute Deviations
30	SE	Standard Error
31	SGD	Stochastic Gradient Descent
32	SLP	Single Layer Perceptron
33	SMO	Sequential Minimal Optimization
34	SVM	Support Vector Machine
35	TN	True Negatives
36	TP	True Positives
37	UCI	University of California, Irvine