A Comparative Analysis of Nonlinear Machine Learning Algorithms for Breast Cancer Detection

Ali Al Bataineh

Abstract—**Breast cancer is a form of invasive cancer and one of the most common health problems for women that is globally responsible for a large number of deaths. Accurately classifying and categorizing breast cancer subtype is an essential task. Automated techniques based on artificial intelligence can significantly save time and reduce error. In this paper, a performance comparison between five nonlinear machine learning algorithms viz Multilayer Perceptron (MLP), K-Nearest Neighbors (KNN), Classification and Regression Trees (CART), Gaussian Nave Bayes (NB) and Support Vector Machines (SVM) on the Wisconsin Breast Cancer Diagnostic (WBCD) dataset is conducted. The primary objective is to evaluate the performance in classifying data with respect to efficiency and effectiveness of each algorithm in terms of classification test accuracy, precision, and recall.**

Index Terms—**Artificial intelligence, machine learning, Multilayer Perceptron (MLP), K-Nearest Neighbors (KNN), Classification and Regression Trees (CART), Gaussian Naïve Bayes (NB), Support Vector Machines (SVM).**

I. INTRODUCTION

Breast cancer has been identified as the second largest cause of cancer deaths among women of age 40 and 55 after lung cancer. According to the World Health Organization's projection, every year the estimated number of breast cancer diagnosis among women is to be 1.5 million [1]. In 2015, 500,000 women died from breast cancer [2]. According to the physicians, an abnormal multiplication of cells in the breast tissue causes breast cancer [3].

Fig. 1 shows the breast cancer classification model with machine learning algorithms, where the breast image database is loaded, features have to be extracted and the classification model can be trained and used for prediction of benign and malignant. Benign cases are considered noncancerous, that is non-life threatening. Malignant cancer starts with abnormal cell growth and might rapidly spread or invade nearby tissue so that it can be life-threatening [4].

For reducing this cause of death, early detection and diagnosis are critical. There are two early detection strategies for the breast cancer which are: early diagnosis and screening. The better diagnostic tools and methods can minimize this fatality rate. The recent development of the breast cancer diagnosis allows identifying the cancer cells; If the diagnosis tools become more efficient, then the detection and prediction can be made more effective [2].

Diagnosis process is employed to significantly and accurately discern between benign and malignant cancerous

patterns. Mammography is one of the conventional approaches for breast cancer diagnosis along with surgical biopsy and fine needle aspirate. The sensitivity results of this diagnosis of malignant lumps approach are, mammography 68%-79%, fine needle aspirate 65%-98%, and the surgical biopsy is about 100% [5]. A surgical biopsy is a costly approach for the diagnosis; however, the effectiveness of it is more. The Fine Needle Aspiration biopsy requires extracting cell samples from the lump and conduct the visual observation under a microscope. Once the diagnosis of breast cancer performed, the prognosis is subsequently determined to predict the future development and characteristics of the cancerous cells on the breast [1].

Machine learning has multiple potential applications in medicine and has been applied to a wide variety of oncology tasks, such as predicting cancer susceptibility, survival rates, and treatments. Machine Learning (ML) is a subfield of Artificial Intelligence (AI) which allows machines to learn without or with the intervention of a human. In the field of AI, ML is one of the most popular models which has been implemented rapidly to train machines and develop predictive models for effective decision making. In classification and prediction problems, machine learning methods are the leading option for obtaining a better outcome. In breast cancer research, the ML methods could be used for identification and prediction of cancer. These ML methods could predict whether cancer is malignant or benign [6].

In this paper, a comparative analysis is presented of five different nonlinear ML-based classification algorithms namely Multilayer perceptron (MLP), K-Nearest Neighbors (KNN), Classification and Regression Trees (CART), Gaussian Nave Bayes (NB) and Support Vector Machines (SVM) on Wisconsin Breast Cancer Diagnostic dataset by measuring their classification test accuracy, recall, and precision. The dataset is split into two, 80% for training and 20% for testing. Results show that all the presented ML algorithms performed well on the classification task.
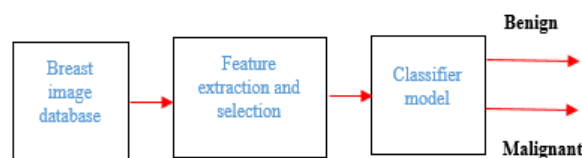

Fig. 1. A basic breast cancer image classification model (regenerated from [4]).

II. NOLINEAR MACHINE LEARNING ALGORITHMS

Machine learning algorithms can be categorized into two types; one is supervised learning, and another is unsupervised learning. In the supervised learning method, a set of data are used to train the machine and are labeled to give the correct

result; however, in unsupervised learning, there are no predetermined data sets and no predefined goals [7].

The field of machine learning establishes remarkable practical success in several different application areas. Some of the most prevalent applications include; medical expert systems, social network, speech recognition, etc.

Machine learning automatically learns how to make precise predictions based on past observations. Classification is a sub-area of machine learning related to supervised learning. As the name indicates, the determination is to classify examples into one of a discrete set of possible classes. Among the most recognized are Multilayer Perceptron (MLP), K-Nearest Neighbors (KNN), Classification and Regression Trees (CART), Gaussian Naive Bayes (NB), and Support Vector Machine (SVM). In machine learning, classification is typically conducted by training a classifier that discriminates against one class from other classes. Additionally, classification is necessarily a mission of predicting the value of a discrete attribute. For the classification problem, the preferred method is supervised learning-based algorithm. In a classification problem, a set of data are used to train the machine to develop the classification model after that it can be used for future predictions.

For the classification problem, the preferred method is supervised learning-based algorithm. In a classification problem, a set of data are used to train the machine to develop the classification model after that it can be used for future predictions.

Five different machine learning algorithms are applied for the classification of the Wisconsin cancer dataset, those five algorithms are:

A. Multilayer Perceptron (MLP)

MLP is a feed-forward artificial neural network; it consists of three layers of nodes. Except for the input node, each node is a neuron which uses a nonlinear activation function. MLP implies the supervised training method called back-propagation training. It adjusts the weight values that are calculated from input-output mappings and minimize the error between the correct output value and the target value. It iteratively computes the values of weight using a gradient descent algorithm [8]. For the Multilayer Perceptron (MLP) with the input vector, the output vector the weight is calculated by gradient descent rule and is given by following equation (1):

$$\Delta w_i = \alpha \, y_j (1 - y_j) \, (t_j - y_j) \, x_j \tag{1}$$

where α is the learning rate, $y_j (1 - y_j)$ is the derivative of activation function and $(t_j - y_j)$ is an error with t_j as a target [8]. Similarly, for the breast cancer classification, this research [9] found an excellent performance of the Multilayer Perceptron even when they reduced the number of features to be classified on their dataset of 663 instances and 25 markers.

Fig. 2 is the general model of MLP with an input layer, hidden layer, and an output layer. This MLP model is been implemented for the classification of breast cancer dataset by Ali Raad *et al.* [10]. They have provided the input of 9 attributes from the database while the MLP has hidden layer with five neurons and the output layer generate two outputs for two class of cancer classification as benign and malignant.

The activation function for the above MLP for classification are: Input Layer: there is no activation function in the input layer Hidden Layer: here is the simplified sigmoid function:

$$\left(\varphi(x) = \frac{1}{1 + e^{-x}} \right) \tag{2}$$

Output Layer: sigmoid function is $\varphi_0(x) = \varphi(x)$ or hyperbolic tangent which is:

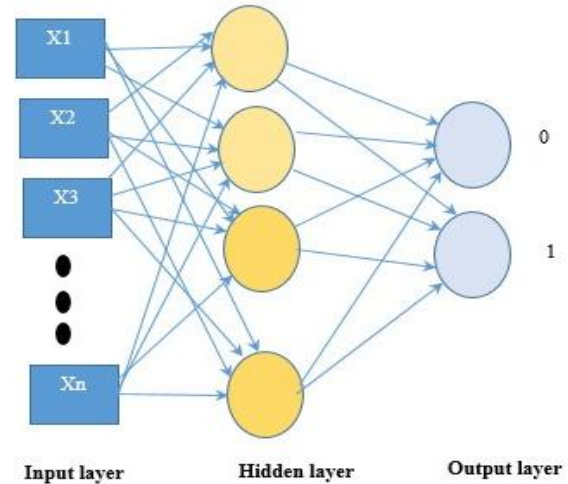$$\varphi_0(x) = \tanh\left(\frac{x}{2}\right) = \frac{1 - e^{-x}}{1 + e^{-x}} \tag{3}$$

Fig. 2. MLP neural network model (regenerated from [10]).

B. K-Nearest Neighbors (KNN)

KNN algorithm is an unsupervised learning algorithm on machine learning used for the classification. K-Nearest Neighbors is one of the most used algorithms in machine learning for the breast cancer classification [11]. This learning method is based on instances, so it doesn't require a learning phase. The KNN model works with the training sample which is associated with a distance function and the choice of the function of the class based on the classes of the nearest neighbors. Before classifying the new element, we must compare it to other elements using a similarity measure then k-nearest neighbors are considered. The class that appears most among the neighbors is assigned to the element to be classified. In KNN, the distance between the neighbor and the new element to be classified is used as the weight of the neighbor [8].

The K-Nearest Neighbors Algorithm can be represented by the following [12]:

Step 1: Choose a value for the parameter k.

Step 2: Input: Give a sample of N examples and their classes.

The classes of sample x are c(x).

Step 3: Give a new sample y.

Step 4: Determine the k-nearest neighbors of y by calculating the distances.

Step 5: Combine classes of these y examples in one class c.

Step 6: Output: The class of y is c(Y) = c.

The distance can be calculated using different ways. Euclidean distance is the most universal, between two vectors x_{ir} and x_{jr}. The Euclidean distance is defined as [12]:

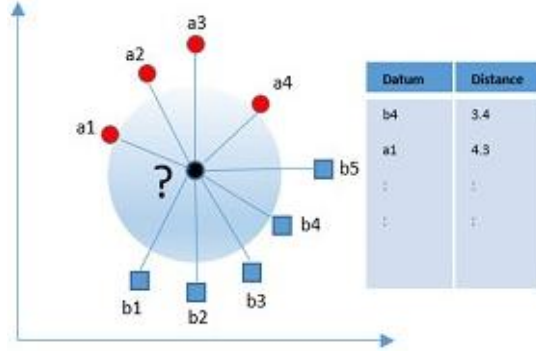$$d(x_i, x_j) = \sqrt{\sum_{r=1}^{n}(x_{ir} - x_{jr})^2} \qquad (4)$$

Fig. 3. Example of the use of the KNN algorithm (regenerated from [11]).

C. Classification and Regression Trees (CART)

The CART has been jointly developed by Leo Breiman, Jerome Friedman, Richard Olsen and Charles Stone [13]. The CART is a common method for developing statistical models from simple feature data. CART deals with data that is not fully finished, data with predicted and input features. The algorithm will consider the set of samples question about the data will lead to the data minimization and continues until some stop criteria is meet [14], [15].
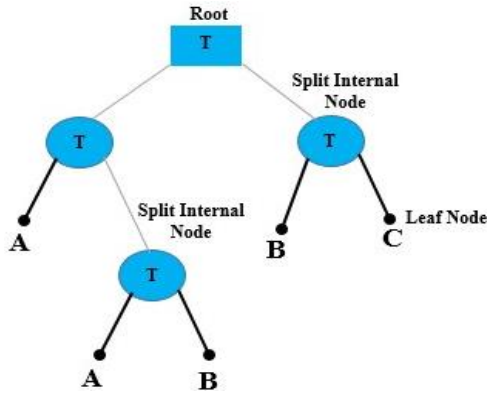

Fig. 4. The Structure of CART (regenerated from [14]).

Fig. 4 is the example of CART tree, when the dependent variable is categorical, CART produces a classification tree and when there is a continuous variable then it will lead to a regression tree.

CART algorithm has been used for prognostic and predictive analysis of breast cancer decision support for the relapse-free survival time, carried out in the subgroup of patients who receive no adjuvant systematic therapy [16].

CART analysis algorithm consists of three steps; the first step — an overgrown tree is built which closely describes the training set. This tree is called the maximal tree and is grown using a binary split-procedure. Second step — the over-grown tree shows that overfitting is being pruned.

During this procedure, a series of smaller subtrees are derived from the maximal trees. In the final step, the tree with the optimal size is selected using a cross-validation (CV) procedure [13].

The algorithm can also be resolved into these three steps: the first step is to split the attributes, second is determining the stopping criteria, and last is how nodes are assigned to classes.

This algorithm can be further explained as [13], [17]:

1) Building the maximal tree and splitting attributes

The maximal tree is build using a binary split procedure, which starts at the tree-root. At each level, the mother group is split into two daughter groups. To choose the most appropriate descriptor for the split value, CART uses an algorithm in which all the descriptors and split value are considered. The split which gives the best reduction in impurity between the mother group (t_p) and the daughter groups $(t_L \ and \ t_R)$ is selected Mathematically this can be expressed as:

$$i(s, t_p) = i_p(t_p) - p_{L_i}(t_L) - pR_i(t_R) \qquad (5)$$

2) Tree pruning

Tree pruning is the process in which a series of smaller sub trees derived from the maximal tree is obtained by successively cutting terminal branches. The different subtrees are then compared to find the optimal one which is based on a cost-complexity measure. In this process both tree accuracy and ,complexity is considered. The cost-complexity parameter $R_\propto(T)$ is used and for each subtree T is defined as follows:

$$R_\propto(T) = R(T) + \propto |\breve{T}| \qquad (6)$$

where $R(T)$ is the average within node sum of squares, $|\breve{T}|$ is the tree complexity and $\propto$ is the complexity parameter. During the pruning process, $\propto$ is gradually increased from 0 to 1. At this scenario, a tree is selected which minimizes $R_\propto(T)$. When $\propto= 0$, $R_\propto(T)$ is minimized by the maximal tree.

3) Select the optimal tree

The most accurate tree is one with the smallest mean CV error, defined the root mean square error of cross-validation (RMSCEV):

$$RMSECV = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)}{n}} \qquad (7)$$

where y_i is the response value of object i , $\hat{y}_i$ is the predicted response value for object i and n is the total number of objects.

4) Variable ranking

In this step, CART allows evaluating the importance of the different explanatory variables to define a split in the selected dataset.

The importance of a variable is defined as;

$$M(x_m) = \sum_{t \in T} \Delta I(\tilde{s}_m, t) \qquad (8)$$

D. Gaussian Naive Bayes (NB)

NB is known as the simple classification algorithm which is based on the Bayes classification algorithm. This paper [3] implements one of the Naïve Bayes algorithms for the breast cancer detection. The weight NB techniques give the sensitivity, specificity and the accuracy values as 99.11%, 98.25 %, and 98.54 % respectively. This research also finds

that the performance of the weighted NB is better than the regular NB method on the breast cancer classification.

Bayes theorem states the following equation for the given class variable and a dependent feature vector [18]:

$$P(y|x_1,\ldots,x_n) = \frac{p(y)p(x_1,\ldots,x_n|y)}{p(x_1,\ldots,x_n)} \qquad (9)$$

The above equation (9) is to calculate the probability of each class y given the values x_i of all the attributes for an instance to be classified, assuming the conditional independence of the attributes given the class [7].

Types of naïve Bayes algorithms are [19]:
1) Gaussian Naïve Bayes.
2) Multinomial Naïve Bayes.
3) Bernoulli Naïve Bayes.

Gaussian Naïve Bayes is preferred when attribute values are continuous and distributed according to the Gaussian distribution.

The Multinomial Naïve Bayes is one of the classification algorithms which is preferred when there is data is of multinomially distributed. This algorithm is used in text classification.

Bernoulli Naïve Bayes is used on the data that is distributed according to multivariate Bernoulli distribution, i.e. each one features is assumed to be a binary-valued variable [18], [19].

Soria *et al.* [9] also implemented the naïve Bayes classification algorithm along with the C4.5 tree and Multilayer Perceptron (MLP) for breast cancer classification. The research found that the performance of the Naïve Bayes achieved a competitive performance even though the assumption of normality of data is strongly violated.

Igor Konenenko [7] has implemented different machine learning algorithms for the medical diagnosis and prognosis. On the medical diagnostic problem, the Naïve Bayesian classifier outperformed on five out of eight problems in terms of the efficiency. He compares different classification algorithm as LFC, Back Propagation, KNN and other in terms of performance, transparency, explanation, reduction and missing data handling capability.

Gaussian NB implements a special type of Naïve Bayes algorithm that is Gaussian Naïve Bayes algorithm for classification. This algorithm is specially used when the features have continuous values, or all the features are following a Gaussian distribution i.e., normal distribution. The likelihood of the features is assumed to be Gaussian [18]:

$$P(x_i|y) = \frac{1}{\sqrt{2\pi\sigma_y^2}} \ \exp\left(-\frac{(x_i - \mu_i)^2}{2\sigma_y^2}\right) \qquad (10)$$

In the above equation (10), x is an attribute of continuous data, the parameters σ_y and μ_i are estimated using maximum likelihood. The mean μ_i and variance σ_y is calculated after the segmentation of data by class.

Fig. 5 shows the probability of Gaussian Naïve Bayes with no calibration, isotonic calibration, sigmoid calibration and empirical. This probability gives the confidence on the prediction. This associated probability can be predicted along with the class label while performing the classification. The predicted probabilities and its calibration are often desirable as a post-processing on classification stages [20].
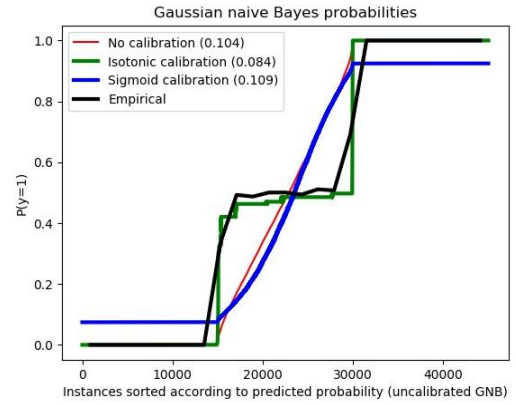

Fig. 5. Gaussian Naïve Bayes (NB) probabilities [20].

E. Support Vector Machine (SVM)

SVM is a machine learning classification technique which implements the supervised model for learning and it is widely used for cancer diagnosis and prognosis field [6]. The SVM method separates two classes by determining the linear classifier that maximizes the margin. This separation is referred to as the optimal separating hyperplane [1].

The regularization parameter and kernel function are the two major components that have to been determined before conducting training. Some of the significant researches employed using the SVM for breast cancer detection utilized heuristics SVM approaches such as the smooth SVM, the linear SVM and general nonlinear SVM [1], [21]. The goal of SVM is to determine a suitable hyperplane with a maximum margin which can be computed as an optimization problem [1]. Fig. 6. shows a scatter plot of two classes with two properties. A linear hyperplane is defined as $ax_1 + bx_2$ and the aim to find a, b and c such that $ax_1 + bx_2 \leq c$ for class 1 and that $ax_1 + bx_2 > c$ for class 2. Different from other techniques, SVM depends on the support vectors, which are the data sets closest to the decision boundary, in their algorithms. This is because removing other data points that are further away from the decision hyperplane will not change the boundary as much as if the support vectors were removed [6].
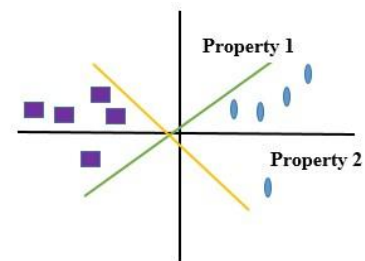

Fig. 6. Linear SVM.

The algorithm of SVM is [22]
Step 1: Load the Dataset
Step 2: Classify Features (Attributes) based on class labels
Step 3: Estimate Candidate Support Value
While (instances! = null)
Do
Step 4: Support Value = similarity between each instance in the attribute

Find total error value
Step 5: if any instance <0
Estimate
Decision value = support value/total error
Repeat for all points until it will empty
End if

III. DATASET

The data used in this work is the Wisconsin Breast Cancer Diagnostic (WBCD) dataset that is obtained from the UCI machine learning repository [23]. This dataset was collected by Dr. Wolberg, Street and Mangasarian and team from the University of Wisconsin Hospitals over three years periodically.

A. Description of the WBCD Dataset

WBCD dataset is a binary classification dataset; it has 569 instances and 32 attributes. All the attributes of this dataset are calculated from digitizing the image of fine needle aspirate(FNA) of a breast mass. The attributes are: the ID, diagnosis (M = malignant, B = benign) and the other 30 real valued input feature attributes comprise the mean, standard error, and worst value for 10 different characteristics of the digitized cell nuclei which are: Radius, Texture, Perimeter, Area, Smoothness, Compactness, Concavity, Concave points, Symmetry, and Fractal dimension. [1], [23]. The dataset has a number of diagnosis of lumps and masses that were found in the patients. Based on the diagnosis the tumor or lump is either classified as malignant (M) or benign (B). The feature values are recoded with four significant digits. The class distribution (Fig. 7) of the samples is such that 357 are benign and 212 are malignant.

TABLE I: THE DESCRIPTION OF THE WISCONSIN DATASET

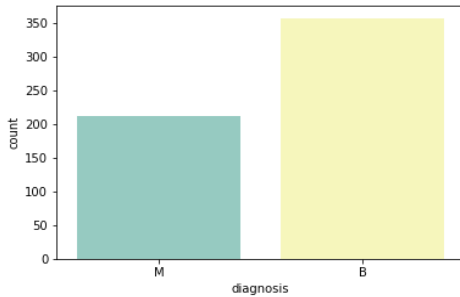| Attributes	Number of Attributes
Samples total	569
Dimensionality	30
Classes	2
Samples per class	Benign: 357 (62.74%) Malignant: 212 (37.26%)
Features	Real, positive

Fig. 7. Class distribution.

Fig. 8 Shows a heat map plot of the correlations between the features. Correlation indicates how related the changes are between the two features in the dataset. If two features change in the same direction, they are positively correlated, but If two features change in opposite directions, then they are negatively correlated. It shows which features have a high correlation with each other. This is valuable to acknowledge because some machine learning algorithms can have poor performance if there are highly correlated input features in the data.
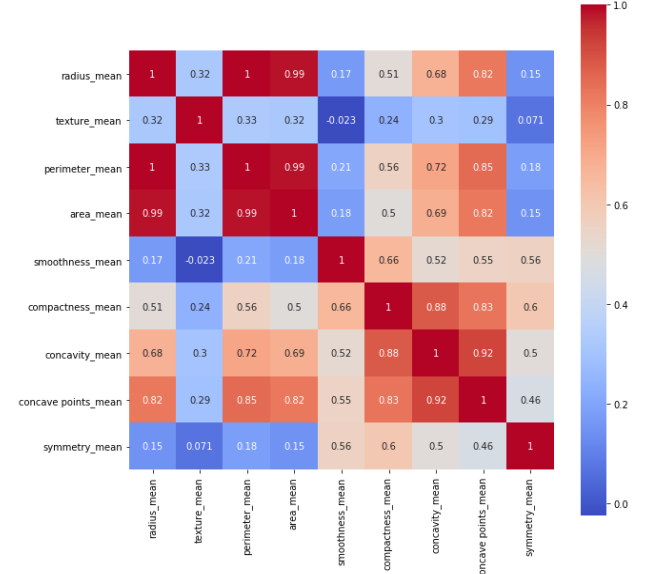

Fig. 8. Correlations between the features.

B. Dataset Preprocessing

In preprocessing, transformations are applied to data before feeding it to the algorithm. Data Preprocessing is a method that is employed to transform raw data into clean data. Standard machine learning algorithms expect the scale of the data to be equivalent to achieve better results. For this work, the WBCD data is standardized. Standardization is a rescaling method that transforms features with a Gaussian distribution and different means and standard deviations to a standard Gaussian distribution with a mean of 0 and a standard deviation of 1. The first feature is the variable named id which represents a unique identifier for each patient in the data is excluded from the dataset . The next variable is the diagnosis value. This variable is the outcome we expect to predict by the proposed algorithms. The standardized value of a feature is called a Z score and is calculated using the following formula.

$$Z = \frac{x - \mu}{\sigma} \tag{11}$$

where x is the feature value that is being standardized, μ is the mean of the distribution, and σ is the standard deviation of the distribution.

IV. RESULTS AND DISCUSSION

This section evaluates and compares the algorithms based on Performance Measures such as accuracy, precision, and recall.

A. Training Accuracy Comparison

The accuracy is measured to evaluate the performance of the five classification algorithms. Accuracy is obtained from the ratio of the correctly predicted instances number divided by the total number of instances from the dataset.

Mathematically, the accuracy of the classification has the following definition:

$$\text{Accuracy} = \frac{x}{n} \times 100\% \tag{12}$$

where x is the number of instances cases correctly classified, and n is the total number of instances cases.

In binary classification, accuracy can be calculated in terms of positives and negatives as follows:

$$Accuracy = \frac{TN+TP}{TN+TP+FN+FP} \times 100\% \qquad (13)$$

where TN = True Negatives (the model predicted benign, and the actual output was malignant), TP = True Positives (the model predicted malignant, and the actual output was also malignant), FN = False Negatives (the model predicted benign, and the actual output was also malignant), and FP = False Positives (the model predicted malignant, and the actual output was benign). Table II shows the accuracy of each model on the training dataset; the dataset is split into two, 80% for training and 20% for testing.

TABLE II: ALGORITHM COMPARISON

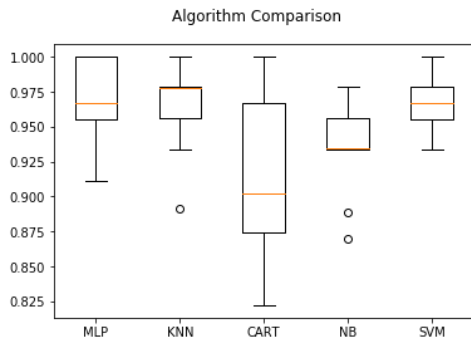| Algorithm	Accuracy (%)
MLP	96.70
KNN	96.27
CART	91.00
NB	93.62
SVM	96.42

Fig. 9. Algorithm and accuracy comparison.

Fig. 9 is the plot of the model evaluation results and comparison of the mean accuracy of each model. The figure displays that the box and whisker plots are squashed at the top of the range, which also displays that numerous samples reaching 100 % accuracy.

B. Performance Comparison

To estimate how each model perform outside the sample to a new dataset also identified as test data, the predictive performance of the models is assessed using Cross-Validation technique. The motivation to use cross-validation techniques is that when we fit a model, we fit it to a training dataset. Without cross-validation, we just have information on how the models perform to training data. Ideally, we would like to test the performance of the models to new data in terms of the accuracy of its predictions. In science, theories are judged by its predictive performance. It is a common technique because it is naive to understand and results in a less biased or less optimistic estimate of the model skill than other methods, such as a simple train/test split. The technique has an individual parameter named k that refers to the number of groups that a given data sample is to be split into. As such, the procedure is often called k-fold cross-validation. It is an excellent method to test and estimate the accuracy of the classifiers; it randomly separates the

training set into k subsets where 1 of the $k-1$ subsets will be used for testing and the rest for training [6]. In this paper, 10-fold cross-validation method is used; it splits the dataset into 10 parts, train on 9 and test on 1 and repeats for all combinations of train-test splits; this technique is instrumental in avoiding overfitting of the training set primarily in a small dataset with a large number of attributes.
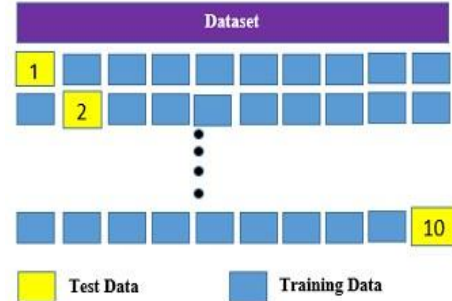

Fig. 9. 10-fold cross validation (regenerated from [6]).

1) Accuracy

The predicted accuracy of the 5 algorithms on the testing data are listed in Table III. We can see that MLP has the highest prediction accuracy of 99.12%. Having high accuracy does not conclude that MLP is the best model. Accuracy is only an excellent measure for symmetric datasets where values of false positive and false negatives are almost the same. Therefore, other performance measures are further tested such as precision and recall to evaluate and compare the performance of the models.

TABLE III: COMPARISON OF PREDICTION ACCURACY ON THE TRAINING DATA

Algorithm	Accuracy (%)
MLP	99.12
KNN	95.61
CART	93.85
NB	94.73
SVM	98.24

2) Precision

Precision is the ratio of correct positive results divided by the number of all total predicted positive observations. Mathematically, it can be expressed as:

$$Precision = \frac{TP}{TP+TF} \times 100\% \qquad (14)$$

Table IV shows the precision values for all five classifiers.

TABLE IV: PRECISION VALUES

	MLP	KNN	CART	NB	SVM
Benign	100%	100%	92%	95%	100%
Malignant	99%	94%	95%	95%	97%
Average	99%	96%	94%	95%	98%

3) Recall

Recall is also named sensitivity, is the ratio of correct positive results to all observations in actual class. Mathematically, it can be expressed as:

$$Recall = \frac{TP}{TP+FN} \times 100\% \qquad (15)$$

Table V shows the recall values for all five classifiers.

TABLE V: RECALL VALUES

	MLP	KNN	CART	NB	SVM
Benign	97%	88%	90%	90%	95%
Malignant	100%	100%	96%	97%	100%
Average	99%	96%	94%	95%	98%

The results presented in Tables III, IV, and V show that MLP classifier has the best performance in terms of accuracy, precision, and recall. These results signify that the MLP model is the best classifier among the five proposed classifiers for classifying a tumor as benign or malignant.

V. CONCLUSION

Machine learning techniques have been applied extensively in various medical applications and well-thought-out as valuable techniques that guide physicians in making decisions on accessible data as well as developing medical expert machines. In the study, five of the popular nonlinear machine learning algorithms are applied for breast cancer detection. The main features and methodology of each algorithm are described. Wisconsin Breast Cancer Diagnostic (WBCD) dataset is used as a benchmark to compare the performance of the proposed models as MLP, KNN, CART, SVM, and NB. The accuracy of MLP on the training data is 96.70% which is better than the other four algorithms. After the estimation, the performance of the predictive models is evaluated in terms of accuracy, precision, and recall on unseen data using k-fold cross-validation technique to test their skills. Results of this study confirm that the MLP model has the highest performance in terms of accuracy, precision, and recall of 99.12%, 99.00%, and 99.00% respectively.

REFERENCES

[1] H. You and G. Rumbe, "Comparative study of classification techniques on breast cancer FNA biopsy data," *International Journal of Artificial Intelligence and Interactive Multimedia*, vol. 1, no. 3, pp. 6-13, 2010.

[2] World Health Organization, "WHO position paper on mammography screening," World Health Organization, Geneva, 2014.

[3] M. Karabatak, "A new classifier for breast cancer detection based on Naïve Bayesian," *Measurement*, vol. 72, pp. 32-26, 2015.

[4] A.-A. Nahid and Y. Kong, "Involvement of machine learning for breast cancer image classification: A survey," *Computational and Mathematical Methods in Medicine*, pp. 1-29, 2017.

[5] P. Pantel, "Breast cancer diagnosis and prognosis," Univerisity of Manitoba, 1998.

[6] I. Kononenko, "Machine learning for medical diagnosis: History, state of the art and perspective," *Artificial Intelligence in Medicine*, vol. 23, no. 1, pp. 89-109, 2001.

[7] D. Bazazeh and R. Shubair, "Comparative study of machine learning algorithms for breast cancer detection and diagnosis," in *Proc. 2016 5th International Conference on Electronic Devices, Systems and Applications (ICEDSA)*, Ras Al Khaimah, 2016.

[8] M. Negnevitsky, *Artificial Intelligence: A Guide to Intelligent Systems, Pearson*, 2005, ch. 6, pp. 175-179.

[9] D. Soria, J. M. Garibaldi, E. Biganzoli and I. O. Ellis, "A comparison of three different methods for classification of breast cancer data," in *Proc. 2008 Seventh International Conference on Machine Learning and Applications*, 2008, pp. 619-624.

[10] A. Raad, A. Kalakech, and M. Ayache, "Breast cancer classification using neural network approach: MLP and RBF," in *Proc. The 13th International Arab Conference on Information Technology ACIT'2012*, 2012.

[11] A. P. Pawlovsky and M. Nagahashi, "A method to select a good setting for the kNN algorithm when using it for breast cancer prognosis," in *Proc. IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)*, Valencia, 2014.

[12] S. A. Medjahed, T. A. Saadi, and A. Benyettou, "Breast cancer diagnosis by using k-nearest neighbor with different distances and classification rules," *International Journal of Computer Applications*, vol. 62, no. 1, pp. 1-5, 2013.

[13] L. Breiman, J. Friedman, C. J. Stone, and R. A. Olshen, *Classification and Regression Trees*, Monterey: Wadsworth & Brooks/Cole Advanced Books & Software, 1984.

[14] J. Morgan, *Classification and Regression Tree Analysis*, Boston University School of Public Health, Boston, 2014.

[15] E. Venkatesan and T. Velmurugan, "Performance analysis of decision tree algorithms for breast cancer classification," *Indian Journal of Science and Technology*, vol. 8, no. 29, 2015.

[16] R. Kates, N. Harbeck, K. Ulm, F. Jaenicke, H. Graeff, and M. Schmitt, "Decision support in breast cancer: Recent advances in prognostic and predictive techniques," *Artificial Intelligence Techniques in Breast Cancer Diagnosis And Prognosis*, World Scientific Publishing, 2000, pp. 55-73.

[17] E. Deconinck, T. Hancock, D. Coomans, D. L. Massart, and Y. V. Heyden, "Classification of drugs in absorption classes using the classification and regression trees (CART) methodology," *Journal of Pharmaceutical and Biomedical Analysis*, vol. 39, pp. 91-103, 2005.

[18] A. McCallum and K. Nigam, "A Comparison of Event Models for Naive Bayes Text Classification," in *Proc. AAAI/ICML-98 Workshop on Learning for Text Categorization*, 1998.

[19] V. Metsis, I. Androutsopoulos, and G. Paliouras, "Spam Filtering with Naive Bayes -- Which Naive Bayes?" in *Proc. Third Conference on Email and Anti-Spam (CEAS)*, 2006.

[20] F. Pedregosa, G. Varoquaux, A. Gramfort *et al.*, "Scikit-learn: Machine Learning in {P}ython," *Journal of Machine Learning Research*, vol. 12, pp. 2825-2830, 2011.

[21] T. Mu and A. K. Nandi, "Breast cancer detection from FNA using SVM with different parameter tuning systems and SOM-RBF classifier," *Journal of the Franklin Institute*, vol. 344, pp. 285-311, 2007.

[22] Kathija and S. Nisha, "Breast cancer data classification using SVM and Naïve Bayes techniques," *International Journal of Innovative Research in Computer and Communication Engineering*, vol. 4, no. 12, 2016.

[23] D. Dheeru and E. K. Taniskidou, *UCI Machine Learning Repository*, 2017.

Ali Al Bataineh is a Ph.D. candidate at the Department of Electrical Engineering and Computer Science at the University of Toledo, Ohio, US. He holds an M.S in computer engineering with distinction from the University of Bridgeport, Connecticut, US, and a B.S. in computer engineering from Yarmouk University, Jordan. His current research interests lie in the areas of artificial intelligence, machine learning, computer vision, metaheuristic optimization, and fuzzy logic.