International Journal of Electrical and Computer Engineering (IJECE)
Vol. 14, No. 2, April 2024, pp. 2181~2190
ISSN: 2088-8708, DOI: 10.11591/ijece.v14i2.pp2181-2190

Performance enhancement of machine learning algorithm for breast cancer diagnosis using hyperparameter optimization

Rashidul Hasan Hridoy[1], Arindra Dey Arni[2], Shomitro Kumar Ghosh[2], Narayan Ranjan Chakraborty[2], Imran Mahmud[1]
[1]Department of Software Engineering, Faculty of Science and Information Technology, Daffodil International University, Dhaka, Bangladesh
[2]Department of Computer Science and Engineering, Faculty of Science and Information Technology, Daffodil International University, Dhaka, Bangladesh

Article Info

Article history:

Received Sep 18, 2023
Revised Dec 16, 2023
Accepted Dec 18, 2023

Keywords:

Breast cancer
Machine learning
Hyperparameter optimization
Grid search
Random search
Logistic regression
K-nearest neighbor

ABSTRACT

Breast cancer is the most fatal women's cancer, and accurate diagnosis of this disease in the initial phase is crucial to abate death rates worldwide. The demand for computer-aided disease diagnosis technologies in healthcare is growing significantly to assist physicians in ensuring the effectual treatment of critical diseases. The vital purpose of this study is to analyze and evaluate the classification efficiency of several machine learning algorithms with hyperparameter optimization techniques using grid search and random search to reveal an efficient breast cancer diagnosis approach. Choosing the optimal combination of hyperparameters using hyperparameter optimization for machine learning models has a straight influence on the performance of models. According to the findings of several evaluation studies, the k-nearest neighbor is addressed in this study for effective diagnosis of breast cancer, which got a 100.00% recall value with hyperparameters found utilizing grid search. k-nearest neighbor, logistic regression, and multilayer perceptron obtained 99.42% accuracy after utilizing hyperparameter optimization. All machine learning models showed higher efficiency in breast cancer diagnosis with grid search-based hyperparameter optimization except for XGBoost. Therefore, the evaluation outcomes strongly validate the effectiveness and reliability of the proposed technique for breast cancer diagnosis.

Corresponding Author:

Rashidul Hasan Hridoy
Department of Software Engineering, Faculty of Science and Information Technology, Daffodil International University
Daffodil Smart City, Birulia, Savar, Dhaka, Bangladesh
Email: rashidulhasanhridoy@gmail.com

1. INTRODUCTION

Breast cancer (BC) originates in breast tissues, where the growth of cells becomes uncontrollable, and cells build a tumor [1]. BC is mostly seen in adults which is the major cause of female deaths in 95% of countries and over 2.3 million (M) incidents of BC happen every year [2]. BC has a detrimental impact on human health and hinders the quality of life. According to several institutions of research, 0.4 M females die for BC every year. Radiation therapy, surgical removal, and medication are often used for controlling the growth and spreading of BC and early identification is highly crucial for effective BC treatment [3]. The early treatment and prevention of BC is an enormous challenge in healthcare and a remarkable improvement carried out in elementary research and clinical treatment in recent decades [4]. Computer-aided technologies used for BC diagnosis and analysis make the treatment of BC easier and reduce fatality [5].

Due to the shortage of proficient physicians and technologies, a notable number of countries are now facing several types of issues, where machine learning (ML) brings immense hope for its effective performance. ML-based tools already yielded satisfactory results and proved their capability in the diagnosis of different types of critical diseases. Applications of ML are now extensively used in healthcare and remarkably enhance the accuracy and speed of physicians' work [6]. The establishment of accurate diagnosis tools for BC utilizing ML techniques in health care can aid doctors in improving the diagnosis rate of BC. ML algorithms utilized for classification follow the supervised ML learning technique where trained models are used for predicting the right label of provided input data. Seven ML classifiers such as logistic regression (LR), k-nearest neighbor (k-NN), multilayer perceptron (MLP), support vector machine (SVM), naive Bayes (NB), XGBoost (XGB), and decision tree (DT) were used in this study to find out the most effective classifier of BC diagnosis.

Parameters of ML models are estimated or learned from the provided data during the training phase, which are basically defined as inner variables of ML models and each algorithm optimizes its parameters using several mechanisms [7]. Generally, training of ML models begins with initializing some values to parameters and these values are updated by utilizing an optimization algorithm. During the training, parameter values are updated continuously by the learning algorithm. On the other hand, hyperparameters are selected and set by the practitioner during the model configuration for controlling the learning process of ML models, which is not similar to parameters. In ML, hyperparameter optimization (HPO) is used for searching the best hyperparameters of an algorithm to ensure the best performance, and grid search (GS) and random search (RS) strategies are used to find these hyperparameters [8]. In this study, both GS and RS were utilized for finding the best hyperparameters for seven ML models. Besides this, these ML models were also trained with default parameters. This study is conducted using the Wisconsin breast cancer (diagnostic) (WBCD) dataset which provides 569 occurrences [9].

After using HPO, the classification accuracy of all ML models was enhanced except XGB. According to the findings of several evaluation studies, the LR model acquired higher accuracy with default parameters and GS and RS-based HPO, 98.25%, 99.42%, and 99.42%, respectively. On the other hand, k-NN got 99.42% accuracy using hyperparameters found by GS, which also obtained a 100.0% recall value. Moreover, GS-based HPO performed well than RS-based HPO. Compared to existing approaches developed with ML techniques for BC diagnosis, the addressed method in this paper is more beneficial, as its performance is improved using HPO. Consequently, the objectives and contributions of this study are as follows: i) An efficient and accurate diagnosis approach for BC is proposed based on ML techniques utilizing HPO; ii) HPO utilizing GS and RS strategies is used to enhance the classification performance of ML models; iii) An investigational comparison of eight ML classifiers such as LR, k-NN, MLP, SVM, NB, XGB, random forest (RF), and DT with HPO utilizing GS and RS is demonstrated in this study using several experimental studies; and iv) The addressed approach showed considerable accuracy which can be integrated easily into computer-aided automated diagnosis tools for effective and reliable BC diagnosis in healthcare.

The remainder of this paper is structured as follows: section 2 summarizes related studies. The materials and suggested approach for BC diagnosis are explained in Section 3. The outcomes and findings of several experiments are covered in section 4. Lastly, section 5 provides the conclusion and plan for future study.

2. RELATED WORK

Data analytics integrated with the strength of ML attracted significant attention in numerous domains due to its superior ability in problem-solving. Numerous researchers are attracted to ML techniques due to their outstanding performance on high-dimensional data and robust scalability. A remarkable number of ML-based BC diagnosis and analysis tools for healthcare have been introduced by researchers worldwide in recent years, where several ML algorithms and techniques showed proficient performance.

Han and Yin [4] proposed a hybrid algorithm for BC prediction, where meta-learning models were used for developing an ANN model that obtained 98.74% accuracy. In this study, recursive feature elimination using cross-validation (RFECV), correlation, and tree-based feature selection techniques were used, and correlation performed better than others, which selected 16 features and achieved 96.49% accuracy. However, RFECV chose fewer features and the tree-based obtained less accuracy than others with 16 features. SVM, k-NN, DT, RF, LR, gradient boosting (GB), XGB, AdaBoost (AB), and MLP were utilized as meta learners, and outputs of these ML models were used as input of ANN that achieved an F1-score of 98.02%. Chen *et al.* [10] addressed an ML-based classification approach for BC, where the classification performance of XGB, RF, LR, and k-NN was analyzed. Pearson's correlation coefficient was used for selecting 15 highly correlated features to the target variable, and data standardization was utilized for eliminating the effect of the different dimensions on ML models. XGB performed better than others and obtained 97.40% accuracy, whereas RF, LR, and k-NN achieved 96.50%, 94.70%, and 91.20% accuracy, respectively. Moreover, the recall value of XGB, RF, and LR was 1, and all classifiers worked well on the training and test set in an 8:2 ratio than 7:3. Akkur *et al.* [11] proposed a hybrid feature selection technique for BC classification, where features were ranked using relief, and binary Harris hawk optimization was utilized for choosing highly discriminating features. The ten-fold cross-validation method was utilized, and SVM showed better performance than other ML models with a 94.56% specificity value. Using the addressed feature selection technique, 4 features were selected from 30 features, and SVM, LR, k-NN, and NB obtained 98.77%, 97.19%, 96.13%, and 97.37% accuracy, respectively. On the other hand, SVM, LR, k-NN, and NB obtained 94.73%, 91.92%, 91.56%, and 92.62% accuracy with 30 features, respectively. Naji *et al.* [12] proposed a diagnosis approach for BC based on ML, where SVM, RF, LR, DT, and k-NN were used. On the training set, RF showed better prediction efficiency and obtained 99.80% accuracy. On the other hand, SVM acquired 97.20% accuracy on the test set, whereas RF, LR, DT, and k-NN got 96.50%, 95.80%, 95.10%, and 93.70% accuracy, respectively. Moreover, SVM outperformed other ML models in terms of area under the ROC curve (AUC), achieving 96.60% AUC, and DT achieved the lowest AUC of 94.50%.

Gopal *et al.* [13] proposed MLP for BC prediction, which obtained 98.00% accuracy, and RF and LR showed less prediction efficiency than MLP. Omondiagbe *et al.* [14] addressed a hybrid approach for classifying BC, where linear discriminant analysis (LDA) was utilized for reducing the high dimensionality of 30 features. Principal component analysis (PCA) was also used for extracting features. For selecting highly important features, correlation-based feature selection (CFS) and recursive feature elimination (RFE) were used. Both SVM and neural network (NN) with LDA obtained 98.82% accuracy, where NN took more computational time than SVM. Using CFS, NN obtained higher accuracy than others, 97.06%, and SVM and NB obtained 96.47% and 91.76% accuracy, respectively. Using RFE, NN, SVM, and NB obtained 98.24%, 96.47%, and 91.18% accuracy, respectively. Bensaoucha [15] used optimized ML models for BC diagnosis, and the prediction performance of DT, NB, SVM, k-NN, and MLP was evaluated. BO and GS methods were used for optimizing ML classifiers, both SVM and MLP obtained 96.52% accuracy. Albadr *et al.* [16] addressed a fast learning network (FLN) for BC diagnosis that obtained 98.83%, 98.44%, and 99.05% accuracy, precision, and specificity, respectively. The performance of several hidden node numbers such as 25, 50, 75, 100, 125, 150, 175, and 200 of FLN were compared, where FLN with 25 hidden nodes got the highest accuracy. Ogundokun *et al.* [17] proposed an ML-based diagnosis approach for BC using hyperparameter optimization, where SVM, CNN, and MLP were utilized. Particle swarm optimization (PSO) was utilized to select features for SVM and MLP, and GS was used to identify the highly efficient combination of hyperparameters. Using PSO, SVM, and MLP obtained 96.50% and 97.20% accuracy, respectively. On the other hand, both SVM and MLP obtained 96.50% accuracy without PSO. To classify BC, Sukmandhani *et al.* [18] analyzed the classification performance of SVM, NB, k-NN, RF, DT, NN, and H2O models. Among ML models, RF obtained the highest accuracy which was 92.26%, and SVM, NB, KNN, and DT got 88.59%, 90.52%, 88.93%, and 90.50% accuracy, respectively. On the other hand, H2O and NN obtained 93.14% and 92.97% accuracy, respectively. NB got higher precision than others, and DL got the highest recall value, 93.89%, and 89.62%, respectively. According to the above-mentioned earlier studies, ML classifiers performed well in BC diagnosis but classifiers need more enhancement to ensure highly reliable performance. Therefore, this study is conducted for establishing a more effective and accurate diagnosis technique for BC.

3. RELATED WORK

This study is divided into five individual parts, and among these, three are the main parts where ML models were trained and evaluated under different conditions. The suggested architecture for accurate BC diagnosis is demonstrated in Figure 1. Firstly, data pre-processing is carried out on the WBCD dataset to prepare the dataset for training. Afterward, ML classifiers were trained using default parameters, and the trained models were evaluated using several performance metrics. Then, ML models were trained again using HPO by utilizing GS and RS. The model's accuracy after HPO was also evaluated. Finally, an ML model is suggested for classifying BC after analyzing the performance of all trained models in depth.

3.1. Dataset

A multivariate WBCD dataset used in this study contains 569 experimental instances with 32 real attributes [7]. The attributes were extracted from cell images, where instances are divided into two labels such as benign and malignant. Among 569 samples, benign (B) and malignant (M) sample numbers are 357 and 212, respectively. In this study, 70% of experimental samples were used for training and 30% for model evaluation purposes.

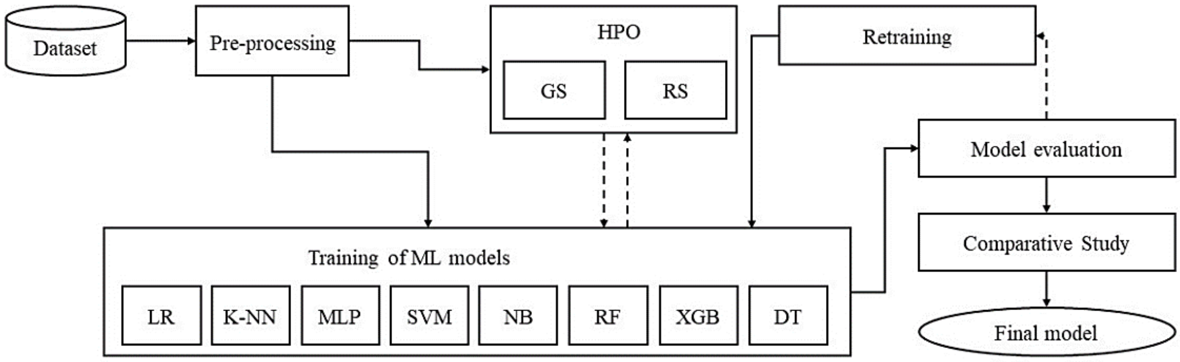

Figure 1. Suggested ML-based BC diagnosis architecture

3.2. Data pre-processing

In ML, data pre-processing is a crucial stage, which mainly focuses on the quality enhancement of data to make it more compatible with the required tasks. It significantly reduces the computing time of ML classifiers and also enhances the performance of classifiers. WBCD dataset contains a categorical attribute, namely, the ID number, which was dropped from the dataset. Afterward, data was normalized utilizing the StandardScaler function, which mainly follows the standard normal distribution technique that sets the mean equal to 0, and then data is scaled to unit variance [19].

3.3. ML classifiers

Several types of supervised ML algorithms were utilized in this study for the accurate diagnosis prediction of BC, including weak and strong classifiers. LR is a robust decision-making tool which mainly utilized for classification-related tasks, it works like a statistical algorithm that analyzes the interrelation between a group of independent and dependent binary variables [20]. k-NN is a non-parametric ML algorithm used for classification that does not provide any decision based on the underlying data [21]. MLP is an NN-based classification algorithm that consists of multiple layers which are fully connected, and it reshapes any input dimension to the required dimension [22]. SVM is widely used for solving classification problems of high-dimensional features, which mainly finds the maximal separating hyperplane between the several classes obtainable in the target feature [11]. NB is an effective probabilistic classifier that works on Bayes' theorem and provides prediction based on the possibility of an object, also used for building fast ML models to give quick predictions [11]. RF is a merged classifier containing a combination of DT, and each DT classifier works individually while making the classification. It provides faster training and less overfitting [4]. XGB is used for scalable and efficient training of ML models, which provides powerful predictions and is efficient in handling missing values. Moreover, it also supports parallel processing that helps to train classifiers on large datasets within a reasonable time [23]. DT generally consists of one root note and multiple inner and leaf nodes and is built by recursively dividing the full samples of data into several child nodes based on the attribute of tests. Easy structure and representation ease make it more powerful and beneficial for classification [24].

3.4. HPO

ML optimization is a challenging task that is now significantly used in ML-based computationally expensive and complex solutions. In ML, parameters are considered from the provided data by a model, whereas hyperparameters of a model are utilized for the estimation of model parameters. Hyperparameters of a model are defined manually to control the learning process by the user. HPO is a technique for determining the best combination of hyperparameters to maximize the performance of an ML model. HPO optimizes the model which minimizes the validation error [8].

In this study, firstly, the parameters of an ML model were analyzed, and then, searching methods such as GS and RS were applied. The model score is generated with the cross-validation technique. Finally, the best hyperparameters are selected after the evaluation of the model score. The method followed in this study for searching for the best hyperparameter is illustrated in Figure 2.

GS utilizes several combinations of all specified hyperparameters with values, and then evaluates the performance of every combination. Afterward, chooses the higher suitable value of hyperparameters, and this method of searching hyperparameters is expensive and time-consuming, which mainly depends on the hyperparameters number [25]. In this study, GridSearchCV is utilized with the WBCD dataset where GS and cross-validation were performed for evaluating the best suitable hyperparameters [26]. In RS, random hyperparameter combinations are selected and utilized for the training of an ML classifier. Based on performance, the best hyperparameter combination is selected [27]. RandomizedSearchCV was used in this study, which randomly delivers a hyperparameters set, and after calculating the score, provides a hyperparameters set that obtained higher scores than other sets.

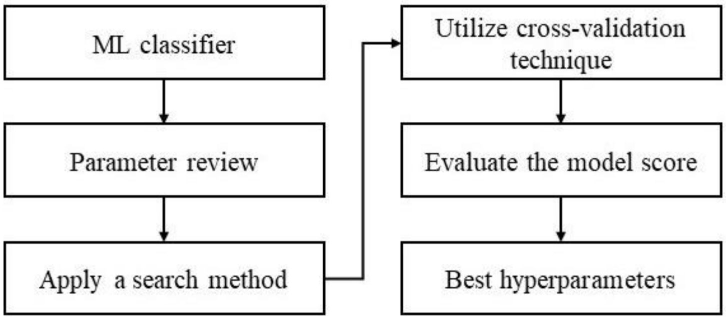

Figure 2. Steps for searching the best hyperparameter for HPO

3.5. Experiments

To find an efficient and accurate ML-based solution for BC diagnosis, this study conducted three different experiments using ML techniques. In the first experiment (Exp 1), parameters of ML algorithms were used to train ML models. To compare the performance of parameters with hyperparameters, Exp 1 is conducted on the WBCD dataset. On the other hand, GS was used for finding highly suitable hyperparameters for ML models in the second experiment (Exp 2). In the third experiment (Exp 3), RS was utilized to search for the best hyperparameters for ML models. Both Exp 2 and 3 were conducted on the WBCD dataset to evaluate and compare the efficiency of GS and RS. Finally, Exp 1, 2, and 3 were analyzed and compared in depth using several performance metrics to find out an optimal ML-based method for effective BC diagnosis. In addition, the average score of each performance metric of ML models for Exp 1, 2, and 3 were calculated to analyze the performance of Exp 1, 2, and 3 for BC diagnosis. All experimental studies were conducted in this study using the Jupyter Notebook of Google Colab with Python 3.10.12.

3.6. Performance matrices

ML models often produce different changes to meet several requirements of real-world applications, so five different performance metrics were used for determining model functionality under the requirements of real-world tasks [8]. Using a variety of performance metrics is crucial for selecting the best suitable model for real-world ML-based solutions. For evaluating the BC diagnosis performance of ML models through prediction, four performance estimation metrics such as accuracy, precision, recall, and F1-score were utilized in this study based on true positive (TP), true negative (TN), false positive (FP), and false negative (FN), which were obtained from confusion matrix of each ML model on the test set of WBCD dataset [23].

Accuracy presents the overall performance of a model, which is the ratio of accurately diagnosed samples among total samples [28]. Precision demonstrates a model's performance based on the correct diagnosis, which represents the percentage of really positive samples among positive predicted samples [29]. The recall is the ratio of the correctly predicted positive samples to all predicted positive samples in the original class, which is also called the rate of true positive [10], [15]. F1-score is the harmonic average of the pre and rec index, which is always less than the value of accuracy as it is calculated with precision and recall value [10], [17].

For comparing the performance of ML models, recall is considered the principal evaluation index in this paper. The higher recall value of a model indicates that it accurately predicted a higher proportion of M BC samples [10]. The arithmetic representation of used evaluation metrics is given in between (1) and (4). To measure the average score of each evaluation index for an ML model i, (5) was used in this study, where tm and piv refer to the total number of models and obtained performance index value, respectively. The average score of each evaluation index of seven used ML models is measured for Exp 1, 2, and 3.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (1)$$

$$Precision = \frac{TP}{TP+FP} \times 100\% \quad (2)$$

$$Recall = \frac{TP}{TP+FN} \times 100\% \quad (3)$$

$$\text{F1} - \text{score} = 2 \times \frac{Precision \times Recall}{Precision+Recall} \times 100\% \quad (4)$$

$$Average\ score\ = \frac{1}{tm} \sum_{i=1}^{tm} piv(i) \quad (5)$$

4. RESULTS AND DISCUSSION

In this study, three different experiments such as Exp 1, 2, and 3 were utilized, and the performance of eight ML models was evaluated on the WBCD dataset for finding the best effective prediction technique for BC. Parameters, GS, and RS-based hyperparameters were used for the training of ML models. The overall performance of ML models over three different experiments is demonstrated in Table 1. In parameter-based experiments, both LR and XGB obtained higher accuracy than other models, 98.25%. NB obtained the lowest accuracy in Exp 1, which was 93.57%. In Exp 2 where HPO was utilized using the best hyperparameters selected by GS, both LR and k-NN acquired 99.42% accuracy, which was higher than other models. NB did not perform well in Exp 2 and obtained 95.32% accuracy which was less than others. NB did not perform well in Exp 2 and obtained 95.32% accuracy which was less than others. In Exp 3 where RS was utilized to find the best hyperparameters for HPO of ML models, both LR and MLP obtained the highest accuracy, 99.42%. Like Exp 1 and 2, NB also obtained less accuracy than other models in Exp 3.

HPO with GS showed a significant achievement in this study, where the accuracy of all models was increased except XGB. In RS-based HPO, the accuracy of LR, k-NN, MLP, and RF models was increased. In this study, XGB did not perform well in both Exp 2 and 3 and showed the best performance on parameter-based training of ML models. In Exp 1, XGB obtained the highest precision value, and DT obtained less than others, 98.39%, and 89.55%, respectively. LR obtained higher recall than other ML models, which was 98.41%. k-NN got the lowest recall value, 88.71%. In terms of the F1-score, LR performed well than other models and obtained a 97.64% F1-score. On the other hand, the F1-score of NB was less than others, 91.20%. The value of the evaluation indexes obtained by ML models in Exp 1 is presented in Table 2.

Table 1. The performance of ML models for three different experiments

Models	Parameters (%)	GS+HPO (%)	RS+HPO (%)
LR	98.25	99.42	99.42
k-NN	95.32	99.42	96.49
MLP	97.66	98.25	99.42
SVM	97.66	98.83	97.66
NB	93.57	95.32	92.98
XGB	98.25	97.08	96.49
RF	96.49	97.08	97.08
DT	94.15	97.08	93.57

Table 2. The statistical outcomes of ML models for Exp 1

Models	Precision (%)	Recall (%)	F1-score (%)
LR	96.88	98.41	97.64
k-NN	98.21	88.71	93.22
MLP	96.83	96.83	96.83
SVM	96.83	96.83	96.83
NB	91.94	90.48	91.20
XGB	98.39	96.83	97.60
RF	96.72	93.65	95.16
DT	89.55	95.24	92.31

In Exp 2, both LR and SVM obtained the highest precision value, 100.0%. k-NN received the highest recall value, 100.0%. LR and k-NN both got a 99.20% F1-score, which was higher than other models. However, NB obtained less precision, recall, and F1-score than other models, 95.08%, 92.06%, and 93.55%, respectively. The value of the evaluation indexes obtained by ML models in Exp 2 is presented in Table 3. In Exp 3, LR and MLP got the highest precision value, 100.0%. LR, MLP, and SVM acquired a 98.41% recall value, which was higher than others. k-NN obtained a 95.16% recall value in Exp 3, whereas it obtained a 100.0% recall value in Exp 2. Both LR and MLP got higher F1-score than other ML models, 99.20%. NB obtained less precision, recall, and F1-score than other models, 91.80%, 88.89%, and 90.32%, respectively. The value of the evaluation indexes obtained by ML models in Exp 3 is presented in Table 4.

For evaluating the performance of Exp 1, 2, and 3 more evidently, the average of evaluation indexes was measured for all experiments in this study, as presented in Table 5. According to the findings of this evaluation, Exp 2 outperformed others which got higher average precision, recall, and F1-score, 97.79%, 96.81%, and 98.99%, respectively. However, Exp 1 received a higher average recall than Exp 3, which was 94.62%. On the other hand, Exp 3 obtained higher average precision and F1-score than Exp 1.

To demonstrate the performance of ML models over Exp 1, 2, and 3 more clearly, the confusion matrix of LR, k-NN, MLP, and RF models is presented in Figure 3.

Considering the peculiarity of the BC diagnosis toll for healthcare, it is anticipated that all BC samples of M can be identified accurately by the prediction technique. To evaluate an ML model's efficiency in diagnosing BC samples of M, recall can play a vital role, which delivers the rate of predicted true positives. Therefore, recall is considered a crucial performance metric to evaluate ML models for BC diagnosis in this paper. If an ML model got a higher recall value, it indicates that a higher number of M BC samples is predicted accurately. In Exp 2, k-NN obtained a 100.0% recall value, demonstrating that k-NN accurately diagnosed all M BC samples. Except for k-NN, no other ML models got a 100.0% recall value, which strongly validates that k-NN is more suitable and effective than other models for BC diagnosis. After k-NN, the recall value of LR, SVM, and MLP was higher than the rest models. On the other hand, the average recall value of Exp 2 was significantly higher than Exp 1 and 3, which were 2.19% and 2.38%, respectively.

Table 3. The statistical outcomes of ML models for Exp 2

Models	Precision (%)	Recall (%)	F1-score (%)
LR	100.0	98.41	99.20
k-NN	98.41	100.0	99.20
MLP	96.88	98.41	97.64
SVM	100.0	96.83	98.39
NB	95.08	92.06	93.55
XGB	95.31	96.83	96.06
RF	98.33	98.33	95.93
DT	98.33	93.65	95.93

Table 4. The statistical outcomes of ML models for Exp 3

Models	Precision (%)	Recall (%)	F1-score (%)
LR	100.0	98.41	99.20
k-NN	95.16	95.16	95.16
MLP	100.0	98.41	99.20
SVM	95.38	98.41	96.88
NB	91.80	88.89	90.32
XGB	96.72	93.65	95.16
RF	98.33	93.65	95.93
DT	93.33	88.89	91.06

Table 5. Average values of evaluation indexes for all experiments

Experiments	Avg. precision (%)	Avg. recall (%)	Avg. F1-score (%)
Exp 1	95.67	94.62	95.10
Exp 2	97.79	96.81	98.99
Exp 3	96.34	94.43	95.36

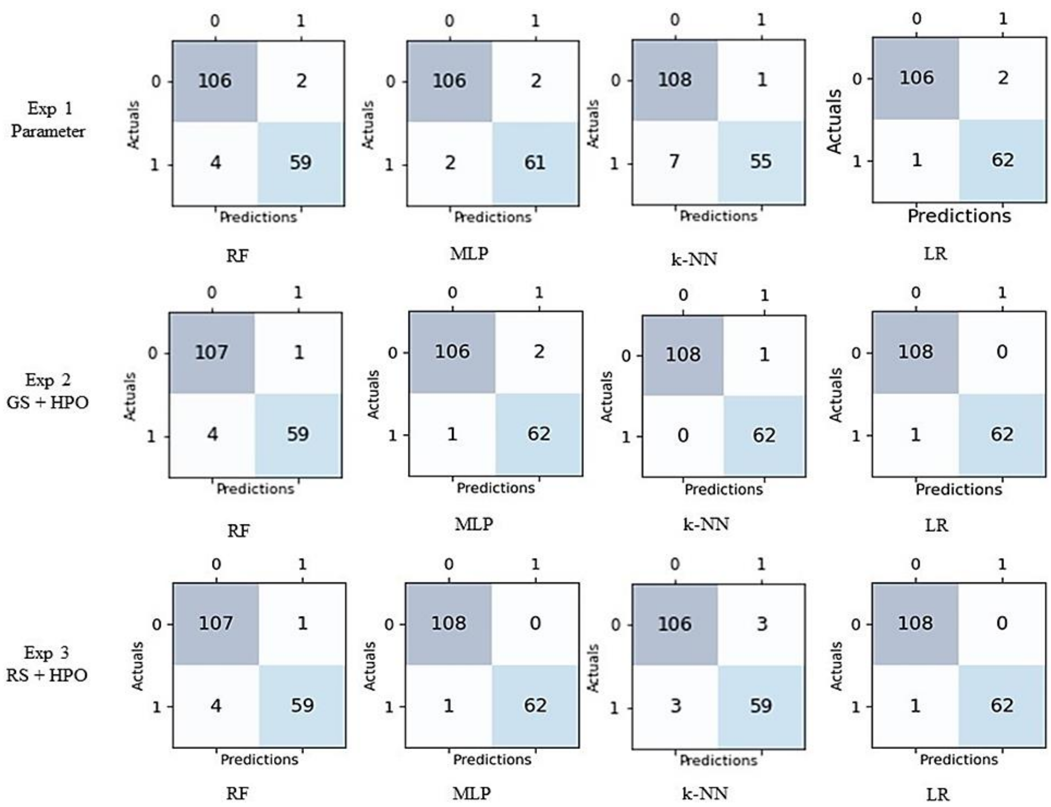

Figure 3. Confusion matrix of ML models (1: M, 0: B)

Based on these findings mentioned earlier, a framework where k-NN is trained with hyperparameters found by GS is suggested in this paper for accurate BC diagnosis. In addition, the LR model performed well in this study in all three experiments in terms of accuracy and NB got less accuracy in all experiments. However, the performance of XGB was decreased after utilizing HPO in this study. By inspiring the increasing demand for automated disease diagnosis toll in healthcare, the designed framework mentioned above is addressed in this study after several evaluations of effectiveness to aid doctors or medical experts in timely providing better treatments and plans to BC patients. Moreover, the outcome of this study is also compared based on recall and accuracy with recent studies conducted on the WBCD dataset for BC diagnosis which is illustrated in Table 6. Chen *et al.* [10] proposed XGB for BC diagnosis, which also got a 100.0% recall value, but it obtained 2.02% less accuracy than ours, which clearly indicates that the addressed technique in this paper is more effective and suitable for BC diagnosis than similar studies conducted in the literature.

Table 6 Comparison of the outcome with recent studies in the literature

Author	Model	Recall (%)	Accuracy (%)
Han and Yin [4]	ANN	98.41	98.74
Chen *et al.* [10]	XGB	100.0	97.40
Akkur *et al.* [11]	SVM	99.52	98.77
Naji *et al.* [12]	SVM	*NG*[1]	97.20
Gopal *et al.* [13]	MLP	97.00	98.00
Omondiagbe *et al.* [14]	SVM	98.41	98.82
Bensaoucha [15]	SVM, MLP	*NG*	96.52
Albadr *et al.* [16]	FLN	99.40	98.83
Ogundokun *et al.* [17]	MLP	97.80	97.20
Sukmandhani *et al.* [18]	H2O	89.62	93.14
Our study	k-NN	100.0	99.42

NG[1]: Not given

5. CONCLUSION

BC is one of the leading reasons for women's deaths worldwide, mostly due to the lack of timely diagnosis and appropriate treatment. A variety of ML-based approaches have been introduced for BC diagnosis in recent years. In this study, three different experiments were conducted using parameters and hyperparameters on the WBCD dataset to find out an effective technique for BC diagnosis. The prediction performance of eight ML classifiers was evaluated using different evaluation indexes, where k-NN obtained 100.0% recall and 99.42% accuracy with GS-based HPO. HPO demonstrated a noteworthy performance in this study, which increased the accuracy of all ML models except XGB. HPO increased the recall value of k-NN from 88.71 % to 100.0%, which strongly validates the effectiveness of HPO in BC diagnosis. Among three experiments, Exp 2 showed significant performance and obtained the highest average score of all evaluation indexes, which utilized GS-based hyperparameters to train ML models. According to the outcomes of several evaluation studies, k-NN trained with GS-based hyperparameters is proposed in this paper to design an automated system for BC diagnosis. In future work, the scope of this study will be expanded by introducing more ML and DL-based techniques with large datasets to develop a fully efficient and robust diagnosis technique for BC. Moreover, a complete internet of things (IoT) based diagnosis system will be designed and evaluated for healthcare to reduce diagnosis expenses and errors.

REFERENCES

[1] S. Łukasiewicz, M. Czeczelewski, A. Forma, J. Baj, R. Sitarz, and A. Stanisławek, "Breast cancer—epidemiology, risk factors, classification, prognostic markers, and current treatment strategies—An updated review," *Cancers*, vol. 13, no. 17, Aug. 2021, doi: 10.3390/cancers13174287.
[2] WHO, "WHO launches new roadmap on breast cancer," *World Health Organization*, 2023. https://www.who.int/news/item/03-02-2023-who-launches-new-roadmap-on-breast-cancer (accessed Jan. 20, 2023).
[3] WHO, "Breast cancer," *World Health Organization*, 2023. https://www.who.int/news-room/fact-sheets/detail/breast-cancer (accessed Jan. 22, 2023).
[4] L. Han and Z. Yin, "A hybrid breast cancer classification algorithm based on meta-learning and artificial neural networks," *Frontiers in Oncology*, vol. 12, Nov. 2022, doi: 10.3389/fonc.2022.1042964.
[5] A. A. Abdul Halim *et al.*, "Existing and emerging breast cancer detection technologies and its challenges: a review," *Applied Sciences*, vol. 11, no. 22, Nov. 2021, doi: 10.3390/app112210753.
[6] M. Javaid, A. Haleem, R. Pratap Singh, R. Suman, and S. Rab, "Significance of machine learning in healthcare: Features, pillars and applications," *International Journal of Intelligent Networks*, vol. 3, pp. 58–73, 2022, doi: 10.1016/j.ijin.2022.05.002.
[7] G. Luo, "A review of automatic selection methods for machine learning algorithms and hyper-parameter values," *Network Modeling Analysis in Health Informatics and Bioinformatics*, vol. 5, no. 1, May 2016, doi: 10.1007/s13721-016-0125-6.

[8] L. Yang and A. Shami, "On hyperparameter optimization of machine learning algorithms: Theory and practice," *Neurocomputing*, vol. 415, pp. 295–316, Nov. 2020, doi: 10.1016/j.neucom.2020.07.061.
[9] A. Frank, "UCI machine learning repository," UCI. http://archive.ics.uci.edu/ml (accessed Dec. 25, 2021).
[10] H. Chen, N. Wang, X. Du, K. Mei, Y. Zhou, and G. Cai, "Classification prediction of breast cancer based on machine learning," *Computational Intelligence and Neuroscience*, vol. 2023, pp. 1–9, Jan. 2023, doi: 10.1155/2023/6530719.
[11] E. Akkur, F. Türk, and O. Eroğul, "Breast cancer classification using a novel hybrid feature selection approach," *Neural Network World*, vol. 33, no. 2, pp. 67–83, 2023, doi: 10.14311/NNW.2023.33.005.
[12] M. A. Naji, S. El Filali, K. Aarika, E. H. Benlahmar, R. A. Abdelouhahid, and O. Debauche, "Machine learning algorithms for breast cancer prediction and diagnosis," *Procedia Computer Science*, vol. 191, pp. 487–492, 2021, doi: 10.1016/j.procs.2021.07.062.
[13] V. N. Gopal, F. Al-Turjman, R. Kumar, L. Anand, and M. Rajesh, "Feature selection and classification in breast cancer prediction using IoT and machine learning," *Measurement: Journal of the International Measurement Confederation*, vol. 178, Jun. 2021, doi: 10.1016/j.measurement.2021.109442.
[14] D. A. Omondiagbe, S. Veeramani, and A. S. Sidhu, "Machine learning classification techniques for breast cancer diagnosis," *IOP Conference Series: Materials Science and Engineering*, vol. 495, no. 1, Jun. 2019, doi: 10.1088/1757-899X/495/1/012033.
[15] S. Bensaoucha, "Breast cancer diagnosis using optimized machine learning algorithms," *2021 International Conference on Recent Advances in Mathematics and Informatics (ICRAMI), Tebessa, Algeria, 2021, pp. 1-6*, doi: 10.1109/ICRAMI52622.2021.9585977.
[16] M. A. A. Albadr, M. Ayob, S. Tiun, F. T. AL-Dhief, A. Arram, and S. Khalaf, "Breast cancer diagnosis using the fast learning network algorithm," *Frontiers in Oncology*, vol. 13, Apr. 2023, doi: 10.3389/fonc.2023.1150840.
[17] R. O. Ogundokun, S. Misra, M. Douglas, R. Damaševičius, and R. Maskeliūnas, "Medical internet-of-things based breast cancer diagnosis using hyperparameter-optimized neural network," *Future Internet*, vol. 14, no. 5, May 2022, doi: 10.3390/fi14050153.
[18] A. A. Sukmandhani, Lukas, Y. Heryadi, W. Suparta, and A. Wibowo, "Classification algorithm analysis for breast cancer," *E3S Web of Conferences*, vol. 388, 2023, doi: 10.1051/e3sconf/202338802012.
[19] L. B. V. de Amorim, G. D. C. Cavalcanti, and R. M. O. Cruz, "The choice of scaling technique matters for classification performance," *Applied Soft Computing*, vol. 133, Jan. 2023, doi: 10.1016/j.asoc.2022.109924.
[20] A. G, B. Ganesh, A. Ganesh, C. Srinivas, Dhanraj, and K. Mensinkal, "Logistic regression technique for prediction of cardiovascular disease," *Global Transitions Proceedings*, vol. 3, no. 1, pp. 127–130, Jun. 2022, doi: 10.1016/j.gltp.2022.04.008.
[21] M. A. jabbar, B. L. Deekshatulu, and P. Chandra, "Classification of heart disease using k-nearest neighbor and genetic algorithm," *Procedia Technology*, vol. 10, pp. 85–94, 2013, doi: 10.1016/j.protcy.2013.12.340.
[22] M. A. Moreno-Ibarra, Y. Villuendas-Rey, M. D. Lytras, C. Yáñez-Márquez, and J. C. Salgado-Ramírez, "Classification of diseases using machine learning algorithms: a comparative study," *Mathematics*, vol. 9, no. 15, Jul. 2021, doi: 10.3390/math9151817.
[23] K. Budholiya, S. K. Shrivastava, and V. Sharma, "An optimized XGBoost based diagnostic system for effective prediction of heart disease," *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 7, pp. 4514–4523, Jul. 2022, doi: 10.1016/j.jksuci.2020.10.013.
[24] M. M. Ali, B. K. Paul, K. Ahmed, F. M. Bui, J. M. W. Quinn, and M. A. Moni, "Heart disease prediction using supervised machine learning algorithms: performance analysis and comparison," *Computers in Biology and Medicine*, vol. 136, Sep. 2021, doi: 10.1016/j.compbiomed.2021.104672.
[25] M. Feurer and F. Hutter, "Hyperparameter optimization," in *The Springer Series on Challenges in Machine Learning*, Springer International Publishing, 2019, pp. 3–33.
[26] D. M. Belete and M. D. Huchaiah, "Grid search in hyperparameter optimization of machine learning models for prediction of HIV/AIDS test results," *International Journal of Computers and Applications*, vol. 44, no. 9, pp. 875–886, Sep. 2022, doi: 10.1080/1206212X.2021.1974663.
[27] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization," *Journal of Machine Learning Research*, vol. 13, pp. 281–305, 2012.
[28] M. A. Elsadig, A. Altigani, and H. T. Elshoush, "Breast cancer detection using machine learning approaches: a comparative study," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 13, no. 1, pp. 736–745, Feb. 2023, doi: 10.11591/ijece.v13i1.pp736-745.
[29] R. H. Hridoy, A. D. Arni, and A. Haque, "Improved vision-based diagnosis of multi-plant disease using an ensemble of deep learning methods," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 13, no. 5, pp. 5109–5117, Oct. 2023, doi: 10.11591/ijece.v13i5.pp5109-5117.

BIOGRAPHIES OF AUTHORS

Rashidul Hasan Hridoy received the B.Sc. degree in computer science and engineering from Daffodil International University, Daffodil Smart City, Dhaka, Bangladesh, in 2021. Currently, he is a lecturer in the Department of Software Engineering, at Daffodil International University. Previously, he worked as a lecturer at the Department of Computer Science and Engineering, at Daffodil International University. He also works as a peer reviewer in several international journals. His research interests include computer vision, machine learning, natural language processing, speech recognition, biomedical image processing, bioinformatics and computational biology, and design of scalable algorithms. He is currently engaged in several research projects for ensuring sustainability in healthcare and agriculture. He published his research outputs in various peer-reviewed international journals and conferences. He can be contacted at email: rashidulhasanhridoy@gmail.com.

Arindra Dey Arni completed her B.Sc. degree in computer science and engineering from Daffodil International University, Daffodil Smart City, Dhaka, Bangladesh, in 2021. Currently, she is involved in different computer vision-based projects, which are introduced for developing efficient and rapid approaches for the healthcare and agricultural industries to ensure sustainability. She published several research articles in different peer reviewed international journals and conferences. She is interested in machine learning, computer vision, cryptography, steganography, human-computer interaction, robotics, and cyber security. She can be contacted at email: arindradeyarni28@gmail.com.

Shomitro Kumar Ghosh is currently a postgraduate student at Trine University, Detroit, MI, USA, and studying information science. He completed his B.Sc. degree in computer science and engineering from Daffodil International University, Daffodil Smart City, Dhaka, Bangladesh, in 2020. His research interests encompass human-computer interaction, human-centered design, computer vision, machine learning, augmented reality, virtual reality, and neural networks. Currently, he is involved in different research projects for ensuring sustainability in the healthcare and agricultural sectors. He published his research outputs in several peer reviewed international journals and conferences. He can be contacted at email: shomitro.kumar19@gmail.com.

Narayan Ranjan Chakraborty is working as an associate professor at the Department of Computer Science and Engineering, Daffodil International University, Daffodil Smart City, Dhaka, Bangladesh. His research interests include topics such as cyber security, information security, text mining, machine learning, and natural language processing. He has published his research outcomes in several international peer-reviewed journals and conferences. He can be contacted at email: narayan@daffodilvarsity.edu.bd.

Imran Mahmud received his Ph.D. in Technology Management from Universiti Sains Malaysia in 2017. In 2008. Dr. Mahmud completed Masters in Software Engineering from the University of Hertfordshire, UK. He is currently working as head and associate professor at the Department of Software Engineering at Daffodil International University, Daffodil Smart City, Dhaka, Bangladesh. He is also working as a visiting professor at the Graduate School of Business at Universiti Sains Malaysia. Previously, he worked as a senior lecturer at the Graduate School of Business at Universiti Sains Malaysia. Dr. Imran was also a visiting lecturer at the Institute of Technology, Bandung (Indonesia) and Hong Kong Management Association (Hong Kong). At the moment, Dr. Mahmud is working with technostress, games addiction, and online learning continuance intention. He can be contacted at email: imranmahmud@daffodilvarsity.edu.bd.