International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Diagnosis of Breast Cancer Using Improved Machine Learning Algorithms Based on Bayesian Optimization

Zeynep CEYLAN[1]

Submitted: 19/07/2020 Accepted : 07/08/2020

Abstract: Breast cancer is one of the most common types of cancer and is the second main cause of cancer death in females. Early detection of breast cancer is crucial for the survival of a patient as well as for the quality of life throughout cancer treatment. The aim of this study is to develop improved machine learning models for early diagnosis of breast cancer with high accuracy. In this context, a performance comparison of machine learning algorithms including Support Vector Machines, Decision Trees, Naive Bayes, K-Nearest Neighbor, and Ensemble Classifiers was performed on a dataset consisting of routine blood analysis combined with anthropometric measurements to diagnose breast cancer. Neighborhood component analysis was applied as a feature selection method to reveal relevant biomarkers that can be used in breast cancer prediction. In order to assess the performance of each proposed classifier model, two different data division procedures such as hold-out and 10-fold cross-validation were employed. Bayesian Optimization algorithm was applied to all classifiers for the maximizing the prediction accuracy. Different performance criteria such as accuracy, precision, sensitivity, specificity, and F-measure were used to measure the success of each classifier. Experimental results showed that the Bayesian optimization-based K-Nearest Neighbor performs better than other machine learning algorithms under the hold-out data division protocol with an accuracy of 95.833%. The results obtained in this study may provide a new perspective on the application of improved machine learning techniques for the early detection of breast cancer.

Keywords: *Bayesian optimization, breast cancer, classifier, feature selection, machine learning*

1. Introduction

Cancer is the second leading cause of death worldwide with 9.6 million deaths per year. There are many types of cancer, including prostate, lung, liver, ovarian and breast and so on, depending on the organ it penetrates in the body. According to the World Health Organization (WHO) report, among the 18.1 million cancer cases detected worldwide in 2018, breast cancer ranks second after lung cancer with a rate of 11.6%. It is responsible for 33% of all cancers in women and 20% of cancer-related deaths [1].

Breast cancer is a disease caused by the change and uncontrolled proliferation of one of the cell groups that make up the breast tissue. The proliferation and growth of cells that cause breast cancer take quite some time. However, after proliferation, the cells can spread to other organs of the body through lymph and blood. The most important factor in breast cancer is the diagnosis of cancer before it spreads to the blood and lymphatic organs. The success of treatment with a diagnosis made at this stage is very high. Therefore, as with all cancer types, early diagnosis and correct treatment of breast cancer are crucial to reduce mortality.

There are many methods for early diagnosis of the disease, depending on the age of the patient. For example, in the twenties, the patient can diagnose the disease by breast control on his own, while a doctor's examination is mandatory at an older age. In the case of stiffness and swelling in the breast tissue, clinical breast examination is necessary. Clinical examination is performed by ultrasonography, fine needle biopsy, and mammography techniques [2]. Another diagnostic method is a computer-assisted diagnosis, a decision support system used to determine whether a person has breast cancer.

Machine learning (ML) techniques are used commonly to assist specialists and doctors to make the right decision in computer-assisted diagnostic studies and contribute to the development of health practices [3–7]. ML is a branch of artificial intelligence (AI) that includes a variety of statistical, probabilistic, and optimization techniques, and allows computers to quickly identify patterns within complex and large data sets by learning from existing data. ML techniques have been studied for a long time in breast cancer data sets that are publicly available in the UC Irvine Machine Learning Repository [8]. In particular, Wisconsin Breast Cancer Dataset (WBCD), the Wisconsin Prognosis Breast Cancer (WPBC), and the Wisconsin Diagnosis Breast Cancer (WDBC) which provides size and other characteristic data for tumors, has been widely studied [9].

Recent studies have shown that a routine blood test that evaluates the body's immune response to the substances produced by tumor cells can be used as an effective early detection method. The medical findings indicated that diagnoses made using this type of dataset can detect earlier than breast cancer diagnosis based on the size and other characteristics of the tumor, and may also be an important reference to new researches, such as the relationship between obesity and cancer [10]. In recent years, many different ML approaches have been tried in the literature for the detection of breast cancer using routine blood and hormone data combined with anthropometric measurements such as body mass index, age, visfatin, leptin, insulin, resistin, adiponectin and etc.

The input data used in various studies and the number of subjects examined are summarized in Table 1.

[1]*Department of Industrial Engineering, Samsun University, Samsun, Turkey; ORCID ID: 0000-0002-3006-9768*
**Corresponding Author E-mail: zeynep.ceylan@samsun.edu.tr*

For example, Kang et al. [11] evaluated the relationship of between serum adiponectin and resistin level with breast cancer risk in biopsy-proven breast cancer patients. Hwa et al. [12] developed logistic regression (LR) models using new predictors such as serum levels of tissue polypeptide-specific antigen, breast cancer-specific cancer antigen 15.3 (CA15–3), and insulin-like growth factor binding protein-3 (IGFBP-3).

Santillán-Benítez et al. [13] analyzed serum levels of leptin, adiponectin and CA 15-3, as well as anthropometric and biochemical parameters as biomarkers for breast cancer. Provatopoulou et al [14] investigated the role of irisin in breast cancer quantitatively determining serum levels of irisin in patients with invasive ductal breast cancer and healthy individuals. Assiri et al. [15] investigated the correlation of resistin, visfatin, adiponectin, and leptin with BC risk in pre- and postmenopausal females using multivariate logistic regression (MLR) analysis. Assiri and Kamel [16] examined the serum leptin, resistin and visfatin levels as risk factors for postmenopausal breast cancer. As shown in Table 1, extensive studies have been conducted on the Breast Cancer Coimbra (BCC) dataset in recent years [4,17–21]. Patrício et al. [17] constructed LR, random forest (RF), and support vector machine models (SVM) models using glucose, age, resistin, and BMI features.

Li [18] used five different classification models including RF, SVM, decision tree (DT), artificial neural network (ANN) and logistics regression (LR) for breast cancer detection using glucose, insulin, homeostasis model assessment (HOMA), chemokine monocyte chemoattractant protein 1 (MCP-1), leptin, adiponectin, resistin, age and BMI data. Differently from this study, Aslan et al. [19] developed an extreme learning machine (ELM) and KNN models for classification using the same dataset. Singh [4] improved various machine learning models with different feature selection and data division strategies using clinical and anthropometric features.

Akben [20] applied the DT algorithm to the BCC dataset to determine the value ranges of data. Silva Araújo et al. [21] first divided the variables in the BCC dataset into 6 clusters to determine which of the most appropriate factors, and then applied fuzzy neural network (FNN) models to clusters for the prediction of breast cancer.

Each of the above-mentioned ML algorithms contains some set of hyperparameters that play an important role in the algorithm performance. In order to obtain excellent performance, these parameters need to be set carefully. To the best of the authors' knowledge, it was determined that the models developed in the previous studies on the BCC dataset did not perform an appropriate optimization technique.

In view of this, in this study, different ML models have been improved with the Bayesian optimization (BO) based approach to improve model accuracy. Recently, BO developed as an effective parameter optimization tool used in a wide range of applications because it effectively explores the possible hyperparameter area and manages a large set of experiments for hyperparameter settings [22,23]. Considering the benefits provided by the BO algorithm, the contribution of this study is outlined as follows;

i. The feasibility of new hybrid models based on ML algorithms and BO in order to diagnose breast cancer at an earlier stage is investigated using a recent dataset on breast cancer.
ii. The performance results of the individual ML models are compared with both the improved models by BO algorithm and thus the importance of parameter optimization is demonstrated.
iii. A comprehensive comparative analysis between the developed ML models and other methods used in the literature is conducted.

Table 1. Datasets and the number of patients used in studies

Reference	***Dataset***	***Study Populations (Breast Cancer /Control)***	***Anthropometric and laboratory values***
[2]	Coimbra University Hospital Centre, Portugal[1]	(64/52)	Age, BMI, Glucose, Insulin, HOMA, Leptin, Adiponectin, Resistin, and MCP-1
[3]	Konyang University Hospital, Korea	(41/43)	Glucose, Age, BMI, Resistin, Menopausal Status, Adiponectin
[4]	Medical Centre, Taiwan	(55/39)	Age, Menopausal status, BMI, HRT, CEA, CA 15-3, TPS, sIL-2R, and IGFBP-3
[5]	Maternal Perinatal Hospital Mónica Pretelini, Mexico	(40/48)	BMI, Leptin, L/A ratio, CA 15-3, Age, Weight, Height, WC, Glucose, Cholesterol, Triglycerides, HDLC, LDLC, Uric acid, Haemoglobin, Adiponectin
[6]	Hippokratio Hospital, Athens, Greece	(101/51)	Age, Menopausal status, BMI, BMI Status, Irisin, leptin, adiponectin, resistin, CEA, CA 15–3, and Her2/neu
[7]	King Abdulla Medical City and El-Noor Hospital, Makkah, KSA	(82/68)	Age, BMI, Resistin, Visfatin, WC, SBP, Adiponectin, Leptin, DBP, Glucose, TC, TG, LDLC, HDLC.
[8]	King Abdulla Medical City and El-Noor Hospital, Makkah, KSA	(209[2]/89)	Age, Nulliparus status, WC, hsCRP, CA 15-3, Leptin, Resistin and Visfatin
[9]	University Hospital Centre of Coimbra, Portugal	(64/52)	Resistin, Glucose, Age and BMI
[10]	Coimbra University Hospital Centre, Portugal[2]	(64/52)	Age, BMI, Glucose, Insulin, HOMA, Leptin, Adiponectin, Resistin, and MCP-1
[11]	Coimbra University Hospital Centre, Portugal[2]	(64/52)	Age, BMI, Glucose, Insulin, HOMA, Leptin, Adiponectin, Resistin, and MCP-1
[12]	Coimbra University Hospital Centre, Portugal[2]	(64/52)	Age, BMI, Glucose, Insulin, HOMA, Leptin, Adiponectin, Resistin, and MCP-1
[13]	Coimbra University Hospital Centre, Portugal[2]	(64/52)	Resistin, Glucose, Age and BMI

BMI, body mass index; CA15-3, breast cancer-specific cancer antigen 15.3; CEA, carcinoembryonic antigen; HRT, hormone replacement therapy; IGFBP-3, insulin-like growth factor binding protein-3; sIL-2R, soluble interleukin-2 receptor; TPS, tissue polypeptide-specific antigen, L/A, leptin/adiponectin; WC; waist circumference, HDLC, high-density lipoprotein cholesterol; LDLC, low-density lipoprotein cholesterol; SBP, systolic blood pressure; DBP, diastolic blood pressure; TC, total cholesterol; hsCRP, high sensitive c reactive protein; HOMA, homeostasis model assessment; MCP-1; chemokine monocyte chemoattractant protein 1.

[1]BCC dataset which is taken from UCI Machine Learning Repository, [2]The data includes 99 females with benign breast lesion

2. Material and Methods

2.1. Dataset

In this study, the BCC dataset taken from the UCI ML Repository was used [8]. The BCC dataset includes blood, hormone, demographic and anthropometric characteristics of the 64 breast cancer patient and 52 healthy volunteers.

The clinical features of all participants in this dataset were observed and measured under similar conditions, by the same research physician and during the initial consultation in the Coimbra University Hospital between 2009-2013. The diagnosis of breast cancer was first determined according to mammography results and then histologically confirmed by expert physicians. On the other hand, expert medical doctors have approved that there is no infection or other acute diseases or comorbidities in the non-cancer (control) group. The statistical characteristics of the attributes used in this study can be seen in Table 2 and detailed information of the dataset can be found in the study by Patrício et al [17]. As seen in Table 2, each parameter has different scales and measurement ranges. At the pre-processing step, normalization is performed to make the features comparable and to prevent any features from having more impact on the classification task than others. Thus, by the input data was scaled between 0 and 1 by applying the max-min normalization as given in Eq. [1].

$$d_{norm} = \left[\frac{d_i - d_{min}}{d_{max} - d_{min}}\right] \quad (1)$$

where d_{norm} is the normalized data, d_i represents the experimental value of value for i_{th} data point, d_{max} and d_{min} show the maximum and minimum values of the data, respectively. Hold-out and 10-fold cross-validation were used as data division strategy. In the hold-out division, the dataset was randomly divided into two groups with 30 % held out.

In addition, training and testing processes were performed 5 times for both data division strategy and then average performance was calculated. The analysis was performed with nine different inputs (Table 2) and then feature selection was made using neighbourhood component analysis (NCA) for classification. NCA is a non-parametric feature selection method that is used to maximize the prediction accuracy of classification or regression algorithms [24].

2.2. Machine Learning Techniques for Classification

Classification techniques are generally divided into two categories namely, supervised ML and unsupervised ML. The main difference between these two learning algorithms is whether the samples given to the learning algorithm are labelled or not. In supervised ML algorithms, the training set is provided with class labels, while unsupervised ML algorithms are applied to unlabelled samples [25]. In this study, popular supervised ML algorithms namely SVM, KNN, NB, DT, and ensemble classifiers (EC) were tested and compared. BO algorithm was performed to automatically adjust the hyperparameter values of the methods used. A brief description of these classifiers is presented in the following sections.

2.2.1. Support Vector Machine

Support Vector Machine (SVM) also called kernel machine is a supervised ML algorithm that can be used for both classification or regression problems [26]. It is widely used in many different applications such as nonlinear time series predictions and financial forecasting, natural language processing, speech and image recognition, monitoring network design, and computer vision, and so on [27].

The main objective of SVM is to find an optimal decision hyperplane in an N-dimensional space which maximizes the separation margin between a set of objects having different class memberships. SVM has some advantages which make it superior to artificial intelligence methods. For example, SVM can efficiently perform nonlinear classification or regression tasks using the kernel trick. Unlike the neural networks, the SVM ensures guaranteed optimality based on the convex optimization method. Thus, the solution is guaranteed to be a global minimum and not a local minimum. Furthermore, since the complexity of the training dataset in SVM is often characterized by the number of support vectors rather than the dimensionality, it can work effectively on small as well as high dimensional data spaces.

SVM algorithms have a set of mathematical functions that are called as kernels. The role of the kernel function is to take data as input and convert it into the required form. Different SVM algorithms use different kernel functions types. In this study, three common kernels including linear, polynomial, Gaussian radial basis function (RBF) are provided as follows:

Table 2. Descriptive statistics of BCC dataset

Group	*Variable*	*Mean*	*SE Mean*	*St. Dev*	*Variance*	*Range*	*Skewness*	*Kurtosis*
1	Age (Years)	58.08	2.63	18.96	359.41	65.00	-0.28	-1.26
2		56.67	1.69	13.49	182.07	52.00	0.53	-0.76
1	BMI (kg/m^2)	28.32	0.75	5.43	29.46	19.91	0.15	-1.19
2		26.99	0.58	4.62	21.35	18.74	0.06	-0.83
1	Glucose (mg/dl)	88.23	1.41	10.19	103.87	58.00	0.30	1.24
2		105.56	3.32	26.56	705.30	131.00	2.16	5.32
1	Insulin (μU/mL)	6.934	0.67	4.86	23.62	23.50	2.41	6.18
2		12.51	1.54	12.32	151.73	56.03	1.96	3.77
1	HOMA	1.55	0.17	1.218	1.48	6.645	2.69	8.73
2		3.62	0.57	4.589	21.06	24.54	2.91	9.67
1	Leptin (ng/mL)	26.64	2.68	19.33	373.83	79.17	1.15	0.79
2		26.60	2.40	19.21	369.12	83.95	1.47	2.17
1	Adiponectin (μg/mL)	10.33	1.06	7.63	58.24	35.85	2.09	4.62
2		10.06	0.77	6.19	38.31	32.10	1.37	2.33
1	Resistin (ng/mL)	11.61	1.59	11.45	131.04	78.81	4.80	28.69
2		17.25	1.58	12.64	159.69	52.01	1.53	2.15
1	MCP.1 (pg/dL)	499.70	40.50	292.20	85405.50	1210.20	0.74	0.09
2		563.00	48.00	384.00	147457.20	1608.40	1.57	2.63

Control group;1, Breast cancer patients; 2

- **Linear**

$$K(x_i, x_j) = \langle x_i . x_j \rangle \quad (2)$$

- **Polynomial:**

$$K(x_i, x_j) = (\alpha\langle x_i + x_j \rangle + c)^d, \gamma > 0 \quad (3)$$

- **Gaussian Radial Basis Function (RBF):**

$$K(x_i, x_j) = e^{\gamma \|x_i - x_j\|}, \gamma > 0 \quad (4)$$

where α is the slope, c is the constant term and d is the polynomial degree. $K(x_i, x_j)$ is the kernel function, x_i and x_j are the training and test patterns, respectively.

2.2.2. Naive Bayes

Naive Bayes (NB) is one of the efficient and effective inductive ML algorithms widely used in clustering and classification tasks. In ML, NB classifiers are a collection of "probabilistic classifiers" based on Bayes' theorem. NB classifiers find wide applications for many real-world problems such as email spam filtering, automatic medical diagnosis, text categorization, and document classification. It is easy to build an NB model because there is no complex iterative parameter estimation. Despite its simplicity and naive design, the NB classifier often outperforms more than sophisticated ML classifiers as it requires only a small amount of training data to estimate the relevant parameters required for classification. Moreover, the results can be interpreted easily compared to other AI methods. In the present study, the NB models are divided into two groups according to the probability distributions used. If the Gaussian distribution (or normal distribution) is used to model its continuous features, these types of NB models are called Gaussian NB. On the other hand, if kernel distribution is used instead of Gaussian distribution to model its continuous features, these types of NB models are called Kernel NB. Although Kernel NB models do not require a strong assumption like the normal distribution, they require more computation time and more memory than the normal distribution.

2.2.3. Decision Tree

Decision tree (DT) is one of the most widely used and practical ML methods due to its ability to create well-defined rules from the dataset. DT models have significant advantages because they are easy to understand and have low memory usage. DT algorithms are widely used in real-life applications such as in customer relationship management, healthcare management, fraud detection, and fault diagnosis, and so on.

The DT is a structured tree, consisting of a root node and several internal and leaf nodes. In a DT, each internal node represents a test on a feature (attribute), each leaf node represents a class (target) value and each branch represents the outcome of the test. In order to evaluate the results of a classification tree, the tree is followed from the first root node to a leaf node. At each node, which branch to follow using the rule associated with that node is decided. This process continues until it reaches a leaf node.

In this study, three different DT models according to their flexibility were used namely; Coarse Tree, Medium Tree, and Fine Tree. Model flexibility increases with the maximum number of splits set. For example, in the Coarse Tree model, the model flexibility is low as few leaves are used to make coarse distinctions between classes. In the medium tree model, the model flexibility is medium as a medium number of leaves are used for finer distinctions between classes. On the other hand, in the fine tree model, the model flexibility is high as many leaves are used to make many fine distinctions between classes.

2.2.4. The K-Nearest Neighbor

K-nearest neighbor (KNN) is a simple, easy to implement and effective supervised ML algorithm that can be used to solve both classification and regression problems. It is widely used in different applications such as text categorization, handwriting detection, speech and pattern recognition. Since the KNN is a lazy learner algorithm, it is much faster than other ML algorithms that require training.

In the KNN algorithm, the K value represents the number of nearest neighbors that should be determined appropriately because it affects algorithm performance. The smaller K value may lead to a larger variance, while using the larger K value may decrease the influence of the variance, however, it needs a more expensive calculation. KNN classification task is carried out by calculating the distances from the test instance to all training instances. Distance metrics are used to find the similarity and dissimilarity between data points. Distance metrics play a vital role in determining the final classification output. Various distance metrics such as Chebyshev, Cosine, Euclidean, Hamming, Jaccard, Mahalanobis are used to search the difference between training and testing samples. Although Euclidean distance is used as the most widely used distance metric in KNN classifications, the performance of other measurements should be evaluated.

2.2.5. Ensemble Classifiers

Ensemble classifiers (EC) are popular ML methods that combine a set of many weak individual classifiers into one high-quality predictive model to decrease the variance, bias or improve the predictive force. This approach provides better predictive performance compared to an individual classifier model. There are a variety of ensemble techniques in the literature, including bagging (bootstrapaggregation), boosting, and stacking (stackedgeneralization).

All boosting and bagging algorithms are based on decision tree learners. Generally, boosting algorithms are configured with weak learners and use very shallow trees. This structure uses relatively little time or memory. However, boosted trees may require more ensemble members than bagged trees for effective predictions. The bag algorithms usually construct deep trees. The bag algorithms can estimate the generalization error without additional cross-validation. This structure leads to relatively slow predictions as it is both time-consuming and memory-intensive. Thus, it is not always clear which algorithm class is superior.

2.3. Hyperparameter Optimization with Bayesian Optimization

Each model in ML techniques includes different hyperparameters that need to be set to obtain an excellent result. For example, parameters such as box constraint (C), kernel parameter (γ) and epsilon (ε) play an important role in the success of the SVM algorithm [28]. In another example, the proper selection of number of terminal nodes in the ensemble trees and the regularization parameter is important for stochastic gradient boosting algorithm performance. In another example, the number of terminal nodes and the regularization parameter in the ensemble trees directly

affects the performance of the stochastic gradient boosting algorithm. In ML studies, proper selection of the hyperparameter is a major challenge because most of the hyperparameters are continuous variables with only loose constraints on their numerical ranges. This task is usually performed using trial and error or by experts who use previous knowledge. However, such an approach may suffer from human bias. At the same time, trial and error requires a very long time and often does not lead to an optimized solution. Thus, robust optimization techniques are required.

In literature, three techniques are frequently used to optimizing ML hyperparameters; grid search, random search, and Bayesian optimization. The grid search requires a large number of parameter evaluations and long processing time when the dimensionality of the hyperparameter space is high [29]. The other parameter setting technique is random search. The random search hyperparameter searches the values randomly, then evaluates the model accuracy. In this technique, there is no guarantee that the next assessment will be better than the previous setup, as in the grid research. Due to the shortcomings mentioned above, a strong optimization technique has recently been needed as an alternative to grid or random search.

In this context, the Bayesian Optimization algorithm, which has a wide range of applications, has emerged as an effective tool for parameter search [29]. Compared to other methods, the BO algorithm requires less time with the smallest number of evaluations to find the best possible parameter values. This is a very useful strategy when the evaluation of the objective function is very expensive. Table 3 summarizes the hyperparameters and their search spaces of the proposed models.

2.4. Performance Evaluation Criteria (PEC)

In the literature, various metrics were used in order to evaluate the performance of classification models. Accuracy is widely applied as the main performance criteria to evaluate the model. However, solely accuracy measurement is not enough in order to evaluate a model with the imbalanced distribution of the class. Therefore, other measures such as precision, sensitivity (recall), specificity and F-score must be used as an alternative [30].

$$Accuracy = \frac{n_{tp} + n_{tn}}{n_{tp}+n_{tn} + n_{fp} + n_{fn}} \quad (5)$$

$$Precision = \frac{n_{tp}}{n_{tp}+n_{fp}} \quad (6)$$

$$Sensitivity = \frac{n_{tp}}{n_{tp}+n_{fn}} \quad (7)$$

$$Specificity = \frac{n_{tn}}{n_{tn} + n_{fp}} \quad (8)$$

$$Fscore = \frac{2\times Precision\times Recall}{Precision+Recall} \quad (9)$$

The symbols, n_{tp} (true positive) is the number of positive cases that are correctly identified as positive, n_{fp} (false positive) is the number of negative cases that are incorrectly identified as positive, n_{tn} (true negative) is the number of negative cases that are correctly identified as negative, n_{fn} (false negative) is the number of positive cases that are incorrectly identified as negative classified by each classifier.

3. Results and Discussion

3.1. Results of Improved ML Models

In this study, a total of 22 different classifier models such as (i) DT; fine tree, medium tree and coarse tree, (ii) NB; Gaussian NB, Kernel NB, (iii) SVM; linear SVM, quadratic SVM, cubic SVM, fine Gaussian SVM, coarse Gaussian SVM, (iv) EC; Boosted Trees, Bagged Trees, Subspace Discriminant, Subspace KNN, and RUSBoosted Trees, (v) KNN; fine KNN, medium KNN, coarse KNN, cosine KNN, cubic KNN, and weighted KNN were constructed for breast cancer detection under the two data division procedures. All classification models were carried out on the MATLAB 2018b platform. The symbolic notations used for 22 classifiers and 5 performance evaluation criteria are given in Table 4.

Table 5 summarizes the performance of all classifiers under 10-fold and hold-out data division procedure without using the NCA feature selection technique. All nine features in the BCC dataset were used as inputs of each classifier. The results depicted that the medium Gaussian SVM exhibited a relatively high level of performance than other classifiers under both division protocols. The highest classification accuracy of 85.294% was obtained under the hold-out data division procedure.

Table 3. Hyperparameters and their search space of the proposed models

Classification Algorithm	***Hyperparameters***	***Search Range***
DT	Maximum number of splits	[1-115]
	Split Criterion	Gini's diversity index, Maximum Deviance Reduction
NB	Distribution names	Gaussian, Kernel
	Kernel Type	Gaussian, Box, Epanechnikov, Triangle
SVM	Kernel Function	Gaussian, Linear, Quadratic, Cubic
	Kernel Scale	[0.001-1000]
	Box Constraint level	[0.001-1000]
	Standardize data	True/False
EC	Ensemble Method	Bag, GentleBoost, LogitBoost, Adaboost, RUSBoost
	Number of Learners	[10-500]
	Learning Rate	[0.001-1]
	Maximum number of splits	[1-115]
KNN	Number of Neighbors	[1-58]
	Distance Metric	City Block, Chebyshev, Correlation, Cosine, Euclidean, Hamming, Jaccard, Mahalanobis, Minkowski, Spearman
	Distance Weight	Equal, Inverse, Squared Inverse
	Standardize data	True/False

Table 4. Symbolic notations for classifiers and performance evaluation criteria

Classifiers	*Model*	*Notation*
DT	Fine Tree	CM1
	Medium Tree	CM2
	Coarse Tree	CM3
NB	Gaussian Naïve Bayes	CM4
	Kernel Naïve Bayes	CM5
SVM	Linear SVM	CM6
	Quadratic SVM	CM7
	Cubic SVM	CM8
	Fine Gaussian SVM	CM9
	Medium Gaussian SVM	CM10
	Coarse Gaussian SVM	CM11
EC	Boosted Trees	CM12
	Bagged Trees	CM13
	Subspace Discriminant	CM14
	Subspace KNN	CM15
	RUSBoosted Trees	CM16
KNN	Fine KNN	CM17
	Medium KNN	CM18
	Coarse KNN	CM19
	Cosine KNN	CM20
	Cubic KNN	CM21
	Weighted KNN	CM22
PEC	Accuracy (%)	CR1
	Precision (%)	CR2
	Sensitivity (%)	CR3
	Specificity (%)	CR4
	F-measure (%)	CR5

In contrast, coarse gaussian SVM, boosted tree, and coarse KNN models provided the lowest classification accuracy among all classifiers. Then, NCA, a supervised method, was applied to detect the relevant features by regularizing the feature weights. The most important property of the NCA is to produce non-negative weights for all features.

As you can see from Fig. 1, the weights of the irrelevant features such as insulin, HOMA, Leptin, and MCP.1 were determined to be very close to zero. Therefore, these features were removed from the dataset. On the other hand, glucose, age, resistin, adiponectin, and BMI were found to have a significant effect on breast cancer prediction with the weights 1.843, 1.806, 1.744, 0.935, and 0.551, respectively. BMI, glucose, age, and resistin have been identified as important biomarkers in previous studies, too. So, it can be said that the results of this study coincide with the results of previous studies [17,20].

Table 6 presents the accuracy of various classifiers applying the features selected by the NCA. It is seen that the accuracy values are in the range of 55.172%–82.759% under 10-fold cross-validation procedure between 55.882%–91.176% under hold-out cross-validation procedure. It is found that the medium KNN classifier achieved the highest classification accuracy of 82.759% and 91.176% respectively under 10-fold and hold-out cross-validation procedures, respectively. Moreover, the higher F-measure value indicates that the model performs better. Accordingly, the medium KNN model achieved the highest F-measure values of 80.392 % and 90.909 % under 10-fold and hold-out cross-validation procedures, respectively.

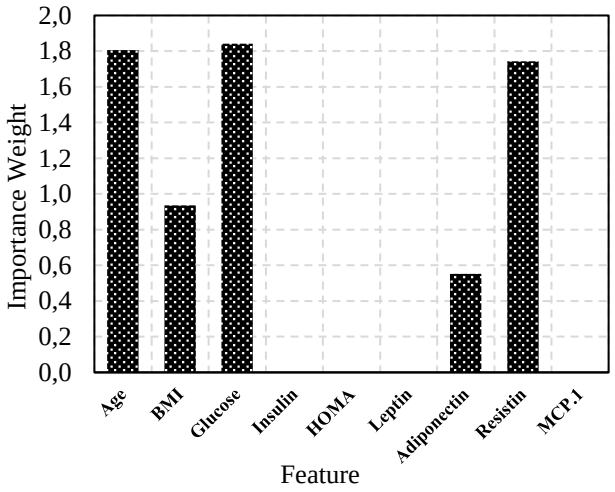


Fig. 1. The importance weight of features

Table 5. Performance of various classifiers without using feature selection under different data division protocols

Classifier	*Performance Measures (%)*									
	10-Fold					*Hold-out*				
	CR1	*CR2*	*CR3*	*CR4*	*CR5*	*CR1*	*CR2*	*CR3*	*CR4*	*CR5*
CM1	74.138	71.154	71.154	76.563	71.154	64.706	53.333	61.538	66.667	57.143
CM2	74.138	71.154	71.154	76.563	71.154	64.706	53.333	61.538	66.667	57.143
CM3	71.552	55.769	74.359	70.130	63.736	73.529	53.333	80.000	70.833	64.000
CM4	62.069	88.462	54.762	81.250	67.647	70.588	100.000	60.000	100.000	75.000
CM5	66.379	71.154	60.656	72.727	65.487	76.471	73.333	73.333	78.947	73.333
CM6	71.552	75.000	66.102	77.193	70.270	82.353	93.333	73.684	93.333	82.353
CM7	68.103	65.385	64.151	71.429	64.762	70.588	80.000	63.158	80.000	70.588
CM8	70.690	69.231	66.667	74.194	67.925	70.588	73.333	64.706	76.471	68.750
CM9	63.793	25.000	81.250	61.000	38.235	61.765	13.333	100.000	59.375	23.529
CM10	75.000	75.000	70.909	78.689	72.897	85.294	86.667	81.250	88.889	83.871
CM11	57.759	9.615	71.429	56.881	16.949	55.882	0.000	-	55.882	-
CM12	55.172	0.000	-	55.172	-	55.882	0.000	-	55.882	-
CM13	71.552	67.308	68.627	73.846	67.961	58.824	60.000	52.941	64.706	56.250
CM14	70.690	71.154	66.071	75.000	68.519	79.412	80.000	75.000	83.333	77.419
CM15	73.276	75.000	68.421	77.966	71.560	67.647	60.000	64.286	70.000	62.069
CM16	73.276	67.308	71.429	74.627	69.307	70.588	66.667	66.667	73.684	66.667
CM17	68.966	69.231	64.286	73.333	66.667	67.647	60.000	64.286	70.000	62.069
CM18	70.690	78.846	64.063	78.846	70.690	76.471	80.000	70.588	82.353	75.000
CM19	55.172	0.000	-	55.172	-	55.882	0.000	-	55.882	-
CM20	64.655	80.769	57.534	76.744	67.200	82.353	86.667	76.471	88.235	81.250
CM21	68.103	71.154	62.712	73.684	66.667	70.588	73.333	64.706	76.471	68.750
CM22	72.414	76.923	66.667	78.571	71.429	82.353	80.000	80.000	84.211	80.000

Table 6. Performance of classifiers using features selected by NCA under different data division protocols

Classifier	*Performance Measures (%)*									
	10-Fold					*Hold-out*				
	CR1	*CR2*	*CR3*	*CR4*	*CR5*	*CR1*	*CR2*	*CR3*	*CR4*	*CR5*
CM1	79.310	75.000	78.000	80.303	76.471	79.412	80.000	75.000	83.333	77.419
CM2	79.310	75.000	78.000	80.303	76.471	79.412	80.000	75.000	83.333	77.419
CM3	70.690	59.615	70.455	70.833	64.583	70.588	80.000	63.158	80.000	70.588
CM4	67.241	82.692	59.722	79.545	69.355	70.588	93.333	60.870	90.909	73.684
CM5	75.862	73.077	73.077	78.125	73.077	73.529	80.000	66.667	81.250	72.727
CM6	71.552	76.923	65.574	78.182	70.796	85.294	100.000	75.000	100.000	85.714
CM7	78.448	76.923	75.472	80.952	76.190	73.529	80.000	66.667	81.250	72.727
CM8	79.310	73.077	79.167	79.412	76.000	73.529	66.667	71.429	75.000	68.966
CM9	70.690	40.385	87.500	66.304	55.263	70.588	46.667	77.778	68.000	58.333
CM10	81.897	75.000	82.979	81.159	78.788	88.235	93.333	82.353	94.118	87.500
CM11	65.517	36.538	73.077	63.333	48.718	64.706	20.000	100.000	61.290	33.333
CM12	55.172	0.000	-	0.000	-	55.882	0.000	-	55.882	-
CM13	73.276	73.077	69.091	77.049	71.028	79.412	93.333	70.000	92.857	80.000
CM14	75.000	65.385	75.556	74.648	70.103	82.353	93.333	73.684	93.333	82.353
CM15	75.000	67.308	74.468	75.362	70.707	73.529	86.667	65.000	85.714	74.286
CM16	77.586	71.154	77.083	77.941	74.000	73.529	80.000	66.667	81.250	72.727
CM17	73.276	65.385	72.340	73.913	68.687	79.412	80.000	75.000	83.333	77.419
CM18	82.759	78.846	82.000	83.333	80.392	91.176	100.000	83.333	100.000	90.909
CM19	55.172	0.000	-	0.000	-	55.882	0.000	-	55.882	-
CM20	74.138	90.385	65.278	88.636	75.806	76.471	100.000	65.217	100.000	78.947
CM21	76.724	69.231	76.596	76.812	72.727	79.167	80.952	73.913	84.000	77.273
CM22	81.897	73.077	84.444	80.282	78.351	85.294	86.667	81.250	88.889	83.871

On the other hand, under all data division strategies, the lowest classification accuracy (55.172% and 55.882%) and the lowest F-measure values were obtained with the boosted tree and coarse KNN, respectively. The highest classification accuracy (91.176%) was achieved with medium KNN model using the NCA feature selection method. In the BO algorithm, the maximum number of objective function evaluations was set to "100" as the termination criteria. The optimization was run 5 times to find the best parameter values of each classifier with its specific search spaces. Table 7 summarizes the best configurations for all classifiers. As shown in Table 8, the integration of the BO algorithm achieved better classifier performance and higher accuracy results. For the KNN model, accuracy increased from 82.759% to 86.207% under 10-fold data division protocol and from 91.176% to 95.833% under hold-out data division protocol. As a result, the best classification performance with an accuracy of 95.833% is obtained by the BO-KNN model using the BO algorithm after the NCA feature selection. In Fig. 2, the performance of classifiers was examined more closely by plotting a receiver operating characteristics (ROC) curve. The ROC curve shows the true positive ratio (or sensitivity versus 1 specificity) to the false positive rate for different thresholds of the classifier output. Larger area under the curve (AUC) values indicate better classifier performance. The highest AUC value of 1.00 was obtained using glucose, age, resistin, BMI, and adiponectin features with BO- KNN classifier under hold-out data division protocol.

3.2. Summary of Results and Comparison with Literature

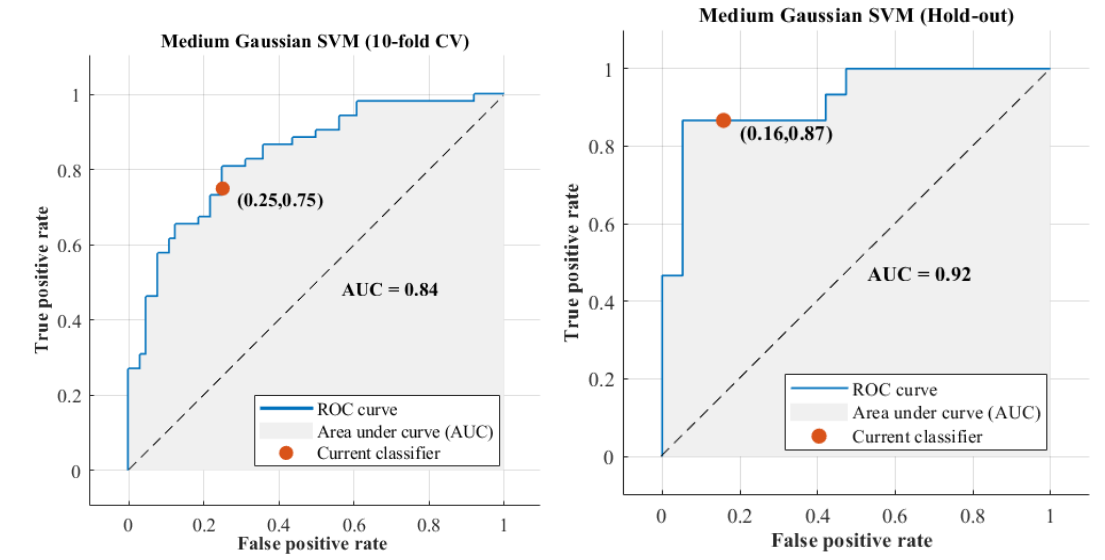
Early diagnosing of breast cancer is crucial for the patient's treatment option and planning, as well as the quality of life during treatment. In this context, in this study, different machine learning models based on the BO optimization method were constructed using glucose, age, resistin, BMI, and adiponectin features as predictors. It was seen that from Table 8, the highest accuracy value was achieved with the BO-KNN classifier after NCA under the hold-out data division. The accuracy, precision, sensitivity, specificity, and F-measure values were obtained as 95.833%, 100.000%, 91.304%, 100.000%, and 95.455%, respectively. In addition, as shown in Fig. 2, the highest AUC value was 1.00.

Table 7. The best-selected hyperparameter values of different classifiers

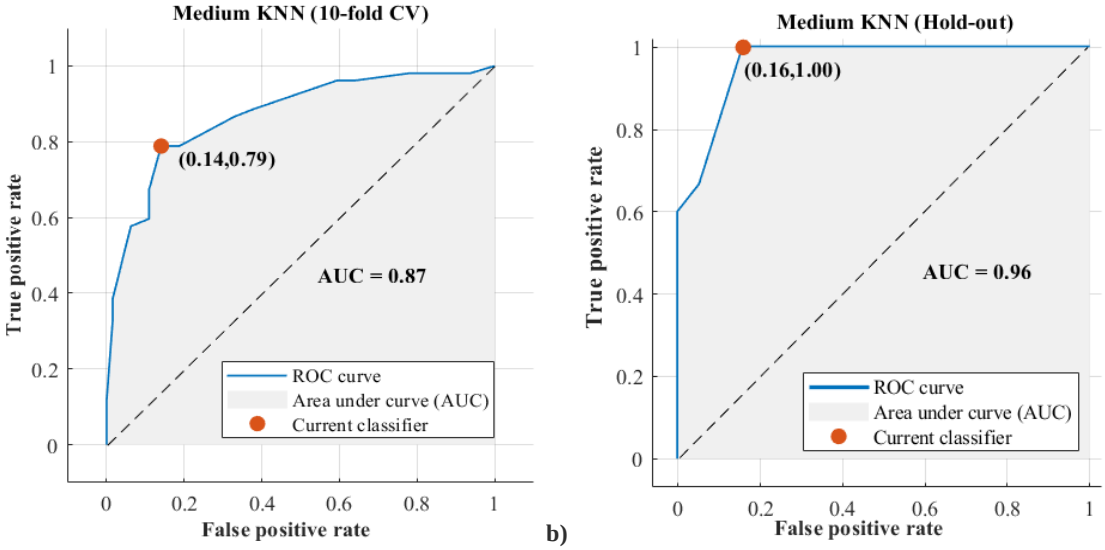
Classifier	*Optimized Hyperparameters*	*Sampling Strategy*	
		10-fold	*Hold-out*
BO_DT	Maximum number of splits	15	11
	Split criterion	Gini's index	Gini's index
BO_NB	Distribution type	Kernel	Kernel
	Kernel type	Triangle	Box
BO_SVM	Kernel function	Cubic	Gaussian
	Kernel scale	-	0.7116
	Box constraint level	0.0334	2.1987
	Standardize data	True	False
BO_EC	Ensemble method	Adaboost	GentleBoost
	Number of learners	10	10
	Learning rate	0.0351	0.2000
	Maximum number of splits	14	1
BO_KNN	Number of neighbors	5	10
	Distance metric	Minkowski(cubic)	Euclidean
	Distance weight	Equal	Squared Inverse
	Standardize data	True	True

Table 8. Performance of the improved ML algorithms using Bayesian Optimization

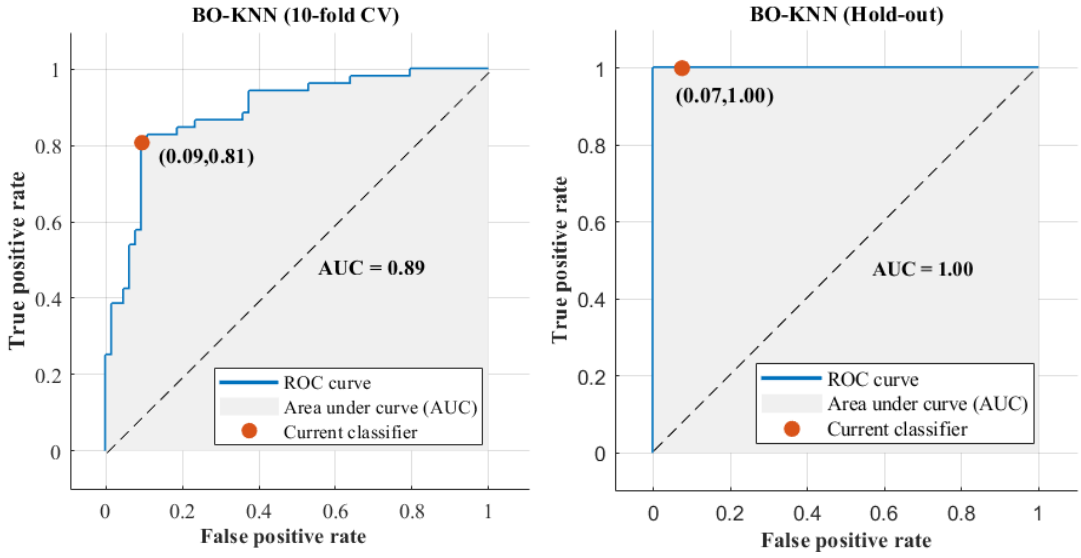
Classifier	*10-Fold*					*Hold-out*				
	CR1	*CR2*	*CR3*	*CR4*	*CR5*	*CR1*	*CR2*	*CR3*	*CR4*	*CR5*
BO_DT	79.310	75.000	78.000	80.303	76.471	79.412	80.000	75.000	83.333	77.419
BO_NB	77.586	75.000	75.000	79.688	75.000	76.471	80.000	70.588	82.353	75.000
BO_SVM	83.621	75.000	86.667	81.690	80.412	91.176	93.333	87.500	94.444	90.323
BO_EC	78.448	71.154	78.723	78.261	74.747	85.294	100.000	75.000	100.000	85.714
BO_KNN	86.207	80.769	87.500	85.294	84.000	95.833	100.000	91.304	100.000	95.455



a)



b)



c)

Fig. 2. ROC plots of the best classifiers a) without using feature selection, b) using NCA, c) using Bayesian Optimization

Table 9. Comparison of ML algorithms, achieved results and sampling strategies used in the studies

Reference	*Algorithms*	*Performance Measures*						*Sampling Strategy*
		CR1	*CR2*	*CR3*	*CR4*	*CR5*	*AUC*	
[2]	SVM, NB, DA, LR, **KNN**, and RF	92%		95%	88%		0.92	Hold-out (67%-33%)
[9]	LR, RF, and **SVM**	-	-	82-88%	84-90%	-	0.87- 0.91	Monte Carlo cross-validation
[10]	DT, **RF**, SVM, ANN, and LR	74%	-	-	-	78%	0.78	Random Subsampling (70%–30%)
[11]	ANN, SVM, KNN, and **ELM**	80%	-	-	-	-	-	Random Subsampling (80%–20%)
[12]	NB, SVM, KNN, **DT (Gini)**, DA, MLR, AdaBoost, and ANN	91%	-	92%	90.4%	-	-	Random Subsampling (90%–10%)
[13]	**FNN**, NB, MLP, RF, ZeroR, and DT	78-81%	-	78-82%	71-81%	-	0.76-0.81	Random Subsampling (70%–30%)
This Study	SVM, KNN, NB, DT, and EC							
	BO_SVM, **BO_KNN**, BO_NB, BO_DT, and BO_EC	96%	100%	91%	100%	95%	1.00	Hold-out (70%-30%)

ANN, Artificial Neural Network; LR, Logistic Regression; BT, Bayes Theorem; RF, Random Forest; SVM, Support Vector Regression; EC, Ensemble classifiers; DT, Decision Trees; AdaBoost, Adaptive Boosting; KNN, K-Nearest Neighbor; ELM, Extreme Learning Machine; NB, Naive Bayes; DA, Discriminant Analysis; MLR, Multiple Linear Regression; FNN, Fuzzy Neural Network; MLP, Multilayer Perceptron; BO, Bayesian Optimization.

Note: *Bold indicates the one that presented the best performance, All CR values are in (%)*

In recent years, many different approaches have been tried in the literature for the detection of breast cancer using the BCC dataset (Table 1). The results of the BO-KNN model were compared with the models developed by Patrício et al. [17], Li [18], Aslan et al. [19], Singh [4], Akben [20], and Silva Araújo et al. [21]. In all of these studies, the same number of subjects (116) were used with different input combinations and data division strategies. Table 9 shows a summary of the methods used in each research study and the performance results (highlighted in bold) of the best approach. The results were measured in terms of accuracy (CR1), precision (CR2), sensitivity (CR3), specificity (CR4), F-score (CR5), and AUC values. The data sampling strategies were also depicted. As can be seen in Table 9, the proposed BO-KNN model shows superior performance with 100% precision and 100% specificity among all models. Furthermore, it was observed that the accuracy and F-measure values obtained by classification methods in other studies were lower than our results. In particular, Singh [4] has achieved the closest performance values (accuracy≅ 92%, AUC≅0.92) under the hold-out data division protocol with medium KNN model. However, Li [18] has the lowest classification accuracy (≅74%) with the RF method. To summarize, the BO-KNN model developed in this study performed consistently and significantly better than the other models discussed in this study. From the results, we can conclude that the integration of the NCA and BO algorithm with ML models plays an important role in classification performance. In future studies, more classifiers with different feature selection strategies and optimization techniques can be studied to improve performance for diagnosing breast cancer.

4. Conclusion

Breast cancer is one of the most important health problems for women. The earlier breast cancer is detected, the better patient's chance of getting successful treatment. In this study, automatic breast cancer detection was performed using blood analysis and anthropometric data of 116 subjects. Different statistical techniques, such as cross-validation, feature selection, and performance evaluation, were applied. Twenty-two different ML models were constructed under two different data division procedures, namely, hold-out and 10-fold cross-validation. NCA based feature selection method was performed to select the most relevant biomarkers. Bayesian optimization method, which can efficiently optimize the hyperparameters of the ML algorithms, was applied to reduce the computation redundancy and enhance the performance of the models. Based on the accuracy and F-score values, the proposed BO-KNN classifier showed superior performance than other ML methods to detect breast cancer using the five powerful biomarkers, namely glucose, age, resistin, BMI, and adiponectin. The results of the study showed that the BO-KNN model has the potential to detect breast cancer early, and it can help researchers and doctors at the diagnosis stage.

References

[1] W. The International Agency for Research on Cancer (IARC) report, "Latest global cancer data: Cancer burden rises to 18.1 million new cases and 9.6 million cancer deaths in 2018," Int. Agency Res. Cancer, no. September, pp. 13–15, 2018.

[2] S. Sapate, S. Talbar, A. Mahajan, N. Sable, S. Desai, and M. Thakur, "Breast cancer diagnosis using abnormalities on ipsilateral views of digital mammograms," Biocybern. Biomed. Eng., pp. 1–16, 2019, https://doi: 10.1016/j.bbe.2019.04.008.

[3] Z. Ceylan and E. Pekel, "Comparison of Multi-Label Classification Methods for Prediagnosis of Cervical Cancer," Int. J. Intell. Syst. Appl. Eng., vol. 5, no. 4 SE-Research Article, Dec. 2017, doi: 10.18201/ijisae.2017533896.

[4] B. K. Singh, "Determining relevant biomarkers for prediction of breast cancer using anthropometric and clinical features: A comparative investigation in machine learning paradigm," Biocybern. Biomed. Eng., vol. 39, no. 2, pp. 393–409, 2019, doi: 10.1016/j.bbe.2019.03.001.

[5] J. Gong, X. Bai, D. -a. Li, J. Zhao, and X. Li, "Prognosis Analysis of Heart Failure Based on Recurrent Attention Model," IRBM, 2019, doi: https://doi.org/10.1016/j.irbm.2019.08.002.

[6] M. Toğaçar, B. Ergen, and Z. Cömert, "A Deep Feature Learning Model for Pneumonia Detection Applying a Combination of mRMR Feature Selection and Machine Learning Models," IRBM, 2019, doi: https://doi.org/10.1016/j.irbm.2019.10.006.

[7] R. D. Badgujar and P. J. Deore, "Hybrid Nature Inspired SMO-GBM Classifier for Exudate Classification on Fundus Retinal Images," IRBM, vol. 40, no. 2, pp. 69–77, 2019, doi: https://doi.org/10.1016/j.irbm.2019.02.003.

[8] A. Asuncion and D. Newman, “UCI machine learning repository.” 2007.

[9] S. Liu et al., “Quantitative analysis of breast cancer diagnosis using a probabilistic modelling approach,” Comput. Biol. Med., vol. 92, no. November 2017, pp. 168–175, 2018, doi: 10.1016/j.compbiomed.2017.11.014.

[10] R. N. Das, Y. Lee, S. Mukherjee, and S. Oh, “Relationship of body mass index with diabetes and breast cancer biomarkers,” vol. 9, pp. 1–6, 2019.

[11] J. H. Kang, B. Y. Yu, and D. S. Youn, “Relationship of serum adiponectin and resistin levels with breast cancer risk,” J. Korean Med. Sci., vol. 22, no. 1, pp. 117–121, 2007, doi: 10.3346/jkms.2007.22.1.117.

[12] H. L. Hwa et al., “Prediction of breast cancer and lymph node metastatic status with tumour markers using logistic regression models,” J. Eval. Clin. Pract., vol. 14, no. 2, pp. 275–280, 2008, doi: 10.1111/j.1365-2753.2007.00849.x.

[13] J. G. Santillán-Benítez et al., “The Tetrad BMI, Leptin, Leptin/Adiponectin (L/A) Ratio and CA 15-3 are Reliable Biomarkers of Breast Cancer,” J. Clin. Lab. Anal., vol. 27, no. 1, pp. 12–20, 2013, doi: 10.1002/jcla.21555.

[14] X. Provatopoulou et al., “Serum irisin levels are lower in patients with breast cancer: Association with disease diagnosis and tumor characteristics,” BMC Cancer, vol. 15, no. 1, pp. 1–9, 2015, doi: 10.1186/s12885-015-1898-1.

[15] A. M. A. Assiri, H. F. M. Kamel, and M. F. R. Hassanien, “Resistin, visfatin, adiponectin, and leptin: Risk of breast cancer in pre- and postmenopausal saudi females and their possible diagnostic and predictive implications as novel biomarkers,” Dis. Markers, vol. 2015, 2015, doi: 10.1155/2015/253519.

[16] A. M. A. Assiri and H. F. M. Kamel, “Evaluation of diagnostic and predictive value of serum adipokines: Leptin, resistin and visfatin in postmenopausal breast cancer,” Obes. Res. Clin. Pract., vol. 10, no. 4, pp. 442–453, 2016, doi: 10.1016/j.orcp.2015.08.017.

[17] M. Patrício et al., “Using Resistin, glucose, age and BMI to predict the presence of breast cancer,” BMC Cancer, vol. 18, no. 1, pp. 1–8, 2018, doi: 10.1186/s12885-017-3877-1.

[18] Y. Li, “Performance Evaluation of Machine Learning Methods for Breast Cancer Prediction,” Appl. Comput. Math., vol. 7, no. 4, p. 212, 2018, doi: 10.11648/j.acm.20180704.15.

[19] M. F. Aslan, Y. Celik, K. Sabanci, and A. Durdu, “Breast Cancer Diagnosis by Different Machine Learning Methods Using Blood Analysis Data,” Int. J. Intell. Syst. Appl. Eng., vol. 6, no. 4 SE-Research Article, Dec. 2018, doi: 10.18201/ijisae.2018648455.

[20] S. B. Akben, “Determination of the Blood, Hormone and Obesity Value Ranges that Indicate the Breast Cancer, Using Data Mining Based Expert System,” Irbm, vol. 40, no. 6, pp. 355–360, 2019, doi: 10.1016/j.irbm.2019.05.007.

[21] V. Silva Araújo, A. Guimarães, P. de Campos Souza, T. Silva Rezende, and V. Souza Araújo, “Using Resistin, Glucose, Age and BMI and Pruning Fuzzy Neural Network for the Construction of Expert Systems in the Prediction of Breast Cancer,” Mach. Learn. Knowl. Extr., vol. 1, no. 1, pp. 466–482, 2019, doi: 10.3390/make1010028.

[22] E. Brochu, V. M. Cora, and N. de Freitas, “A Tutorial on Bayesian Optimization of Expensive Cost Functions, with Application to Active User Modeling and Hierarchical Reinforcement Learning,” 2010.

[23] B. Shahriari, K. Swersky, Z. Wang, R. P. Adams, and N. de Freitas, “Taking the Human Out of the Loop: A Review of Bayesian Optimization,” Proc. IEEE, vol. 104, no. 1, pp. 148–175, 2016, doi: 10.1109/JPROC.2015.2494218.

[24] W. Yang, K. Wang, and W. Zuo, “Neighborhood component feature selection for high-dimensional data,” J. Comput., vol. 7, no. 1, pp. 162–168, 2012, doi: 10.4304/jcp.7.1.161-168.

[25] W. Bao, N. Lianju, and K. Yue, “Integration of unsupervised and supervised machine learning algorithms for credit risk assessment,” Expert Syst. Appl., vol. 128, pp. 301–315, 2019, doi: 10.1016/j.eswa.2019.02.033.

[26] M. R. Salmanpour et al., “Optimized machine learning methods for prediction of cognitive outcome in Parkinson’s disease,” Comput. Biol. Med., vol. 111, no. February, p. 103347, 2019, doi: 10.1016/j.compbiomed.2019.103347.

[27] P. H. Abreu, M. S. Santos, M. H. Abreu, B. Andrade, and D. C. Silva, “Predicting breast cancer recurrence using machine learning techniques: A systematic review,” ACM Comput. Surv., vol. 49, no. 3, 2016, doi: 10.1145/2988544.

[28] I. O. Alade, M. A. Abd Rahman, and T. A. Saleh, “Predicting the specific heat capacity of alumina/ethylene glycol nanofluids using support vector regression model optimized with Bayesian algorithm,” Sol. Energy, vol. 183, pp. 74–82, May 2019, doi: 10.1016/J.SOLENER.2019.02.060.

[29] L. Cornejo-Bueno, E. C. Garrido-Merchán, D. Hernández-Lobato, and S. Salcedo-Sanz, “Bayesian optimization of a hybrid system for robust ocean wave features prediction,” Neurocomputing, vol. 275, pp. 818–828, 2018, doi: 10.1016/j.neucom.2017.09.025.

[30] J. Han, J. Pei, and M. Kamber, Data mining: concepts and techniques. Elsevier, 2011.