International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11
Article Received: *10 September 2023* ***Revised:*** *20 October 2023* ***Accepted:*** *30 October 2023*

Hyperparameter-Optimized Machine Learning Techniques for Mammogram Classification

Sreekala K K[a], Jayakrushna Sahoo[b], Amarjit Roy[c]
[a] Department of Computer Science and Engineering, Indian Institute of Information Technology Kottayam, Kerala, India, sreekalaphd2019@iiitkottayam.ac.in
[b] Department of Computer Science and Engineering, Indian Institute of Information Technology Kottayam, Kerala, India, jsahoo@iiitkottayam.ac.in
[c] Department of Electrical Engineering (ECE Specialization), Ghani Khan Choudhury Institute of Engineering and Technology, West Bengal, India, amarjit@gkciet.ac.in

Abstract: Computer technology has employed Machine Learning models in a variety of applications to improve performance. The hyperparameter of a machine learning model must be adapted to overcome learning limitations and increase its performance. In this research, the hyperparameters of machine learning classifiers are tuned to identify cases of benign or malignant breast abnormalities. An experimental investigation was conducted using the Wisconsin Diagnosis Breast Cancer (WDBC) Dataset. A fusion model, Bayesian Optimization Hyper Band-Naïve Bayes (BOHB-NB) is employed, which is combined with conventional classification approaches like Logistic Regression (LR), Naive Bayes (NB), and Support Vector Machine (SVM). The proposed methods are compared to cutting-edge models like SVM, NB, LR, K-Nearest Neighbour (KNN), Random Forest, and Decision Tree using a wide range of parametric measures, such as Precision, Recall, Specificity, F-measure, Accuracy, True Positivity Rate (TPR), and False Positivity Rate (FPR). The results show that the proposed methods outperform the leading models.

Keywords: Machine Learning, Hyperparameter, supervised Learning, Classification, optimization, Breast Cancer.

1. Introduction

Machine Learning models are excellent choice for breast cancer detection due to its adaptability, repeatability, and ability to maintain original data accuracy [1]. In medical field, machine learning is vital for various purposes, including disease diagnosis, prevention, health check-ups, major disease screenings, health management, early diagnosis, disease severity evaluation, treatment method selection, treatment effect evaluation, and recovery[2-3].

Breast cancer is a major health concern for women in Indian metropolitan cities too, with high rates of morbidity and mortality[4]. India is estimated to carry one third of the global breast cancer burden. Between 2008 and 2018, the incidence and mortality rates of breast cancer in India increased by 11.5% and 13.82% respectively, which may be attributed to a lack of breast cancer screening programs, late diagnosis, and inadequate medical facilities. Breast cancer is a heterogeneous tumor with diverse clinical characteristics and response to treatment, and tumors are complex tissues made up of various cell types that interact with each other. The normal cells in the tumor-associated stroma actively participate in carcinogenesis, contributing to cancer hallmarks. The course of the disease is strongly influenced by local microenvironmental factors, as well as systemic factors such as age, body mass index, menopausal status, and overall immunity.

In the case of advanced classification methods such as Artificial Neural Network (ANN) and SVM, the classification's performance is affected largely by the dimension of feature vectors; also, the training time of the protocol is determined. A crucial task for feature selection is to extract and select required and useful features. After the selection of features, they are fed into a classifier, which categorizes the available lesions as benign or malignant. Punitha, Al-Turjman, and Stephan [5] proposed an automated breast cancer detection method using feature extraction and parameter optimization in ANNs. Some techniques are classifying lesions and non-lesions that is called binary lesion classification. Before classifying a lesion, an optimization technique is essential for the classifier for better performance The process of searching for a vector in a function that generates an optimal solution is called the process of optimization. The available solutions refer to all of the feasible values whereas the optimal solution refers to the extreme value. Generally, for solving these optimization problems, the optimization algorithms are applied. A simple manner of classification for such algorithms is taking into account the nature of the algorithms, whereby they can be classified as either stochastic or deterministic algorithms. The chief goal of machine learning is to develop an efficient procedure to automatically learn effective and accurate models from experience [6-7]. It provides methodology and procedure tools to aid in the resolution of prognosis and

diagnosis issues in many problems in medical fields. It is argued that the effective employment of learning tactics will aid in the introduction of computer-based systems in the healthcare setting, thus facilitating and improving the work of medical specialists and, eventually, improving the efficiency of the health care industry. To minimize diagnosis time and increase diagnosis accuracy, it has become critical to build dependable and efficient medical decision support systems to assist with the increasingly complex diagnosis decision process. [8-9]. Machine learning has emerged as the predominant technique for addressing data-related issues, and it is now commonly used in a variety of applications. To simulate the machine learning schemes to real-world problems, their Hyperparameters must be calibrated to match particular datasets. hyperparameter is very complex and costly; it has become critical to automatically optimize hyperparameter [10].

Previously, several machine learning methods were used to identify breast cancer, as Meriem. Amrane et al. [11] proposed two distinct classifiers: In order to find breast cancer, the NB classifier and the KNN classifier were utilized. To assess their precision, cross-validation was be used. The error rate (96.19 %). Abdel-Ilah L and Ahinbegovi H [12] used the WDBC, an open library dataset, in the majority of their techniques. The dataset had 699 samples that were split into two: 100 items for testing and 599 items for training. Breast fine-needle aspirates (FNAs), which each display nine features, were also employed as inputs for the network. Omar Ibrahim Obaid [13] has done another research in which the results of machine learning algorithms, including SVM, KNN, and Decision Tree. SVM is the most accurate and have the lowest false discovery rates, with an accuracy of 98.1%. A system for effectively differentiating between benign and malignant breast tumours has been developed by Habib Dhahri et al. [14] using a genetic algorithm that robotically finds the better model by linking features pre-processing methods and classifier algorithms.

Asri et al. [15] suggest that breast cancer can be detected using an interbred classifier that is based on a knowledge extraction approach. The main objective of this study was to analyze and compare various machine learning algorithms, including SVM, NB, K-NN, and Decision Tree. WDBC data set was employed to carry out the algorithms. Their primary goal was to assess the algorithms' performance across a wide range of parameters to develop a new fusion approach with optimal execution. Sadhukhan et al. [16] used a model created with machine learning and based on cell nucleus properties. Their examination revealed that their efforts were directed toward developing an algorithm capable of determining whether a tumour is malignant or benign. Benbrahim et al. [17] compared the effectiveness of 11 machine learning algorithms used for classification using the WDBC dataset. Using fine needle aspirate of a breast mass, the post-diagnosis photographs, the researchers proposed a technique for creating two classifiers that could identify between benign and malignant breast tumors. In order to determine which machine learning algorithm yields the best results, the researchers investigated and evaluated 11 different algorithms. The neural network had the highest accuracy index of the 11 with 96.49 percent, according to the research. Osmanovi et al. [18] used the University of California at Irvine (UCI) machine learning repository to extract data from 699 samples. An ANN was used (the lump nature) with nine neurons (number of features). The statistics showed that if the ANN was functionally implemented, there was a 99 % chance of proper diagnosis. They therefore had a 97.6% chance of receiving a bad reputation. In their [19] study, Negi et al developed four different machine learning based techniques for breast cancer diagnosis: Bayesian network, KNN, SVM, and Random Forest. To detect and classify breast cancer, the authors developed a comprehensive method. In terms of accuracy and distinctness, the SVM method surpassed the random forest methodology, with the former having the best chance of effectively recognising tumours. The WDBC dataset was used to assess the effectiveness of various breast cancer recognition algorithms, such as the Multilayer Perceptron, Kai-Neuron Networks, Classification and Regression Trees, and Natural Bayesian networks [20]. Research examined each algorithm individually in order to assess its reliability, accuracy, and precision as a data classification algorithm. The MLP outscored other algorithms with a score of 96.70 % on the training data. Asri et al. [21] compared and assessed four categorization algorithms: SVM, Naive Bayes, KNN, and Decision Tree. They attempted to assess the efficiency and usefulness of the algorithms by applying performance criteria such as recall, precision, specificity, and accuracy.

Yu et al. [22] introduced a light-weighted, fully convolutional network recognition system named DisepNet that recognizes breast abnormality. Hua Li et al. projected a DenseNet-II neural network model in their paper [23], an enhancement to the existing DenseNet model. Initially, pre-processing is done on mammography images. Second, the pre-processed mammography datasets are provided to the DenseNet-II neural network model. The model achieves an average accuracy of 94.55% when tested against the neural network models DenseNet, GoogLeNet, VGGNet, and AlexNet. Pratheep et al. [24] proposed a scheme for breast cancer classification by merging the merits of hyperparameter tuned random forest and feature weight. In KernelNeutrosophic c-Means classification, minor useful features are assigned smaller weights, while valuable features are assigned greater weights. This scheme uses a Bayesian

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11
Article Received: *10 September 2023* ***Revised:*** *20 October 2023* ***Accepted:*** *30 October 2023*

optimization algorithm to optimize the Random decision forest model. Figlu Mohantyel al. [25] proposed a work categorizing digital mammograms as malignant, benign and ordinary. The cross diagonal texture matrix (CDTM) approach is used to determine the mammography Region of interest (ROI), then Haralick's characteristics are extracted from each ROI . Then, the extracted feature vector's size is lessened using the kernel Principal Component Analysis (PCA) method. At last, the author introduced the kernel Extreme Learning Machine (ELM) method. In this method, the Kernel ELM parameters are optimized using the grasshopper optimization principle. For the DDSM and MIAS datasets, the model had corresponding accuracy levels of 92.61% and 97.49%.

Alshayeji M. H et al. [26] identified and predicted breast cancer without feature extraction or optimization techniques. Datasets from MIAS and DDSM were used for the creation of the method. This approach demonstrates an accuracy of 99.76% (normal vs. abnormal) from MIAS dataset and 98.80% for DDSM (benign vs. malignant) dataset. By using the Lizard optimization Algorithm, Subasreee et al. [27] suggested the Enhanced Recurrent Neural Network (RERNN) for breast cancer classification (LOA). Noises in the input images are removed by employing the pre-processing technique "Altered Phase Preserving Dynamic Range Compression (APPDRC)." An RERNN classifier is then used to extract radiomic characteristics, such as grayscale statistics, Haralick texturing, and morphological features (ELBP), based on the Entropy-Based Local Binary Pattern. Muduli D et al. [28] suggested a method for selecting characteristics from ROI mammography images using the lifting wavelet transform (LWT). The size of extracted features is decreased by combination LDA with PCA. Following that, breast cancers are classified using the moth flame optimization method (MFO-ELM). In this technique, MFO is utilized to optimize the characteristics of the ELM hidden nodes. According to this model, 99.76% of normal and abnormal classes are detected by the MIAS dataset, whereas 98.80% of benign and malignant classes are detected by the DDSM dataset.

Despite the use of advanced Artificial neural network techniques for early detection of breast cancer, researchers still face difficulties in accurately categorizing cancers as malignant or benign. As a result, they invest significant effort into finding a combination of techniques that can recover their results. The goal of this work is to develop a hyper-parameter optimization model for machine learning classifiers with a high degree of accuracy for early breast cancer detection, resulting in earlier diagnosis. A series of experiments were carried out using various hyperparameter optimization techniques on the selected classifiers. A technique combining BOHB and NB algorithms is designed and found as a high performing model.

As we have already came across the generic overview of Breast cancer and some of the recent works over those in Section 1, the rest of the paper is as shadows: Section 2 depict the dataset preparation, Section 3depict the proposed methodology, Section 4depict the results and discussion, and finally concluded with conclusion of this work in Section 5.

2. Dataset

The dataset, WDBC is used in this study for breast cancer identification process. WDBC dataset is a machine learning repository. This dataset contains 569 samples. Malignant and benign are the two classes of this dataset [26]. There are 357 benign samples and 212 malignant samples. The points in the benign class are taken as inliers, while the malignant type of this dataset is down-sampled to 21 points, with observed outliers. For the experimental purpose these 699 samples are split into two, a training set and a test set.

2.1 Dataset Preprocessing

An important step in data mining is data pre-processing, which transforms raw data into clean, precise, and understandable forms. In pre-processing activities, data is cleaned, transformed, mapped, reduced, organized, and selected. Using the Wiener filter, an ideal guess of the actual picture is made by limiting the mean square error between the estimated and unique images. The ideal filter is the wiener filter. The mean square error is decreased with a wiener filter. In addition to handling degradation functions, wiener filters also handle noise.

The degradation model estimates the error between F(M, N) and F(m, n) as:

$$E(M, N) = F(M, N) - F(m, n)$$

The square of the error is provided by

$$[F(M, N) - F(m, n)]^2$$

According to, the mean square error is

$$E\{F[(M, N)-F(m, n)]^2\}$$

Among every challenge, feature selection is the most crucial. The dataset is standardized by using equation 1 to avoid incorrect tasks and imbalanced data.

$$Z = X - \mu\ \sigma \qquad (1)$$

Where X is designated as a characteristic that has to be standardized, its mean value is given as μ and its SD is denoted as σ.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11
Article Received: *10 September 2023* ***Revised:*** *20 October 2023* ***Accepted:*** *30 October 2023*

3. Proposed methodology

Figure 1 illustrates the proposed approach for the classification of breast cancer. In this methodology, the WDBC dataset is used for experimentation. Initially, the data are pre-processed using the pre-processing scheme. The pre-processed data is given to the PCA. This PCA approach selects the best features only, which is then applied to train a machine learning classifier model, starting from the most prominent components of the original dataset. Before that, the machine learning model's parameters are optimized using the Bayesian Optimization- Hyperband (BO-HB) technique. The machine learning classifiers utilized are LR, SVM, and Naive Bayes. In the classification process, data samples are classified as benign or malignant. Finally, compare the performance metrics with the existing model, from the comparison our proposed works achieve better performance scores.

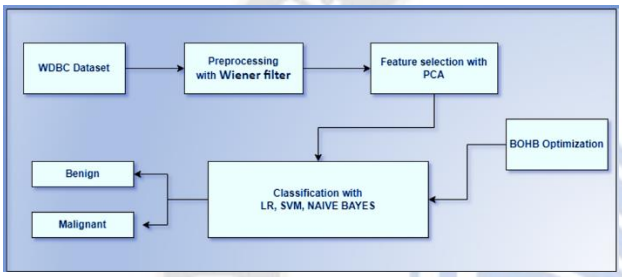

Figure 1: Block Diagram of the Projected Procedure

3.1 Feature Selection

Feature selection process involves extracting only relevant features while discarding unnecessary data, thus reducing the amount of data to be processed. Picking relevant characteristics is the first step in developing a classification model. The total number of input parameters should be kept to a minimum in a classifier to maximize the model's classification performance. The classification process is carried out with the same or greater precision by employing a small number of dominating characteristics rather than the entire attribute collection. By selecting a subset of attributes, one can reduce computation time and model complexity in addition to improving precision. The primary challenge of machine learning is dimensionality reduction. As a bonus, it may also suggest inherent traits and provide assistance in elucidating the biological manifestation of the machine learning "black box" [30].

3.1.1 Principal Component Analysis

By modifying data using feature points, PCA analyzes feature points. It just uses the core fundamental elements and ignores the interval. By projecting data onto a low-dimensional subspace, data from a space can be accessed. With more diverse data, it aims to keep the essential components while reducing superfluous components with less diversity [31].

This article suggests methods for selecting features based on PCA. A theoretical frequency distribution correlates to the appropriate features utilizing PCA as an empirical frequency distribution for this selection. A set of measurements that could be clustered can be transformed using an orthogonal transformation as part of the scientific process known as PCA to create a set of linearly uncorrelated values. The amount of PCA is equal to or less than the sum of primary values.

The mean values of Cell concavity, texture, symmetry, fractal dimension, area, smoothness, radius, and perimeter are defined as features of cancer, and their importance in percentage (%) value is given in Table 1. These variables' maximum value is linked to cancerous tumors. ii) The average values for symmetric, smoothing, texturing, or fractal dimensions do not reveal a preference for one diagnosis over another.

By discovering a new, smaller set of m variables, and keeping the majority of the data information, or the variance in the data, PCA is being applied to the data set with the ultimate goal of reducing its dimensionality in this case. Keeping the first m principal components should preserve the majority of the data info while decreasing the dimensionality of the data set since the principal components obtained from PCA are sorted in terms of variance. Considering that the eigenvectors of the symmetric covariance matrix make up the primary components, they are orthogonal. By reducing the number of eigenvectors, the dimensions of the new set of data are reduced.

Table 1: Feature Importance [31]

Features	**Importance (%)**
Radius	11.5543
Perimeter	20.45614
Area	17.3446
fractal dimension	0.9043
Concavity	18.2823
Texture	3.8492
Symmetry	1.6371
Smoothness	1.9023
Compactness	4.3632
Concave points	19.7067

3.2 Machine Learning Classifiers

The information is automatically categorized or organized into one or more of a set of "classes" using a classifier algorithm in machine learning. Automating

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11
Article Received: *10 September 2023* ***Revised:*** *20 October 2023* ***Accepted:*** *30 October 2023*

procedures that typically require manual work is made possible by machine learning algorithms. In addition to saving a large sum of money and time, they can improve the effectiveness of problem-solving. The following subsections provide the key model for categorizing breast cancer photos.

3.2.1 Linear Regression Model

Depending on whether they are used to predict target variables, supervised learning prototypes are categorized as regression or classification techniques.

Regression is a modelling technique that examines the interaction between dependent and independent factors. With linear regression, the goal is to forecast the amount of the dependent variable from a collection of independent variables or features rather than classifying it as LR does. This is accomplished by employing a sigmoid function to calculate the likelihood that a given instance belongs to a specific class.

A popular regression model that forecasts a target y is called linear regression. It uses the formula used in equation 2:

$$\hat{y}(W,X) = w_0 + w_1x_1 + \cdots + w_px_p \qquad (2)$$

Assume the target variable as y which is a linear arrangement of p whose input characteristics as x = (x_1…. x_p). The weight vector w = (w_1....w_p) is identified as an attribute 'coef' in the sklearn linear model, while w_0 is labelled as an additional variable 'intercept '. In most cases, no hyperparameters must be calibrated in Logistic Regression.

The initial LR models were created using Ridges regression, which penalizes variables and minimizes objective functions as shown in equation 3.

$$\alpha\|\omega\|_2^2 + \sum_{i=1}^{p}(y_i - w_i.x_i)^2 \qquad (3)$$

Where, $\|\omega\|_2$ is the L2-norm of vector, and α is identified as the regularization strength. The factors are more resistant to collinearity when the value is larger since a higher value indicates a greater degree of shrinkage as in Equation 4.

$$\alpha\|\omega\|_1 + \sum_{i=1}^{p}(y_i - w_i.x_i)^2 \qquad (4)$$

As a result, in both the ridge and lasso regression models, regularisation intensity is a major hyperparameter.

A linear model is the classification model used to classify data, which is based on the regularization technique used for the penalization, the cost function in LR can differ. Regularization methods in LR are classified into three types: L1 and L2-norm and elastic-net regularization. To control the overfitting in the LR model, the regularization penalties scheme is used as model complexity.

3.2.2 Support Vector Machine

Regarding classification and regression issues, the SVM classifier is a type of supervised machine learning. A hyperplane is then built as a classification limit for the partition information after SVM algorithms linearly different data points by shifting them from small to big areas. The objective function of an SVM, as shown in equation 5, presupposes that there are n data points.

$$\text{argmin} = \left\{\frac{1}{n}\sum_{i=1}^{n} max\{0, 1 - y_i f(x_i)\} + Cw^T w\right\} \qquad (5)$$

Where C denotes the penalty parameter of the error term—a crucial hyperparameter in all SVM models—and w denotes a normalization vector. As a result, the kernel form will be a critical hyperparameter to tune. The examples of common kernel forms in SVM are as follows:

The many kernel functions are categorized on the basis.

1. Linear kernel: $f(x) = x_i^T x_j;$

2. Polynomial kernel: $f(x) = \left(\gamma x_i^T x_j + r\right)^d$

3 RBF kernel: $f(x) = exp(-\gamma\|x - x^{!}\|^2)$

4. Sigmoid kernel: $f(x) = \left(tanh\left(\gamma x_i^T x_j + r\right)\right)$

Following the selection of a kernel kind, a few other diverse hyperparameters must be tuned, as seen in the kernel function equations.

3.2.3 Naive Bayes

The field of medical data mining has benefited greatly from the application of Naive Bayes classification. It has a high level of accuracy while maintaining the characteristics' autonomy. Values are constantly being lost in clinical data.

Misplaced values are treated in this work as plainly accidental. This method employs the use of zeros to represent sparse numerical data, while vector is retained. It is assumed that nested columns with missing data are sparse. When dealing with columns of a simple data type, missing values are assumed to be absent at random. To lower the cardinality, columns can be deleted as needed.

NB algorithm depends on the Bayesian methods, which are used in the classification purpose of this study. NB algorithms that determine explicit hypothesis probabilities, like the classification Naive Bayes, are one of the best practices for certain types of problems $x_{1,} \cdots x_n$ and a target y variable will denote the impartial function of nave Bayes provided the n function as shown in equation 6, is dependent on x_1:

$$\hat{y} = argmax = P(y)\prod_{i=1}^{n} P(x_i/y) \qquad (6)$$

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11
Article Received: *10 September 2023* ***Revised:*** *20 October 2023* ***Accepted:*** *30 October 2023*

If $P(y)$ is the y value probabilities, and $P(x_i/y)$, the y values are given to the posterior x_i probabilities. Considering the various assumptions of $P(x_i/y)$ distribution, which complement the four major forms of the NB model.

For Gaussian NB, as in equation 7, the Gaussian distribution is followed as:

$$P\left(\frac{x_i}{y}\right) = \frac{1}{\sqrt{2\pi\sigma_y^2}} exp\left(-\frac{(x_i-\mu_y)^2}{2\sigma_y^2}\right) \quad (7)$$

By employing the maximum probability form, it is not necessary to tune a hyperparameter to evaluate the mean value μ and the variance y for Gaussian NB. The Gaussian distribution governs the output of the Gaussian NB model.

Multinomial NB is planned for the distribution of multinomial data based on the algorithm of NB. If there are n characteristics and a function value i is invoked at the data point belonging to class 'y' which is the distribution of any value from the target mutable $P(x_i/y)$. It is equal to the conditional probability$P(x_i/y)$, as given in equation 8, which can be calculated by an updated version of θ_{yi}, based on the idea of relative frequency counts.

$$\hat{\theta}_{yi} = \frac{N_{yi}+\alpha}{N_y+\alpha n} \quad (8)$$

If N_{yi} is signified as the sum of times in which function i is in the y-type data point, $N_y$is the total of N_{yi} $(i = 0; 1; 2; \cdots; n)$. For functionalities not included in the learning examples, the smoothing priors $\alpha = 0$. When α = 1, Laplace smoothing is referred to; when α < 1 Lidstone smoothing.

NB is developed for the classification purpose for the binary class or multi-class classification, which can handle imbalanced data. NB enables the use of binary-assessed functional vectors for the data to comply with multivariate Bernoulli distributions. The smoothing parameter for the additive (Laplace/Lid stone) is the central hyperparameter to be tuned.

3.3 Hyperparameter Optimization Technique

A technique called hyperparameter optimization helps with the challenge of optimizing the hyperparameters of algorithms. Outstanding machine learning models include a wide range of different and advanced hyperparameters that create a vast search space. Many start-ups employ deep learning as the foundation of their pipelines and the search space for deep learning methods is considerably greater than that of typical machine learning algorithms.

Parameter tuning on a large search space is a challenging task. Then the data-driven strategies must be employed to overcome hyperparameter optimization issues. Manual approaches do not work. Figure 2 depicts the overall flowchart of the optimization methods in which the Bayesian Optimization and Hyperband techniques are discussed and thereby bring an integrated version of both i.e., BOHB to the machine learning algorithms such as LR, SVM, and NB for effective performance.

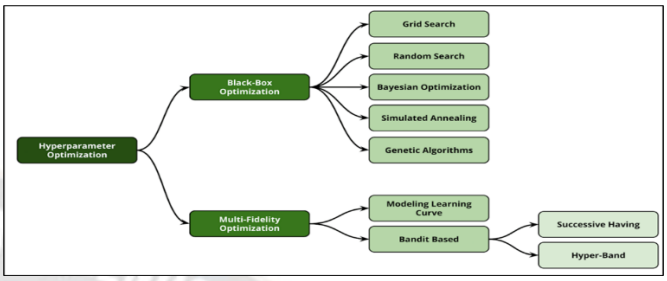

Figure 2: Block Diagram of Hyperparameter Optimization Procedures

3.3.1 Bayesian Optimization

Bayesian Optimization (BO) is a common iterative procedure for hyperparameter optimization difficulties. BO bases potential assessment points on previously obtained data. The substitution model attempts to match the objective function to all currently observed locations. BO models are tailored to exploration and exploitation operations to identify likely appropriate regions while avoiding overfitting to better configurations in unknown areas.

A closed-form expression-less black-box function is denoted by the letter f as in equation 9. In addition, evaluating this black-box function is costly. Let f: XR denote a well-behaved function on the X Rd subset. This project's purpose is to tackle the global optimization challenge outlined below.

$$x^* = arg\ \max(f(x)) \quad (9)$$

Bayesian optimization finds the optimal black-box function f(xglobal) by building a probabilistic model for the function and utilizing it to determine where in X to analyze the function next while taking uncertainty into account (x). This results in a method for finding the fewest number of complex non-convex functions with the fewest number of evaluations, but at the cost of doing further math to determine the next place to test [35–37]. Algorithm 1 and Figure 3 provide a summary of the BO representation and process.

Algorithm 1: Bayesian Optimization [33]

Input: initial data D_0, #iteration T
Output: x_{max}, y_{max}
Begin
Step 1: Find a black-box function without closed-form expressions.
Step 2: Find a well-behaved function on the X Rd subset

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11
Article Received: *10 September 2023* ***Revised:*** *20 October 2023* ***Accepted:*** *30 October 2023*

Step 3: Identify the global optimization challenge
For t= 1 to T do,
Fit a GP from D_t and get (9)
End for
Step 4: obtain the predictive delivery for a new remark x
Step 5: calculate mean and variance
Step 6: Evaluate yt = f(xt) and $D_t = D_{t-1} \cup (xt, yt)$
End.

By inferring f through evaluations, Bayesian optimization generates a Gaussian process [38]. With the help of this adaptable distribution, it may associate a randomly distributed parameter with any location in the continuous input space. Given the expected distribution, which is also a Gaussian distribution, a new observation x is calculated with the following mean and variance:

$$\mu(x') = k(x', X)K(X, X)^{-1}y$$

$$\sigma^2(x') = k(x', x') - k(x', X)K(X, X)^{-1}k(x', X)^T$$

The element (i, j) of the covariance matrix K(U, V) is computed as ki,j = k(xi, xj), where xi is U and xj is V.

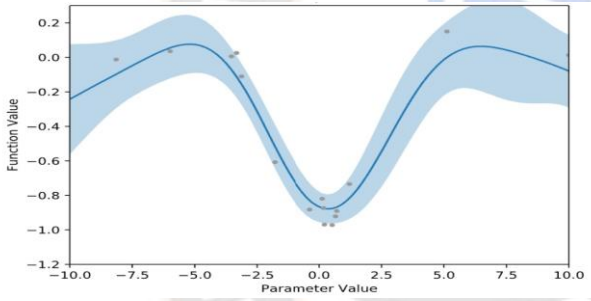

Figure 3: Bayesian Hyperparameter Optimization

Because it is costly to compute the original function f(x), a less costly function called the gaining function (x) will be utilized to select the next point to be assessed. This function may be created using the surrogate model. To find the next point, it function xt+1 = arg max xXt rather than the original function (x). This auxiliary maximisation is well-known, and conventional numerical techniques can be used to optimise it. The acquisition capabilities were purposefully created to mix the usage of existing attractive sites with the search space research. Despite the availability of a variety of acquisition functions [39–41], no particular acquisition strategy provides the optimum overall effectiveness.

As a result, following each objective function assessment, BO updates the substitution model. BO is more efficient than Grid Search and Randomized Search because it can identify the best hyperparameter combinations using a substitute model and is often less expensive than completing the entire target function. The model can make irregular and parallel sequence methods in certain iterations. It is usually possible to detect nearly optimal hyperparameter combinations in certain iterations only.

Here in this model, BO is conducted because it converges faster and less expensive than the Grid Search and Randomized Search optimization techniques. It gives nearly optimal hyperparameter combinations within certain iterations itself, which makes the models more efficient.

3.3.2 Hyperband

The solution proposed by Hyperband involves constantly selecting the appropriate factors for the classification algorithms. This attempts to find a balance between the sum of configurations of hyperparameters (n) and their budget allocations (b = B = n) and to assign the total budget (B) to n parts for each configuration ($b = B = n$). Successive halving is used as a subroutine on random configurations to remove underperforming hyperparameter configurations and increase performance. Algorithm 2 depicts the key stages of Hyperband algorithms.

Algorithm 2: Hyperband[29]

Input: b_{max}, b_{min}
Begin
step 1:$s_{max} = log\left(\frac{b_{max}}{b_{min}}\right)$
step 2: for $s \in \{b_{max}, b_{min} - 1, \ldots\ldots, 0\}$ do
step 3: $n = DetermineBudget(s)$
step 4: $\gamma = SampleConfigurations(n)$
step 5: $SuccessiveHalving(\gamma)$
step 6: end for
step 7: return: The finest configuration so far.
End

The samples are reliant on n and b in phases 4 and 5, and the results are then communicated to the subsequent half-model. This discards the 30 malfunctioning configurations found and goes ahead with the next iteration with the good configurations. It is a frequent process until the optimal configuration of the hyperparameter is established. The complexity of Hyperband, represented as o is caused by the successive half search method ($nlogn$).

3.3.3 BOHB

BO is a way to optimize a CNN, which is a recent hyperparameter optimization technology that integrates BO and Hyperband, to benefit from both while avoiding their disadvantages. A low-efficiency random search is carried out in the original Hyperband for the hyperparameter configuration space. To attain high performance and low running time by effectively making use of parallel resources for optimizing all types of hyperparameters, BOHB replaces the random search approach with BO. Tree Parzen Estimator (TPE) is the conventional substitution model in BOHB,

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11
Article Received: *10 September 2023* ***Revised:*** *20 October 2023* ***Accepted:*** *30 October 2023*

although it employs multidimensional kernel density estimators.

Model-based searchers in BOHB are used instead of the configurations that are randomly chosen at the beginning of each HB loop to determine how many configurations to analyze with which budgets. Following the achievement of the necessary iteration, these configurations are decreased using the conventional sequential halving procedure. It keeps track of the efficiency of all objective functions g (x, b) + $ of configurations x on all expenditures b and utilizese this information as the basis for our models in subsequent rounds. The budget selection method suggested by HB is still used, but it substitutes a BO component for the representative selection to direct the results. In an effort options that have been previously assessed, build a model and use BO.

This technique, as summarised by the pseudocode in Algorithm 3, will be explained in the remainder of this section. The main difference between the BO element of BOHB and TPE is that, in contrast to the TPE hierarchy of one-dimensional KDEs, a single multidimensional KDE is chosen to more effectively handle impacts in the input space. A minimal amount of points, Nmin, is needed to fit acceptable KDEs; Nmin is equal to d + 1, where d is the number of hyperparameters (in line 4 of Algorithm 3). It immediately starts modeling before the sum of measurements for budget b, Nb = |Db|, is enough to content q • NbNmin. Instead (line 3), it selects the Nmin + 2 random configurations that have the lowest

$$N_{b,l} = max(N_{min}, q \cdot N_b)$$

$$N_{b,g} = max\left(N_{min}, N_b - N_{b,l}\right) \qquad (10)$$

optimal and least optimal arrangements for modeling the two densities. As a result, even with few measurements provided, both models will have enough data points and the least amount of overlap. It collects a sample of Ns points from l0(x), which has the similar KDE as $l(x)$, but with altogether bandwidths increased by a proportion of bandwidth, to encourage further research around intriguing setups and to boost Expected Improvement, EI (lines 5-6). In the last phases of optimizing, it is discovered that this method enhances convergence by frequently querying but infrequently updating the model with the largest budget. It also only picked out a fixed percentage of the combinations to maintain the theoretical guarantees of HB (line 1). This ensures that in addition to global exploring, our approach has also assessed (on average) m • (smax + 1) random configurations on bmax after m • (smax + 1) SH runs. RS analyzes (smax + 1)-times as many configurations on the maximum budget because each SH run, only consumes up to (smax + 1) • bmax of the budget as in equation 10. This demonstrates that BOHB will converge eventually even in the worst scenario (where the lesser fidelities are deceiving), even if it is this amount shorter than RS. Similar reasoning holds for HB, however, in our tests, BOHB and HB both performed noticeably better than RS [43, 44].

Here in this work, the model is trained with BOHB to attain high performance and low running time by effectively making use of parallel resources for optimizing all types of hyperparameters. It is discovered that this strategy enhances convergence, particularly in the last stages of optimization when methods with the greatest budgets—like SVM, NB, and linear regression—are frequently requested but infrequently changed.

Algorithm 3: Pseudocode for BOHB[35]

Input: *D observations, N number of samples, percentile q, ρ is the percentage of random runs, the minimum number of points Nmin needed to form a model and bandwidth factor bw*
Output: *next configurations to evaluate*
Step 1: *If rand() < ρ then*
Step 2: *Return random configurations*
Step 3: $B = arg\ max\ \{Db : |Db| > \_ - Nmin + 2\}$
Step 4: *If* $b = \theta$ *then*
Step 5: *Return random configuration*
Step 6: *Fit KDEs according to equation* (10)
Step 7: *Draw Ns samples according to* $l'(x)$
Step 8: *Return sample with highest ratio* $l(x)/g(x)$
Step 9: *end*

4. Result and Discussion

The experiment for identifying and categorizing breast cancer is done in this simulation procedure using the WDBC dataset. The experiment runs on a machine with 4GB of RAM running Windows 10 and uses Python as the platform. The next part provides mathematical expressions for each metric used in the machine learning classifier's performance measurement, including accuracy, precision,

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11
Article Received: *10 September 2023* ***Revised:*** *20 October 2023* ***Accepted:*** *30 October 2023*

specificity, recall, f-measure, tpr, fpr, and calculation time. The percentage of accurate predictions is used to determine which classifier to use. Classifier accuracy is determined using a variety of machine learning classifier approaches. In this work, methods are compared —SVM, NB, and LR—to other models already in use. There are numerous ways to alter the split between the training and test data sets. Here, the ratio between the training and testing datasets was 70:30.

4.1 Performance Analysis

True represents the actual instances in the data sets, and predicted means how accurately the classifier would be able to classify B (benign) or M (malignant). Table 2 shows the performance measure used for evaluation. Table 3 depicts the comparison analysis of our 3 models (BOHB-SVM, BOHB-NB, BOHB-LR) with other existing models such as KNN, Regular SVM, Regular NB, Regular LR, Random Forest, Decision tree over various performance measures.

Table 2: Overall Performance Measures

Performance measures	Description
Specificity	$Specificity = TN / (TN + FP)$
Recall	$Recall = TP/(TP + FN)$
Precision	$Precision = TP / (TP + FP)$
Accuracy	$Accuracy = (TP + TN)/(TP + FP + TN + FN)$
F1-score	$F1 - score = 2 * (Recall * Precision) / (Recall + Precision)$
FPR	$FPR = FP / (FP + TN)$
TPR	*TPR = TP/(TP+FN)*

Table 3: Presentation Analysis of Different Machine Learning Classifiers

Models	Accuracy	Specificity	Precision	Recall	F1-score	TPR	FPR
Regular SVM	0.85±0.03	0.93±0.009	0.84±0.02	0.9±0.011	0.868	0.81	0.19
Regular NB	0.83±0.02	0.95±0.03	0.85±0.01	0.93±0.05	0.888	0.76	0.24
Regular LR	0.89±0.0	0.95±0.05	0.88±0.01	0.94±0.06	0.909	0.86	0.14
KNN	0.81±0.04	0.94±0.04	0.8±0.027	0.93±0.01	0.860	0.81	0.19
Decision Tree	0.87±0.02	0.97±0.014	0.88±0.01	0.94±0.02	0.909	0.88	0.12
Random forest	0.84±0.01	0.94±0.02	0.9±0.016	0.96±0.05	0.929	0.9	0.1
BOHB-SVM	0.95±0.06	0.98±0.012	0.95±0.01	0.97±0.03	0.959	0.92	0.08
BOHB-NB	0.98±0.05	0.99±0.006	0.97±0.01	0.98±0.06	0.974	0.95	0.05
BOHB-LR	0.97±0.04	0.99±0.003	0.96±0.01	0.98±0.05	0.969	0.96	0.04

Figure 4 depicts the graphical representation of various state of art models concerning our 3 models over measure Accuracy and specificity in which due to optimization to the 3 MLA, it outperforms with 0.95, 0.97, 0.98 than regular models like SVM, NB, and LR.

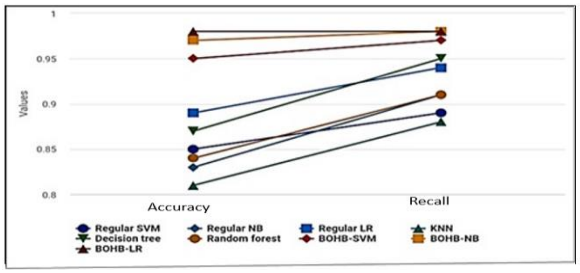

Figure 4: Models vs Accuracy, Recall over WDBC dataset

Figure 5 depicts the graphical representation of various state of art models concerning our 3 models over measure specificity and precision in which most of the algorithms have managed to bring such effectiveness in detecting breast cancer images in which the 3 models that are integrated with BOHB bring a greater performance 0.98,0.98 and 0.99 respectively.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11
Article Received: *10 September 2023* ***Revised:*** *20 October 2023* ***Accepted:*** *30 October 2023*

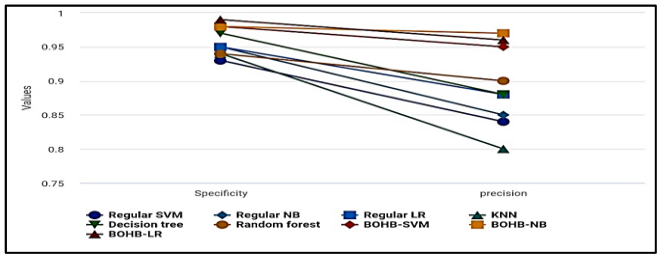

Figure 5: Models vs Specificity, Precision over WDBC dataset

Figure 6 illustrates the graphical representation of various state of art models concerning our 3 models over measure F1 score and recall where the performance of 3 models due to BOHB has been effective when compared to other regular models. Also, models like the random forest, KNN, and decision tree bring sufficient results without tuning and it is based on the structure of algorithms to detect them effectively.

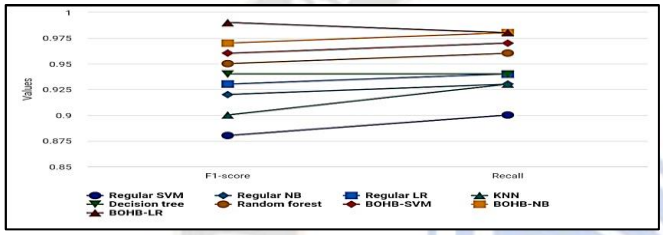

Figure 6: Models vs F1-Score, Recall over WDBC dataset

Figure 7 depicts the TPR and FPR rates of various models where the hyper-tuned models were able to detect most of the images as B and M and so do models like the random forest, KNN, and Decision tree. But regular models such as SVM, NB, and LR didn't give the expected results due to their limited structure and much computation time.

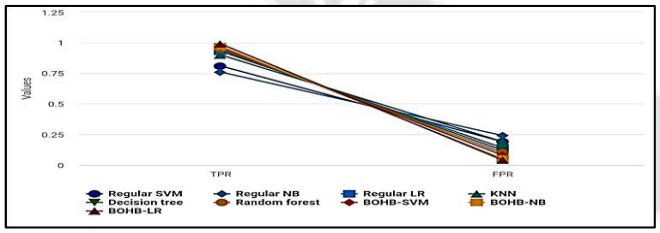

Figure 7: Models vs TPR, FPR over WDBC dataset

Figure 8 depicts the computation time of each model where our model's computation is less (0.9) compared to other models.

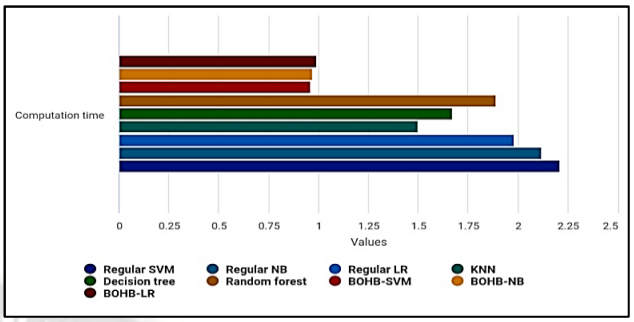

Figure 8: Models vs Computation Time in msec

Experimental results shown in Table 4 indicate the feature selection technique used by ours (PCA) is compared with other selection techniques over measures like entropy, skewness, kurtosis, contrast, correlation, and coarseness in which the importance of PCA is visible. Table 4 shows the effect of the feature selection method on our proposed methods (BOHB-SVM, BOHB-NB, BOHB-LR) in which it is visible on all measures taken for PCA gives a much better performance rate (0.6534) when compared to other feature selection methods. This rate will give an extra boost to the accuracy of each classifier for analyzing the breast cancer images.

To check whether the feature selection used by our system reduces overfitting, improves accuracy, and reduces training time, an evaluation of these selection techniques is required. Entropy, which quantifies the unpredictable nature of random variables and scales the amount of information they effectively share, Skewness, which evaluates how "tailed" distribution of a real-valued random variable is, Kurtosis, which quantifies how close two parameters are to having a linear relationship, and correlation, which quantifies how linearly dependent two factors are and thus how nearly identical their effects are on the dependent variable (synonymous of picture quality). These measurements, which are often referred to as statistical studies, show how texture affects the spatial organization of the gray values and how they interact with the environment.

Table 4: Comparison of Various Feature Selection Methods

Performance Measure	PCA	Chi-square Test	Fisher's Score	RFE	LASSO	Random Forest
Entropy	0.6534	0.9456	3.033	2.099	0.8765	0.4657
Skewness	0.0011	0.0055	0.0987	0.0105	0.0789	0.0984
Kurtosis	2.74E-05	2.34E-06	1.34E-06	2.44E-06	2.04E-05	1.88E-06
Contrast	0.3452	0.2650	0.6574	0.4563	0.9342	0.9845
Correlation	0.9875	0.9864	0.983	0.9674	0.9866	0.9823
Coarseness	13.856	8.876	11.897	10.876	11.998	9.865

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11
Article Received: *10 September 2023* ***Revised:*** *20 October 2023* ***Accepted:*** *30 October 2023*

Figures 9,10 and 11 depict the graphical representation of various feature selection techniques with our method (PCA) over measures like entropy, skewness, kurtosis, contrast, correlation, and coarseness in which PCA gives effective feature selection performance compared to other models over breast cancer image.

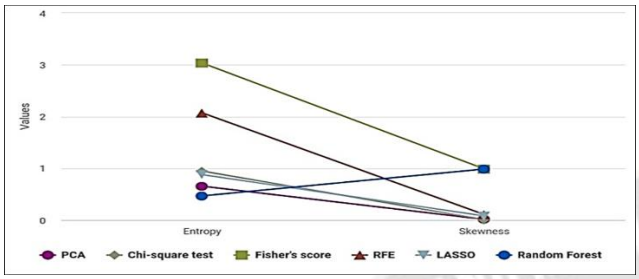

Figure 9: Feature Selection Technique vs Entropy, Skewness

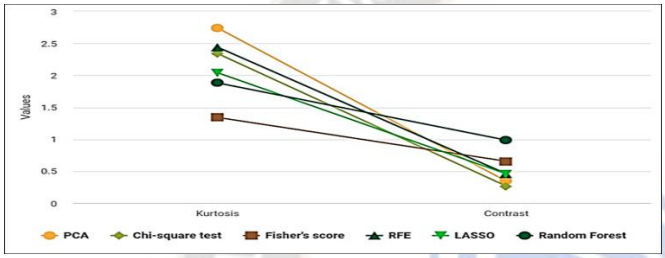

Figure 10: Feature Selection Technique vs Kurtosis, Contrast

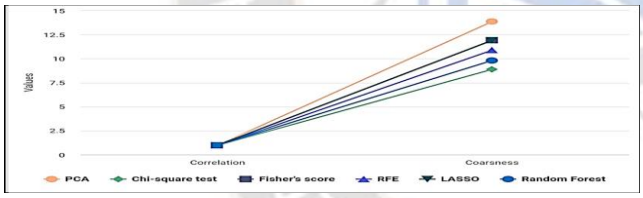

Figure 11: Feature Selection Technique vs Correlation, Coarseness

According to the experimental outcomes provided in Table 5, the PCA coupled with BOHB, as well as numerous models (LR, SVM, and NB), can greatly shorten the time needed for breast cancer detection. This mixture offers a theoretical foundation for the accurate diagnosing of breast cancer as well as the capability to do it quickly and with high accuracy. Table 5 compares the proposed models (BOHB-SVM, BOHB-NB, and BOHB-LR) with other cutting-edge models developed by other scientists based on metrics like accuracy, computation time, detection rate, and memory usage.

Table 5: Comparison of Authors' Models with Proposed Method

Authors	Accuracy	Computation Time	Detection Rate	Memory Utilization
DhahriHabibet al. [12]	0.84±0.48	6.5	84	14.3
Sadhukhan et al. [17]	0.89±0.55	5.7	82	15.5
Benbrahim et al. [18]	0.93±0.63	5.3	88	16.1
Negi et al. [20]	0.87±0.52	5.2	87	8.9
Asri et al. [22]	0.88±0.36	6	84	14.3
BOHB-SVM	0.95±0.60	4.6	91	11.3
BOHB-NB	0.98±0.50	4.5	94	10.7
BOHB-LR	0.97±0.45	4.7	94	8.9

The graphical depiction of several models presented by authors compared to our proposed models for accuracy, computation time, detection rate, and memory usage is shown in Figures 12, 13, 14, and 15.

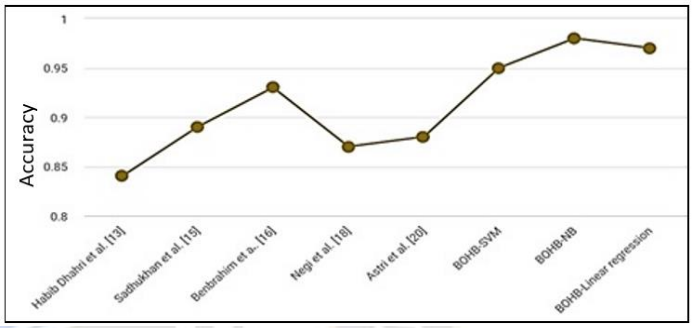

Figure 12: Authors and Proposed Models vs Accuracy

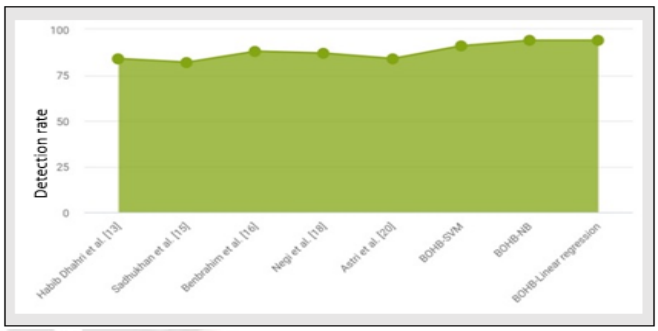

Figure 13: Authors and Proposed Model vs Detection Rate

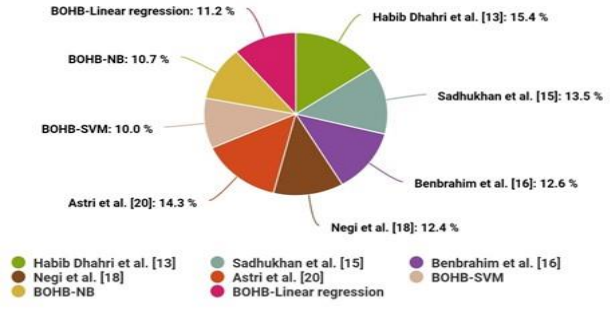

Figure 14: Authors and Proposed Method vs Computation Time

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11
Article Received: *10 September 2023* ***Revised:*** *20 October 2023* ***Accepted:*** *30 October 2023*

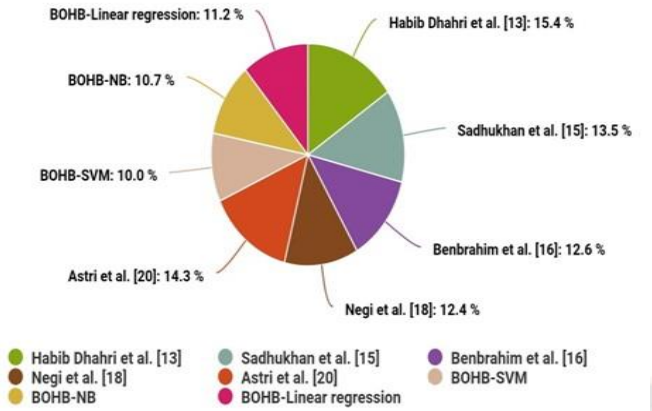

Figure 15: Authors and Proposed Method vs Memory Utilization

From figures 14 and 15, it is analyzed that the proposed approach obtains better accuracy, Computation time, detection rate, and memory utilization compared to other models. Dhahri Habibet al. [16], Sadhukhan et al. [19], Benbrahim et al. [20], Negi et al. [22], and Asri et al. [24] designed the CAD model with accuracy ranges from 84, 89, 93, 87 and 88, computation time ranges from 6.5, 5.7, 5.3, 5.2 and 6, detection rate ranges from 84, 82, 88, 87 and 84 and memory utilization ranges from 14.3, 15.5, 16.1, 8.9 and 14.3.

But our models such as BOHB-SVM, BOHB-NB, and BOHB-LR obtain accuracy ranges from 95, 98, and 97, computation time ranges from 4.6, 4.5, and 4.7, detection rate ranges from 91, 94, and 94 and memory utilization ranges from 11.3, 10.7 and 8.9. From that, it is concluded that the proposed approaches outperform well compared to the conventional model. This method is mostly utilized by oncologists to find and diagnose brain cancer in its early stages.

5. Conclusion

To increase the accuracy of diagnosis breast cancer and mortality rate caused by breast cancer, a variety of machine learning models can be employed to analyze diverse medical datasets. Enhanced machine learning optimization methods can be used to do it. The main challenge for machine learning is to make the prediction classifier for accurate classification in medical diagnosis. Each machine learning method needs variable tweaking to update the model's characteristics and produce a finely tailored one. And need to set their best value to achieve high efficiency.

This research assessed: SVM, NB, and LR. The dataset, WDBC is used for categorization. The classification accuracy of the models is enhanced by combining the BOHB optimization approach and the machine learning classification approach. To identify the most correct categorization in regard to model precision and computation time, this study examined a variety of machine learning algorithms using hyperparameter tweaking techniques. The proposed machine learning scheme as BOHB-NB achieved a better accuracy performance of 98.50% and a better computation time of 0.732 ms.

References

1. Alanazi, Abdullah. Using machine learning for healthcare challenges and opportunities. Informatics in Medicine Unlocked 30 (2022): 100924.
2. Laishram, Romesh, and RinkuRabidas. WDO optimized detection for mammographic masses and its diagnosis: A unified CAD system. Applied Soft Computing 110 (2021): 107620.
3. Wu J, Hicks C. Breast Cancer Type Classification Using Machine Learning. Journal of personalized medicine. 2021 Feb;11(2):61.
4. López, N. C., García-Ordás, M. T., Vitelli-Storelli, F., Fernández-Navarro, P., Palazuelos, C., &Alaiz-Rodríguez, R. (2021). Evaluation of Feature Selection Techniques for Breast Cancer Risk Prediction. International Journal of Environmental Research and Public Health, 18(20), 10670.
5. Punitha, S., Al-Turjman, F., & Stephan, T. (2021). An automated breast cancer diagnosis using feature selection and parameter optimization in ANN. Computers & Electrical Engineering, 90, 106958.
6. Tian, C. J., Lv, J., & Xu, X. F. (2021). Evaluation of Feature Selection Methods for Mammographic Breast Cancer Diagnosis in a Unified Framework. BioMed Research International, 2021.
7. Xie T, Wang Z, Zhao Q, Bai Q, Zhou X, Gu Y, Peng W, Wang H. Machine learning-based analysis of MR multiparametric radionics for the subtype classification of breast cancer. Frontiers in oncology. 2019 Jun 14; 9:505.
8. Bacha, Sawssen, and Okba Taouali. "A novel machine learning approach for breast cancer diagnosis." Measurement 187 (2022): 110233.
9. Rao, K., Gladence, L., & Lakshmi, V. (2019). Research of Feature Selection Methods to Predict Breast Cancer. International Journal of Recent Technology and Engineering (IJRTE), 8, 2353-2355
10. Schratz P, Muenchow J, Iturritxa E, Richter J, Brenning A. Performance evaluation and hyperparameter tuning of statistical and machine-learning models using spatial data. arXiv preprint arXiv:1803.11266. 2018 Mar 29.
11. Amrane M, Oukid S, Gagaoua I, Ensari T. Breast cancer classification using machine learning. In2018 Electric Electronics, Computer Science, Biomedical Engineerings' Meeting (EBBT) 2018 Apr 18 (pp. 1-4). IEEE.
12. Abdel-Ilah L, Šahinbegović H. Using machine learning tool in the classification of breast cancer. InCMBEBIH 2017 2017 (pp. 3-8). Springer, Singapore.
13. Obaid OI, Mohammed MA, Ghani MK, Mostafa A, Taha F. Evaluating the performance of machine learning techniques in the classification of Wisconsin Breast Cancer. International Journal of Engineering & Technology. 2018 Dec;7(4.36):160-6.

14. Dhahri H, Al Maghayreh E, Mahmood A, Elkilani W, Faisal Nagi M. Automated breast cancer diagnosis based on machine learning algorithms. Journal of healthcare engineering. 2019 Nov 3;2019.
15. Asri, H Mousannif, H Al Moatassim, et al. A hybrid data mining classifier for breast cancer prediction Adv Intel Sys Comp, 1103 (2020), pp. 9-16, 10.1007/978-3-030-36664-3_2
16. Sadhukhan, S., Upadhyay, N., & Chakraborty, P. (2020). Breast cancer diagnosis using image processing and machine learning. In Emerging Technology in Modelling and Graphics (pp. 113-127). Springer, Singapore.
17. Benbrahim, H., Hachimi, H., & Amine, A. (2019, July). Comparative study of machine learning algorithms using the breast cancer dataset. In International Conference on Advanced Intelligent Systems for Sustainable Development (pp. 83-91). Springer, Cham.
18. Osmanović, A., Halilović, S., Ilah, L. A., Fojnica, A., &Gromilić, Z. (2019). Machine learning techniques for classification of breast cancer. In World Congress on Medical Physics and Biomedical Engineering 2018 (pp. 197-200). Springer, Singapore.
19. Negi, R., & Mathew, R. (2018, December). Machine Learning Algorithms for Diagnosis of Breast Cancer. In International Conference on Computer Networks, Big Data and IoT (pp. 928-932). Springer, Cham.
20. Al Bataineh, A. (2019). A comparative analysis of nonlinear machine learning algorithms for breast cancer detection. International Journal of Machine Learning and Computing, 9(3), 248-254.
21. Asri, H., Mousannif, H., Al Moatassime, H., & Noel, T. (2016). Using machine learning algorithms for breast cancer risk prediction and diagnosis. Procedia Computer Science, 83, 1064-1069.
22. Yu, X., Xia, K., & Zhang, Y. D. (2021). DisepNet for breast abnormality recognition. Computers & Electrical Engineering, 90, 106961.
23. Li, H., Zhuang, S., Li, D. A., Zhao, J., & Ma, Y. (2019). Benign and malignant classification of mammogram images based on deep learning. Biomedical Signal Processing and Control, 51, 347-354.
24. Kumar, P., & Nair, G. G. (2021). An efficient classification framework for breast cancer using hyperparameter tuned Random Decision Forest Classifier and Bayesian Optimization. Biomedical Signal Processing and Control, 68, 102682.
25. Mohanty, F., Rup, S., & Dash, B. (2020). Automated diagnosis of breast cancer using parameter-optimized kernel extreme learning machine. Biomedical Signal Processing and Control, 62, 102108.
26. Alshayeji, M. H., Ellethy, H., & Gupta, R. (2022). Computer-aided detection of breast cancer on the Wisconsin dataset: An artificial neural networks approach. Biomedical Signal Processing and Control, 71, 103141.
27. Subasree, S., Sakthivel, N. K., Tripathi, K., Agarwal, D., & Tyagi, A. K. (2022). Combining the advantages of radiomic features based feature extraction and hyper parameters tuned RERNN using LOA for breast cancer classification. Biomedical Signal Processing and Control, 72, 103354.
28. Muduli, D., Dash, R., & Majhi, B. (2020). Automated breast cancer detection in digital mammograms: A moth flame optimization based ELM approach. Biomedical Signal Processing and Control, 59, 101912.
29. Li L, Jamieson K, DeSalvo G, Rostamizadeh A, Talwalkar A. Hyperband: A novel bandit-based approach to hyperparameter optimization. The Journal of Machine Learning Research. 2017 Jan 1;18(1):6765-816
30. He, S., Guo, F. and Zou, Q., 2020. MRMD2. 0: a python tool for machine learning with feature ranking and reduction. Current Bioinformatics, 15(10), pp.1213-1221..
31. Song F, Guo Z, Mei D. Feature selection using principal component analysis. In2010 international conference on system science, engineering design and manufacturing informatization 2010 Nov 12 (Vol. 1, pp. 27-30). IEEE.
32. Springenberg, J., Klein, A., Falkner, S., and Hutter, F. Bayesian optimization with robust bayesian neural networks. In Advances in Neural Information Processing Systems, 2016.
33. J. Snoek, H. Larochelle, and R. P. Adams, "Practical Bayesian optimization of machine learning algorithms," in Advances in neural information processing systems, 2012, pp. 2951–2959.
34. 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," Proceedings of the IEEE, vol. 104, no. 1, pp. 148–175, 2016
35. Falkner S, Klein A, Hutter F. BOHB: Robust and efficient hyperparameter optimization at scale. InInternational conference on machine learning 2018 Jul 3 (pp. 1437-1446). PMLR..