ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20240227075152

Original scientific paper
Received: 2024-02-27, Accepted: 2024-08-01

Breast Cancer Diagnosis Using Machine Learning and PSO

Sarita Silaich*, Rajesh Yadav

Abstract: Healthcare systems around the world are facing huge challenges in responding to trends of the rise of chronic diseases. Early detection of breast cancer is essential for successful treatment since it is a common and potentially fatal condition. Based on clinical data, machine learning algorithms have shown potential in the categorization of breast cancer. This work aimed to build classification models Support Vector Machine (SVM), K-Nearest Neighbours (KNN), Random Forest (RF), Logistic Regression (LR), and Artificial Neural Network (ANN) for Diagnostic Wisconsin Breast Cancer Dataset (WDBC) also improves classifier performance by using feature selection optimization and cross validation. The Particle Swarm Optimization (PSO) technique is used to select relevant, irredundant and most informative features that dependent on the performance of a classifier utilizing certain characteristics. The effectiveness of five distinct classifiers is assessed in this work. According to the findings, PSO-based SVM classifier has the greatest mean subset accuracy over a wide variety of training testing ratios.

Keywords: breast cancer; cross validation; Diagnostic Wisconsin Breast Cancer Database (WDBC); machine learning classifiers; Particle Swarm Optimization (PSO)

1 INTRODUCTION

Accurately diagnosing and detecting different ailments and diseases is a major problem in the fascinating and ever-evolving disciplines of bioinformatics and medical research. This effort necessitates not only a wealth of knowledge and skills, but also creative thinking and approach. Disease diagnosis is a challenging and dynamic field in medicine. An abundance of data on medical diagnoses may be accessible at clinics, healthcare facilities, colleges, and various research centers, as well as on the internet [1]. To make the system fully automated and rapid in diagnosing illnesses, classification is vastly used. Diseases are often diagnosed depending on the medical expertise of the medical planning officer. As a result of this, accurate illness diagnosis can be complicated and accompanied by cases of errors and unintended bias [2].

According to the report of WHO 2020, breast cancer is a major health concern and it afflicted 2.3 million individuals and killed 685,000, people. It is the world's most frequent illness and impacted 7.8 million individuals in the same year. Breast cancer kills more women than any other type of cancer worldwide. The disease affects women all around the world after puberty, with rates increasing with age. From 1990 to 2016, the age-standardized incidence rate of breast cancer in women rose by 39.1 % nationwide (with a 95 % uncertainty range of 5.1 to 85.5), according to Globocan statistics Breast cancer accounted for 13.5 % (178,361), of all cancer cases and 10.6 % (90,408), of total deaths in India in 2020 [3]. Breast cancer is a significant health concern, especially for women, as it is the most ubiquitous type of cancer, as indicated by the data presented.

As per UK cancer statistics 41,000 women diagnosed breast cancer each year. Overall, breast cancer is the leading cancer in women globally. Accurately diagnosis breast cancer is crucial since the illness apparent in a variety of ways that need specialized testing and analysis. To deliver accurate diagnosis and personalized treatment regimens customized to the particular requirements of patients, a multidisciplinary approach integrating the experience of medical professionals, researchers, and bioinformaticians is required. To create novel diagnostic tests and treatment techniques, sophisticated computational tools are employed to analyze massive volumes of genomic data in search of patterns and mutations related with breast cancer [4].

As a result of having to examine a substantial quantity of image data during mammography, the degree of accuracy achieved by the test is diminished. This technique requires a considerable amount of time, and in the worst-case situation, it might yield an erroneous diagnosis of the ailment. In this research, we investigate and compare a wide variety of machine learning algorithms, each of which is capable of recognizing the sickness on the basis of the input features. Five distinct supervised machine learning strategies were applied in order to accurately diagnose the problem.

2 RELATED WORKS

The progress that has been made in the field of medical research has led to the development of a number of innovative breast cancer detection methods. The research that is pertinent to this field is briefly summarized in the following paragraphs. The challenge of correctly detecting breast cancer (BC) may be thought of as a classification issue. Researchers are applying a wide array of machine learning (ML) techniques, artificial neural networks, support vector machine algorithms, and a number of other data extraction methods in order to conduct an analysis of it. Because of their capacity to successfully capture complex nonlinear interactions among variables, the ANN, SVM, and PNN models have acquired real-world relevance in the area of classification modeling. This is due to the fact that they may be used to model classification problems. It has been shown that functional weaving is a very effective method for identifying patterns statistically while also providing a nonlinear forecast of their complexity. It may be possible to profit from the use of machine learning in order to acquire a diagnosis that is both reliable and cost-effective. A method of machine learning known as the feed-forward artificial neural network (FFANN) is described here. This study aims to identify and classify breast cancer patients so that further research may be conducted [5].

In order to overcome challenges associated with optimization, metaheuristic optimization makes use of

algorithms that are based on heuristics. Rami N. Khushaba and his colleagues [6] developed a method for selecting features that makes use of a strategy that involves differential evolution (DE) optimization in conjunction with a repair mechanism that is based on feature allocation evaluations. This method was proposed as a means of selecting features. In order to resolve the combinatorial optimization problems that is associated with featureselection, DEFS implements the distributed float number optimization. The construction of a wheel-like structure and the provision of opportunities for the distribution of features were carried out with the intention of making it easier to pick features via the use of the float optimizer. Within data sets that vary in the degree of dimensionality, DEFS has been used to search for optimal segments of characteristic characteristics in an effort to locate optimal solutions.

Gu et al. [7] have recently introduced a novel variant of the particle swarm optimization (PSO) algorithm, termed as competitive swarm optimizer (CSO), which is specifically designed to address the challenges of high- dimensional feature selection in large-scale optimization problems. The CSO, which was initially designed for uninterrupted operation, was included.

The concept of optimization originated with the aim of conducting feature selection, which can be viewed as a problem of combinatorial optimization. An approach for documentation was implemented with the aim of decreasing computationally expenses. The authors conducted experiments on six standard datasets and compared their proposed CSO-based feature decision-making algorithm with a canonical PSO-built algorithm and innovative PSO variants. Their findings showed that the CSO-based algorithm selected a significantly fewer number of features while accomplishing the best classification accuracy.

Moradi et al. [8] proposed the utilization of localized search tactics, integrated into particle swarm optimization, for finding a subset of features that is both less related as well as salient. This approach was referred to as HPSO-LS. The key goal of employing the local hunt technique was to facilitate the algorithm for searching of the particle swarm optimization in selecting distinctive features by considering their correlation data. Additionally, the proposed approach employs a scheme for determining the size of the subset in order to select a reduced set of features.

The usefulness of the proposed tactic has been evaluated on a set of 13 standard classification tasks and compared to five contemporary methods for selecting features. Furthermore, HPSO-LS has been compared to four established filter-based techniques, namely information gain, terms variation, fisher score, and mRMR, as well as five established wrapper-based techniques such as particle swarm optimization, genetic algorithm, simulated annealing, and ant colony optimization . The findings indicate that the proposed approach enhances the classification accuracy to a level comparable to that of filter-based and wrapper-based feature selection techniques [9].

More particularly, a technical taxonomy of the chosen material, including hybridization, improvement, and PSO variations, is examined in this work together with previous research on methodologies and applications published between 2017 and 2019 [10]. They are a targeted application of the algorithm categorized for the actual world. SVM and ANN classification algorithms were used by Bayrak et al. in their work to forecast breast cancer using a Wisconsin Breast Cancer Data Collection. The Sequential Minimal Optimization (SMO) and LibSVM methods were used to classify Support Vector Machines (SVM). The authors used the WEKA programming tool to categorize Support Vector Machine (SVM) using Multilayer Perceptron (MLP) and accepted perception methods. By combining SMO-SVM with the 10-fold cross-validation method, the authors were able to attain a high accuracy of 96.9957 % [11].

Sakri et al. [12] focus was on increasing accuracy by combining the PSO algorithm with the MI algorithms K-NNs, Naive Bayes (NB), and (REP) tree. One of the biggest issues in Saudi Arabia, according to their research, is the prevalence of breast cancer among women. Their research indicates that ladies over the age of 46 are the disease's main victims. The authors utilized five phase-based data analysis approaches to the WBCD dataset. Their final report was based on a comparison of feature selection methods used for classification with and without them. Juneja K, Rana [13] created updated decision tree technique, known as a weight-enhanced decision tree to detect breast cancer on dataset that was retrieved from the UCI repository. Through the use of the Chi-square test, they concluded that they had rated each characteristic and kept those that were relevant to this categorization assignment. Their suggested strategy produced results of around 98 % and 85-90 % precision on the Wisconsin Breast Cancer Diagnosis (WBCD) benchmark dataset, Yue et al. [14] reported thorough analyses of four classifiers, including SVM, K-NNs, ANNs, and Decision Tree approaches for predicting breast cancer. For training and assessment, four-fold cross-validation is used. SVM obtains the best levels of accuracy, specificity, and sensitivity during the training phase, with scores of 97.71 %, 98.9 %, and 97.08 %, respectively. Senapati et al. [15] used local linear wavelet neural network for breast cancer recognition. The Recursive Least Square (RLS) approach enhances the efficacy of training parameters, notably refining the suggested model to unveil the intricate connection weights among neurons in both the hidden and output layers. Through comparison with traditional methods, this approach proves its resilience and robustness, showcasing its superiority in performance.

3 MATERIALS AND METHODS

3.1 Breast Cancer Dataset

The Breast Cancer Dataset, commonly known as the Diagnostic Wisconsin Breast Cancer Database (WBCD), was collected from the UCI machine learning repository. Dr. William H. Wolberg produced the dataset in the late 1980s, and it is accessible to the public via the UCI repository [16]. It includes details about clinical dataset 569 patient breast tissue samples. It includes different attributes calculated from digital photographs of the samples. The objective of the dataset is to predict whether a tissue sample is benign or malignant based on the attributes. In this case 357 (62.74 %) are benign and 212 (37.26 %) are malignant. With the exception of the patient id and diagnosis level, the attributes in the dataset shown in Tab. 1. In our study benign instances

are regarded as the positive class and malignant cases as the negative class. The objective is to create a model that can correctly predict whether a certain instance is benign or malignant based on the values of the various features in the dataset.

Table 1 Attributes and Descriptions WDBC [16]

Sno.	Attributes Name	Description of Attribute
1	Radius	Mean of distance from center to points on cell nucleus perimeter
2	Texture	Standard deviation of gray scale values
3	Perimeter	Perimeter of Tumor
4	Area	Area of Tumor
5	Smoothness	Local variation in radius lengths
6	Compactness	(Perimeter^2/Area) – 1
7	Concavity	Severity of concave partitions of the contour
8	Concave Point	Number of concave partitions of the contour
9	Symmetry	Symmetry in the cell nuclie
10	Fractal Dimension	Coastline Approximation – 1

3.2 Pre-Processing Dataset

The transformation of raw data into a format that is appropriate for machine learning activities is what happens during the data pre-processing stage, which is an essential part of the data analysis pipeline. It seeks to enhance the quality of the data, get rid of errors, deal with missing values, and standardize the data so that it may be used by a variety of algorithms in an efficient manner. Because real-world datasets often include noise, mistakes, and variances, which may have a detrimental influence on the performance and accuracy of machine learning models, pre-processing of the data is very necessary [17].

3.3 Feature Extraction

To handle datasets with a large number of features, we used feature extraction methods to lower the combined dataset's dimensionality. High-dimensional datasets may result in overfitting and longer calculation times. Principal Component Analysis (PCA) and other dimensionality reduction algorithms are examples of feature extraction techniques that assist maintain important information while reducing the number of features. This step is critical in simplifying the dataset and improving classifier performance.

3.4 Data Normalization

Data normalization is a critical pre-processing step to ensure unbiased learning and feature scaling across all features. Since the datasets were combined from different sources, they might have different scales and ranges for their attributes [18]. Data must be transformed during normalization such that each attribute has a mean of 0 and a standard deviation of 1. This method equalizes the scale of all characteristics, preventing certain features from predominating the learning process due to their magnitudes.

$$X_{\text{normalized}} = \frac{X - X_{\min}}{X_{\max} - X_{\min}}.$$

3.5 Particle Swarm Optimization

In the case of breast cancer diagnosis, we used PSO for feature selection to find a subset of significant traits that may differentiate between benign and malignant tumors. This allows us to more accurately diagnose patients with breast cancer. The PSO algorithm's goal is to maximize a fitness function, which is often based on how well the classifier performs while using certain characteristics. In order to begin the process, a swarm of particles must be initialized, each of which will represent a possible feature subset. These particles move around the feature space by calculating velocity using Eq. (1) and update position as in Eq. (2) to search of the best possible combination of features. This process is repeated with the end goal of improving the classification accuracy.

$$v_{ij}^{t+1} = wv_{ij}^{t} + c_1 r_1 \left[pBest_{ij}^{t} - x_{ij}^{t} \right] + c_2 r_2 \left[gBest_{ij}^{t} - x_{ij}^{t} \right], \quad (1)$$

$$x_{ij}^{t+1} = x_{ij}^{t} + v_{ij}^{t+1}, \quad (2)$$

v_{ij}^{t} is velocity of i^{th} partical of j^{th} dimension at time t and x_{ij}^{t} is position of same, w is inertia weight, c_1, c_2 are cognitive learning factor, r_1, r_2 uniformly distributed random number between 0 and 1, *pBest* is its personal best value and *gBest* is global best value. Then position is converted in binary using sigmoid function.

$$x_{ij}^{t+1} = \begin{cases} 0 \text{ if } rand(\ ) \geq Sigmoid\left(v_{ij}^{t+1}\right) \\ 1 \text{ if } rand(\ ) \geq Sigmoid\left(v_{ij}^{t+1}\right) \end{cases}, \quad (3)$$

$$Sigmoid\left(v_{ij}^{t+1}\right) = \frac{1}{\left(1 + e^{-v_{ij}^{t}}\right)}. \quad (4)$$

3.6 Machine Learning

The pre-processed dataset was used in the technical paper that was presented, and five different machine learning classifiers were used to diagnose breast cancer based on the data. The following classifiers that are used.

3.6.1 Support Vector Machine (SVM)

It is a sophisticated and extensively used classification technique that seeks to discover an ideal hyperplane that divides various classes in the feature space. This goal of the method is to find an optimal separation between classes. It performs well with both data that can be separated linearly and data that can be separated non-linearly [20].

3.6.2 K-closest Neighbours (KNN)

It is a basic yet effective classification method that classifies a data point based on the majority class among its k closest neighbours in the feature space. This technique

classifies a data point based on the majority class among its *k* nearest neighbours in the feature space [21].

3.6.3 Random Forest (RF)

It is often known as RF, is a strategy for ensemble learning that mixes numerous decision trees in order to enhance accuracy and decrease overfitting. Every tree receives its training based on a different selection of characteristics and data points [22].

3.6.4 Logistic Regression (LR)

It is a method of binary classification that makes use of a logistic function in order to make forecasts about the likelihood of a given instance belonging to a certain class. These forecasts may be made based on the function. To put it another way, it calculates the chance that a certain category is satisfied by a given occurrence [23].

3.6.5 Artificial Neural Network (ANN)

It is a model of reasoning based on the human brain. A conventional ANN model contains a hierarchy of layers: input layer, hidden layers and output layer which are composed of interconnected neurons containing an activation function for nonlinear transformation. In ANN model input layer receives the data also called features and transmits the data to a hidden layer where data is processed and trained results are provided at the output layer. Some ANN training process may involve long causal chains of computational stage depending on complexity of the problem.

3.7 Methodology

Methodology flow chart is shown in Fig. 1 which starts by collecting Wisconsin Breast Cancer Diagnostic Dataset. After applying normalization, ten cross-validations applied and then the performance of various classification algorithms is evaluated both using PSO as feature selection and without PSO.

A well-known method for accurately evaluating the effectiveness of classifiers is called K-fold cross-validation as it tune the hyper parameters, and it was used here to evaluate each of these classifiers. During 10-fold cross-validation, the dataset is divided into ten subsets known as folds. Each classifier is then trained and tested ten times, with each iteration uses a different fold from the dataset as the test set. This will apply the appropriate values of hyper parameters to produce high performance that are more trustworthy and can be used more broadly.

Performance is evaluated using confusion matrix as shown in Tab. 2.

Table 2 Confusion matrix

		Predicted	
		N	P
Actual	N	True Negative (*TN*)	False Positive (*FP*)
	P	False Negative (*FN*)	True Positive (*TP*)

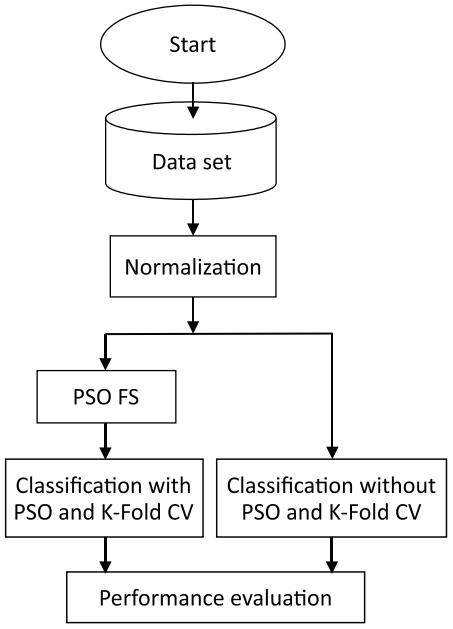

Figure 1 Methodology

A good model is a model that can predict correctly label. Accuracy, recall (also called sensitivity or True Positive Rate), and precision equation are given as below.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}, \tag{5}$$

$$Recall = \frac{TP}{TP + FN}, \tag{6}$$

$$Precision = \frac{TP}{TP + FP}. \tag{7}$$

4 RESULTS

Applying PSO feature selection the Tab. 3 is resultant of the selected features. Tab. 3 show the accuracy using selected feature by PSO (acc_pso) and using all features (acc_all), precision using PSO (pre_pso) and precision using all featues (pre_all), recall using PSO (re_all) and recall using all features (re_pso) for different training ratios. It is observed that except ANN all other remaining classifiers have improved accuracy with PSO. SVM using PSO has highest testing accuracy 97.81 %.

Result shows that using PSO performance of all classifiers is improving when size of test data is increasing that is logical but without PSO performance is not in increasing order as number of samples increasing in test data. Recall or sensitivity is improved for all classifiers means false negative is less, which is more important in medical because more false positive is acceptable but high false negative is more dangerous as disease becomes incurable. Also using PSO resultant less features trained models are simple and more generalized without compromising accuracy.

The graph representations of accuracy analysis of all algorithms with feature selection and without feature

selection are shown in following graphs. That shows that using PSO as FS enhanced the accuracy for several reasons:

Feature Selection: The WDBC dataset likely contains numerous features, and not all of them might be relevant or contribute equally to the classification task. PSO helps in selecting a subset of features that are more discriminative, reducing noise and irrelevant information. This can lead to better classification accuracy by focusing on the most informative features.

Table 3 Test result analysis

SVM	acc_all	acc_pso	pre_all	pre_pso	re_all	re_pso
90-10	96.49	92.98	97.22	97.06	97.22	91.67
80-20	97.75	96.61	98.60	97.18	98.60	95.83
70-30	97.66	97.49	98.13	98.13	98.30	97.20
60-40	96.93	**97.81**	96.58	**97.92**	98.60	**98.60**
KNN						
90-10	98.25	96.49	97.30	97.22	99.00	97.22
80-20	98.25	93.86	97.30	97.10	98.00	93.06
70-30	97.60	94.74	96.40	95.37	98.00	96.26
60-40	96.05	95.18	95.89	95.21	97.90	97.20
RF						
90-10	96.49	94.74	97.22	97.14	97.22	94.44
80-20	94.74	95.61	95.83	97.18	95.83	95.83
70-30	95.32	95.32	95.41	95.14	97.20	97.20
60-40	94.74	96.49	95.17	96.55	96.50	97.90
LR						
90-10	96.49	98.25	97.22	97.30	97.22	99.00
80-20	95.01	96.49	94.67	95.95	98.61	98.61
70-30	94.74	94.74	92.98	92.24	99.07	99.00
60-40	95.80	95.61	94.00	94.04	98.60	99.30
ANN						
90-10	96.49	92.98	97.22	97.06	97.22	91.67
80-20	97.37	95.61	98.59	97.18	97.22	95.83
70-30	98.25	95.91	99.06	96.30	98.30	97.20
60-40	98.68	96.61	99.30	96.50	98.60	96.50

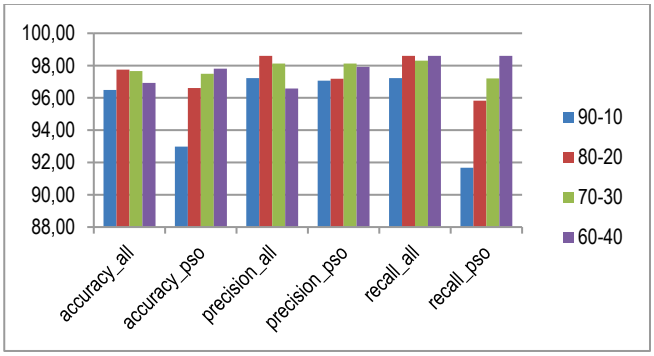

Figure 2 SVM classifier results with and without PSO

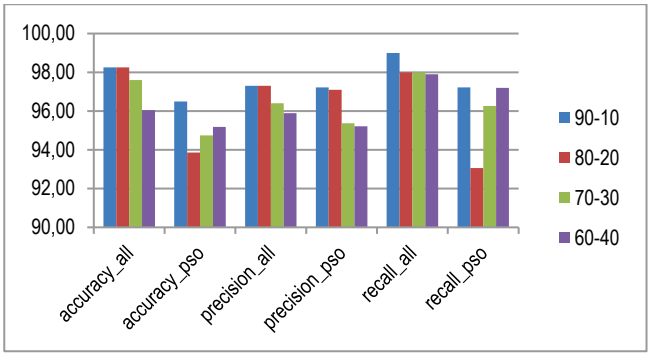

Figure 3 KNN with and without PSO

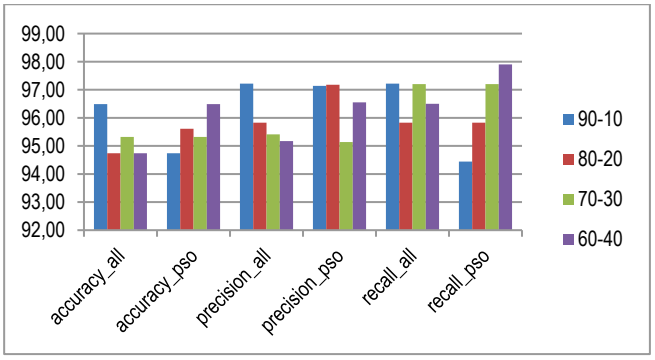

Figure 4 RF classifier results with and without PSO

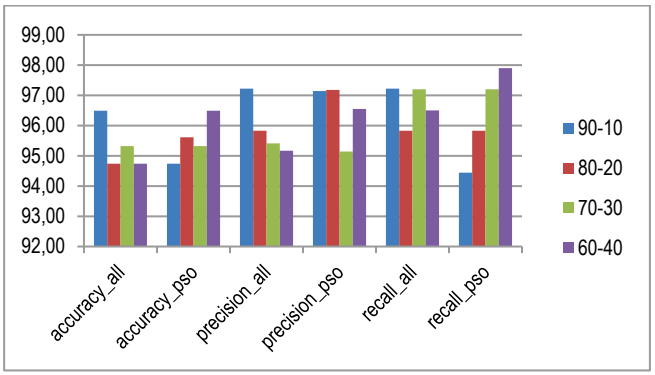

Figure 5 LR with and without PSO

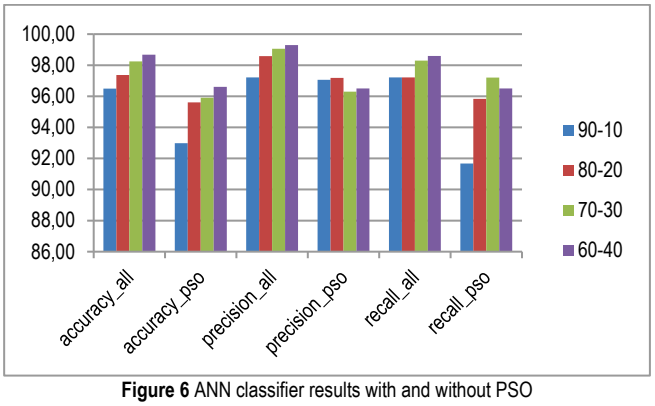

Figure 6 ANN classifier results with and without PSO

More insight into misclassification can be done using confusion matrix. For SVM train test split 60:40 (341 Training Samples and 228 test samples) confusion matrix are as follows:

Table 4 Confusion matrix for SVM with all Features

		Predicted	
		N	P
Actual	N	True Negative (*TN*) = 80	False Positive (*FP*) = 5
	P	False Negative (*FN*) = 2	True Positive (*TP*) = 141

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} = \frac{80+141}{80+141+5+2} = 96.93$$

$$Recall = \frac{TP}{TP+FN} = \frac{141}{141+2} = 98.60$$

$$Precision = \frac{TP}{TP+FP} = \frac{141}{141+5} = 96.58$$

Table 5 Confusion matrix for SVM and PSO Selected Features

		Predicted	
		N	P
Actual	N	True Negative (*TN*) = 82	False Positive (*FP*) = 3
	P	False Negative (*FN*) = 2	True Positive (*TP*) = 141

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} = \frac{82 + 141}{82 + 141 + 3 + 2} = 97.81$$

$$Recall = \frac{TP}{TP + FN} = \frac{141}{141 + 2} = 98.60$$

$$Precision = \frac{TP}{TP + FP} = \frac{141}{141 + 3} = 97.92$$

These results show that using PSO feature selection misclassifications are less as compared to using all features.

Search Space Exploration: PSO explores the search space efficiently by evaluating different combinations of features. It optimizes the feature subset selection by iterating through potential solutions and gradually converging towards an optimal or near-optimal solution. This exploration allows it to find a feature subset that works best for the classification task.

Handling Redundancy and Correlation: Sometimes, certain features in a dataset might be redundant or highly correlated. PSO can handle such cases by identifying and selecting only the most relevant features while discarding redundant or highly correlated ones. This prevents overfitting and improves the generalization of the classifier.

Improving Model Efficiency: By reducing the number of features, PSO can also lead to more efficient models in terms of computational resources and time. With fewer features as shown in Fig. 7, the model complexity decreases, potentially speeding up the training and inference processes.

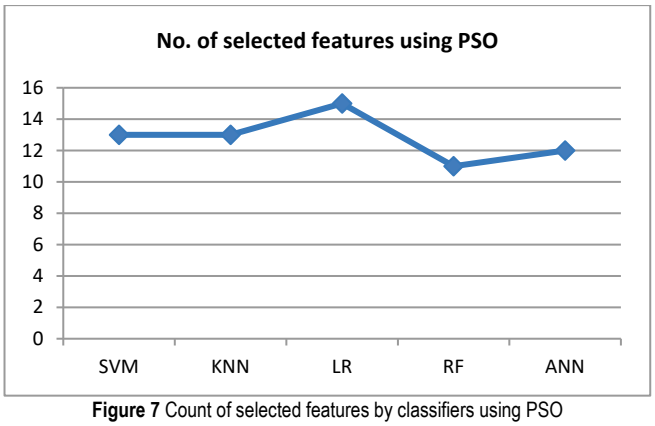

Figure 7 Count of selected features by classifiers using PSO

5 CONCLUSION

Our main aim was to explore how the integration of a feature selection algorithm with classification algorithms impacts breast cancer prognosis. Also 10-fold cross validation used to tune the hyper parameters and different split ratios applied to explore the result. By reducing the number of features, we aimed to highlight the significance and impact of specific features on the final results. We have shared findings from experiments on five widely-used classification algorithms: SVM, KNN, Random Forest, Logistic Regression and ANN for both with and without the PSO feature selection method. We can conclude that using less number of features selected by PSO classifiers accuracy is not compromised much, while using feature selection, models become simple and generalized. From results it can be concluded that SVM performed better using PSO in comparison with remaining classifiers.

6 FUTURE SCOPE

In our study we worked on single objective function as a fitness function on texted dataset. Future work can be to explore multi objectives function as fitness function in combination of PSO with deep learning on complex imaging and very large dataset.

7 REFERENCES

[1] Bhardwaj, R., Nambiar, A. R. & Dutta, D. (2017). A Study of Machine Learning in Healthcare. *The 41st IEEE Annual Computer Software and Applications Conference*, 236-241. https://doi.org/10.1109/COMPSAC.2017.164

[2] Kuehnert, M. J. & Qualls, C. (2019). Managing Healthcare Data Quality. 3rd ed. Chicago, IL: Health Administration Press.

[3] International Agency for Research on Cancer. India Source: Globocan 2020. Available from: https://gco.iarc.fr/today/data/factsheets/populations/356-india-fact-sheets.pdf

[4] Jang, B.-S. & Kim, I. A. (2020). Machine Learning Algorithms and Whole Exome Sequencing Data from Breast Cancer Patients in the UK Biobank Predict Survival. https://doi.org/10.21203/rs.3.rs-115867/v1

[5] Tripathy, S. (2020). Investigation of the FFANN Model for Mammogram Classification Using an Improved Gray Level Co-occurrences Matrix. *International Journal of Advanced Science and Technology, 29*. 4214.

[6] Khushaba, R. N., Al-Ani, A. & Al-Jumaily, A. (2011). Feature subset selection using differential evolution and a statistical repair mechanism. *Expert Systems with Applications, 38*(9), 11515-11526. https://doi.org/10.1016/j.eswa.2011.03.028

[7] Gu, S., Cheng, R. & Jin, Y. (2018). Feature selection for high-dimensional classification using a competitive swarm optimizer. *Soft Computing, 22*, 811-822. https://doi.org/10.1007/s00500-016-2385-6

[8] Moradi, P. & Mozhgan, G. (2016). A hybrid particle swarm optimization for feature subset selection by integrating a novel local search strategy. *Applied Soft Computing, 43*, 117-130. https://doi.org/10.1016/j.asoc.2016.01.044

[9] Xie, S. et al. (2022). Using SVM and PSO-NN Models to Predict Breast Cancer. In: Liu, Q., Liu, X., Cheng, J., Shen, T., Tian, Y. (eds) *Proceedings of the 12th International Conference on Computer Engineering and Networks, CENet 2022*. Lecture Notes in Electrical Engineering, vol 961. Springer, Singapore. https://doi.org/10.1007/978-981-19-6901-0_74

[10] Gad, A. G. (2022). Particle Swarm Optimization Algorithm and Its Applications: A Systematic Review. *Archives of Computational Methods in Engineering, 29*, 2531-2561. https://doi.org/10.1007/s11831-021-09694-4

[11] Mustapha, M. T., Ozsahin, D. U. et al. (2022). Breast Cancer Screening Based on Supervised Learning and Multi-Criteria Decision-Making. *Diagnostics, 12*, 1326. https://doi.org/10.3390/diagnostics12061326

[12] Sakri, S. B., Rashid, N. B. A. & Zain, Z. M. (2018). Particle swarm optimization feature selection for breast cancer recurrence prediction. *IEEE Access, 6*, 29637-29647. https://doi.org/10.1109/ACCESS.2018.2843443
[13] Juneja, K. & Rana, C. (2020). An improved weighted decision tree approach for breast cancer prediction. *Int. j. inf. tecnol. 12*, 797-804. https://doi.org/10.1007/s41870-018-0184-2
[14] Yue, W. et al. (2018). Machine learning with applications in breast cancer diagnosis and prognosis. *Designs, 2*(2), 13. https://doi.org/10.3390/designs2020013
[15] Senapati, M. R. & Mohanty, A. K. et al. (2013). Local linear wavelet neural network for breast cancer recognition. *Neural Computing Appl., 22*(1), 125-131. https://doi.org/10.1007/s00521-011-0670-y
[16] Wolberg, W. (1992). Breast Cancer Wisconsin (Original). UCI Machine Learning Repository. https://doi.org/10.24432/C5HP4Z
[17] Amato, A. et al. (2023). Data preprocessing impact on machine learning algorithm performance. *Open Computer Science, 13*. https://doi.org/10.1515/comp-2022-0278
[18] Zhao, Z. et al. (2019). Capsule Networks with Max-Min Normalization. https://doi.org/10.48550/arXiv.1903.09662
[19] Osareh, A. & Shadgar, B. (2010). Machine learning techniques to diagnose breast cancer. *The 5th International Symposium on Health Informatics and Bioinformatics*, Ankara, Turkey, 114-120. https://doi.org/10.1109/HIBIT.2010.5478895
[20] Singh, N. (2023). Support Vector Machine (SVM).
[21] Syriopoulos, P. K., Kotsiantis, S. B. & Vrahatis, M. N. (2022). Survey on KNN Methods in Data Science. In: Simos, D. E., Rasskazova, V. A., Archetti, F., Kotsireas, I. S., Pardalos, P. M. (eds) Learning and Intelligent Optimization. LION 2022. *Lecture Notes in Computer Science, vol 13621*. Springer, Cham. https://doi.org/10.1007/978-3-031-24866-5_28
[22] Lomakin, N., Pokidova, V. V. et al. (2023). Digital forecast of the efficiency of the enterprise based on the machine learning model Random forest. *Mezhdunarodnaja jekonomika (The World Economics)*. 411-425. (in Russian) https://doi.org/10.33920/vne-04-2306-06
[23] Alanazi, A. (2022). Using machine learning for healthcare challenges and opportunities. *Informatics in Medicine Unlocked, 30*, 100924. https://doi.org/10.1016/j.imu.2022.100924

Authors' contacts:

Sarita Silaich, Assistant Professor
(Corresponding author)
Government Polytechnic College,
W-6, Residency Road, Jodhpur, Rajasthan 342001, India
E-mail: sarita.bits@gmail.com

Rajesh Yadav, Assistant Professor Dr.
Mody University of Science & Technology,
Lakshmangarh 332 311, Dist. Sikar, Rajasthan, India
E-mail: yadav.rajesh27@gmail.com