IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 16 August 2024, accepted 24 September 2024, date of publication 27 September 2024, date of current version 28 October 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3469869

RESEARCH ARTICLE

Machine Learning-Assisted Cervical Cancer Prediction Using Particle Swarm Optimization for Improved Feature Selection and Prediction

EMMANUEL ILEBERI AND YANXIA SUN, (Senior Member, IEEE)
Department of Electrical and Electronic Engineering Science, University of Johannesburg, Johannesburg 2094, South Africa

Corresponding author: Emmanuel Ileberi (emmanuelileberi@gmail.com)

This work was supported in part by the South African National Research Foundation under Grant 141951, Grants nos. 137951 and AJCR230704126719120106.

ABSTRACT Cervical cancer is a common and deadly disease that affects women worldwide. Early diagnosis and treatment can improve the survival and quality of life of patients. Machine learning techniques can help to analyze complex and high-dimensional data related to cervical cancer and provide accurate and reliable predictions. However, selecting the most relevant and informative features from the data is a challenging task that affects the performance and interpretability of machine learning models. This paper proposes a novel method that uses particle swarm optimization (PSO) to perform feature selection and optimization for cervical cancer prediction. PSO is a bio-inspired algorithm that mimics the social behavior of a swarm of particles that search for the optimal solution in the feature space. The use of PSO to select the best subset of features that maximize the classification accuracy of eight machine learning models: Support Vector Machines (SVM), Gaussian Naive Bayes (GNB), Random Forests (RF), Decision Trees (DT), Extreme Gradient Boosting (XGB), Linear Regression (LR), Adaptive Boosting (AdaBoost), and K-nearest neighbor (KNN).To evaluate the method, a publicly available dataset was used, the Cervical Cancer Risk Factors Dataset (CCRFD). Then, compare the results with several state-of-the-art methods that use different feature selection techniques and ML algorithms. The experimental results show that the method achieves superior performance in terms of feature reduction rate, accuracy, precision, and AUC. Specifically, the Adaboost-PSO model performed best in terms of feature reduction rate with a reduction of rate 100% while the RF-PSO model performed best in terms of accuracy and precision, with an accuracy of 98% and precision of 100%.

INDEX TERMS Machine learning, feature selection, cervical cancer, particle swarm optimization.

I. INTRODUCTION

Cervical cancer is a type of cancer that affects the cervix, which is the lower part of the uterus that connects to the vagina. Cervical cancer is one of the most common and deadly cancers among women worldwide, especially in developing countries. According to the World Health Organization (WHO), cervical cancer is the fourth most frequent cancer in women, with an estimated 604,000 new cases and 342,000 deaths in 2020 [1].Global cancer statistics [2] show that cervical cancer affects 493,000 new patients every year, making up 15% of all female cancers. This disease is prevalent in developing countries [3], where it has a high mortality rate of 83% It is especially common in African countries [4], [5]. Cervical cancer is caused by persistent infection with certain types of human papillomavirus (HPV), which can be transmitted through sexual contact. Other risk factors for cervical cancer include smoking, multiple sexual partners, early sexual activity, long-term use of oral contraceptives, and weakened immune system. Some of the symptoms of cervical cancer include abnormal vaginal bleeding, pain during sex, pelvic pain, and vaginal discharge. Cervical cancer can be diagnosed by various methods, such as Pap smear test, HPV test, colposcopy, biopsy, etc. Cervical

The associate editor coordinating the review of this manuscript and approving it for publication was Essam A. Rashed.

cancer can be treated by various methods, such as surgery, radiation therapy, chemotherapy, immunotherapy, etc [6], [7], [8], [9].

Machine learning has been widely applied to various domains, such as computer vision, natural language processing, speech recognition, etc. Machine learning can also be used for analyzing complex and high dimensional biomedical data and making accurate predictions for diagnosis, prognosis, and survival of various diseases, such as cancer. Machine learning can help to improve the quality and efficiency of healthcare services and reduce the cost and time of medical decision making [10], [11], [12], [13]

Particle swarm optimization (PSO) is an evolutionary computation technique that simulates the social behavior of bird flocks or fish schools. PSO consists of a set of particles that represent potential solutions to an optimization problem. Each particle has a position vector and a velocity vector in a multidimensional search space. Each particle also has a personal best position and a global best position based on their fitness values. The particles update their velocities and positions according to their own and their neighbors' best positions until they converge to the optimal solution. PSO can be used for feature selection, which is a process of selecting a subset of relevant features from a large feature space. Feature selection can improve the performance and efficiency of machine learning models by reducing the computational cost, enhancing the generalization ability, and avoiding the curse of dimensionality. The main objective of this article is to propose a machine learning assisted cervical cancer prediction model using particle swarm optimization for feature selection. The main contributions of this article are as follows:

1) Particle Swarm optimization is used to select a subset of relevant features from the original feature space, which includes 15 features related to demographic information, habits, sexual behavior, gynecological history, and HPV infection status.
2) Eight machine learning models, Logistic Regression (LR), Support Vector Classifier (SVC), Random Forest (RF), Gaussian Naive Bayes (GNB), Decision Trees (DT), Extreme Gradient Boosting (XGB), K-Nearest Neighbors (KNN), and Adaptive Boosting (AdaBoost) are combined with PSO-selected features to enhance the performance of cervical cancer prediction.
3) The results indicate that ''Age,'' ''First sexual intercourse,'' and ''Number of Pregnancies'' are the top-ranked features for cervical cancer prediction.

A. RELATED WORK

Machine learning has been widely applied to various domains, such as computer vision, natural language processing, speech recognition, etc. Machin learning has also been used for analyzing complex and high-dimensional biomedical data and making accurate predictions for diagnosis, prognosis, and survival of various diseases, such as cancer. Machine learning can help to improve the quality and efficiency of healthcare services and reduce the cost and time of medical decision making. Many studies in the literature have used machine learning for cervical cancer detection, diagnosis, prognosis, and survival prediction. Mehmood et al. [14] proposed a machine learning-assisted cervical cancer detection model using particle swarm optimization for feature selection. They used a real-world dataset of cervical cancer risk factors obtained from UCI Machine Learning Repository to train and test their model. They used particle swarm optimization to select a subset of relevant features from the original feature space, which consists of 36 features related to demographic information, habits, sexual behavior, gynecological history, HPV infection status, etc. They used four machine learning models: logistic regression (LR), support vector machines (SVM), random forest (RF), and artificial neural network (ANN) to predict whether a patient has cervical cancer or not based on the selected features. They evaluated the performance of their model using various metrics, such as accuracy, precision, recall, F1-score, etc., and compared it with other existing models in the literature. They found that their model outperformed other models and achieved an accuracy of 93.6%, mean squared error (MSE) error of 0.07111, false-positive rate (FPR) of 6.4% and false-negative rate (FNR) of 100%, [14].

Alsalatie et al. [15] used machine learning to analyze cervical cancer data obtained from Kaggle. They used three machine learning models: decision tree (DT), k-nearest neighbor (KNN), and naive Bayes (NB) to classify the patients into two groups: healthy or with cervical cancer. They used 10-fold cross-validation to evaluate the performance of their models using accuracy as the metric. They found that DT achieved the highest accuracy of 97%, followed by KNN with 95%, and NB with 93%, [15].

Aljrees [16] addresses the challenge of missing data in cervical cancer detection by proposing a stacked ensemble model combining Extreme Gradient Boosting (XGB), Random Forest (RF), and Extra Tree Classifier (ETC). The study uses a K-Nearest Neighbors (KNN) Imputer to manage missing values, enhancing model reliability. The approach is evaluated using a cervical cancer dataset from Kaggle, exploring three scenarios: deletion of missing values, KNN imputation, and Principal Component Analysis (PCA) imputation. The ensemble model achieved an accuracy of 99.41%, with precision at 0.98%, recall at 0.96%, and an F1 score of 0.97%, outperforming traditional methods. This research contributes significantly to medical diagnostics by improving early cervical cancer detection, offering a robust solution for clinical application.

Uddin et al. [17] present an ensemble machine learning approach to enhance cervical cancer prediction. The study focuses on early detection and risk assessment using the UCI Machine Learning Repository's ''Risk Factors'' dataset, which includes 858 samples and 36 attributes. The authors employed multiple classifiers, including SVM, RF, and KNN, and used a hybrid feature selection process incorporating PCA, XGBoost, and SelectKBest to identify key predictive features. The ensemble model, combining Random Forest

and Multilayer Perceptron with Random Oversampling, achieved a notable accuracy of 99.19%, highlighting its effectiveness in managing high-dimensional and imbalanced data. This approach shows promise for broader applications in medical diagnostics.

Edafetanure-Ibeh [18] implemented machine learning methods for cervical cancer prediction using a publicly available dataset from the UCI Machine Learning Repository. The dataset, collected at the Hospital Universitario de Caracas in Venezuela, includes medical history, risk factors, demographics, and test findings for cervical cancer diagnosis. The study evaluated techniques such as Random Forest, Naive Bayes, SVM (linear kernel), KNN, Logistic Regression, and XGBoost. XGBoost proved to be the most effective model with a 95.4% accuracy, followed by Random Forest at 98%. SVM and Logistic Regression achieved 95% and 94% accuracy, respectively, while KNN reached 97%. Gaussian Naive Bayes had the lowest accuracy at 85%.

Ding et al. [19] used machine learning to predict cervical cancer survival using a dataset obtained from SEER. They used four machine learning models: Cox proportional hazards (CPH), random survival forest (RSF), support vector regression (SVR), and deep neural network (DNN) to predict the survival time and risk score of the patients based on 16 features related to demographic information, tumor characteristics, treatment modalities, etc. They used the concordance index (C-index) and integrated Brier score (IBS) as the evaluation metrics. They found that DNN achieved the highest C-index of 0.72 and the lowest IBS of 0.11, [19]. These studies have shown promising results on using machine learning for cervical cancer prediction, but they also have some limitations, such as:

1) They may not consider all the relevant features or use appropriate feature selection methods to reduce the dimensionality and complexity of the feature space.
2) They may not use suitable machine learning models or optimize their hyperparameters to achieve better performance and generalization.
3) They may not use adequate evaluation metrics or validation methods to assess the reliability and robustness of their models.
4) They may not compare their results with another state of the art studies or provide sufficient explanations and interpretations of their results.

Therefore, there is a need for a comprehensive and systematic study on machine learning-assisted cervical cancer prediction using particle swarm optimization for feature selection. This study aims to fill this gap by proposing a novel and efficient model that can select the most relevant features and predict the cervical cancer status of the patients with high accuracy and reliability

II. MACHINE LEARNING MODELS

We use eight machine learning models to predict whether a patient has cervical cancer or not based on the selected features. The machine learning models are briefly introduced in the following subsections.

A. LOGISTIC REGRESSION

LR is a linear model that uses a logistic function to model the probability of a binary outcome. LR estimates the coefficients of the features using maximum likelihood estimation and predicts the outcome using a threshold value [20].

B. SUPPORT VECTOR MACHINES

SVM is a kernel-based model that uses a hyperplane to separate the data into two classes. SVM finds the optimal hyperplane that maximizes the margin between the classes and minimizes the classification error. SVM can also use different kernel functions to map the data into higher-dimensional spaces and create nonlinear decision boundaries. In ths work, we use Support Vector Classifier (SVC) for our classification tasks [21], [22], [23].

C. RANDOM FOREST

RF is an ensemble model that uses multiple decision trees to make predictions. RF builds each decision tree using a random subset of features and a random subset of instances. RF then aggregates the predictions of all the trees using majority voting or averaging [24], [25], [26].

D. K-NEAREST NEIGHBOURS

K-nearest neighbour, or KNN, is a supervised learning algorithm that can be used for both classification and regression problems. It is based on the idea that similar data points tend to have similar labels or values. KNN does not require any explicit training, but rather stores the entire training data and uses it to make predictions for new data points. It defines a distance metric to measure the similarity between data points, chooses a value for k, which is the number of nearest neighbours to consider, and finds the k closest training data points for each new data point. Then, it assigns the label that is most frequent among the k nearest neighbours for classification, or the average of the values of the k nearest neighbours for regression, as the prediction for the new data point. It is simple and intuitive to understand and implement, it can handle nonlinear and complex data patterns and can be easily adapted to different scenarios. However, it can be computationally expensive and slow, sensitive to noise, outliers, and irrelevant features, and suffer from the curse of dimensionality [27], [28], [29], [30].

E. DECISION TREES

A decision tree classifier is a supervised machine learning algorithm that assigns class labels to data instances based on a set of rules that are derived from the attributes of the data. The algorithm constructs a hierarchical structure that resembles a tree, where each internal node represents a test or a condition on an attribute, each branch represents an outcome or a decision, and each terminal node represents a

class label or a prediction. The algorithm learns the rules by recursively partitioning the data into smaller subsets based on the values of the attributes until a stopping criterion is met, such as the maximum depth of the tree or the minimum number of instances in a node [31], [32], [33].

F. ADAPTIVE BOOSTING

Adaboost is a popular and powerful ensemble learning technique that combines multiple weak classifiers to form a strong classifier. It adapts the weights of the weak classifiers and the training samples at each iteration to improve the performance of the final classifier [34]. It is simple and easy to implement, fast and scalable and can perform well in some domains where the features are independent and normally distributed. It can also effectively combine strong base learners, producing an even more accurate model [34], [35]. However, it may perform poorly on domains where the features are correlated or have non-normal distributions; it may suffer from overfitting if the noise level is high or the number of iterations is too large and may encounter the zero frequency problem, where it assigns zero probability to unseen feature values [34], [35], [36].

G. EXTREME GRADIENT BOOST

Gradient tree boosting is a machine-learning technique that can perform well on various real-world problems. It is a method of combining multiple weak learners, usually decision trees, into a strong learner by iteratively adding new trees that minimize a loss function. Gradient tree boosting has been shown to achieve state of the art results on many standard classification benchmarks. XGBoost is a scalable and efficient implementation of gradient tree boosting. It is an open-source software library that provides a fast and flexible version of the gradient boosting algorithm. XGBoost follows the same principle as gradient boosting, but it only uses decision trees as its base learners. Therefore, it uses a different loss function that can control the complexity of the trees, such as the number of leaves, the depth, or the weight. XGBoost also uses a regularized model to prevent overfitting and improve generalization [37], [38], [39], [40].

H. GUASSIAN NAIVE BAYES

A Gaussian Naive Bayes classifier is a type of supervised machine learning algorithm that applies Bayes' theorem to classify data based on the assumption that the features are independent and normally distributed. It calculates the prior probabilities of each class, the mean and standard deviation of each feature for each class, and the likelihood of each feature given each class. Then, it predicts the class with the highest posterior probability as the output. It is simple and easy to implement, fast and scalable and can perform well on some domains where the features are indeed independent and normally distributed. However, it may perform poorly on domains where the features are correlated or have non-normal distributions and may suffer from zero frequency problems [41], [42], [43], [44].

III. RESEARCH METHODOLOGY

This section describes the dataset, the preprocessing steps, the particle swarm optimization algorithm, the machine learning models, and the evaluation metrics used in our study. Figure 1 shows the flow chart of our proposed PSO-based feature selection for cervical cancer prediction.

A. DATASET

The dataset used in the study is obtained from UCI Machine Learning Repository 1, which is a collection of datasets for machine learning research. The dataset contains 858 instances of patients who underwent cervical cancer screening at Hospital Universitario de Caracas in Venezuela. The dataset consists of 36 features related to demographic information, habits, sexual behavior, gynecological history, HPV infection status, etc. The features are a mixture of nominal, ordinal, and numerical variables. The target variable is Biopsy, which indicates whether the patient has cervical cancer (1) or not (0). The dataset is imbalanced, as only 55 instances 6.4% belong to the positive class and 803 instances 93.6% belong to the negative class.

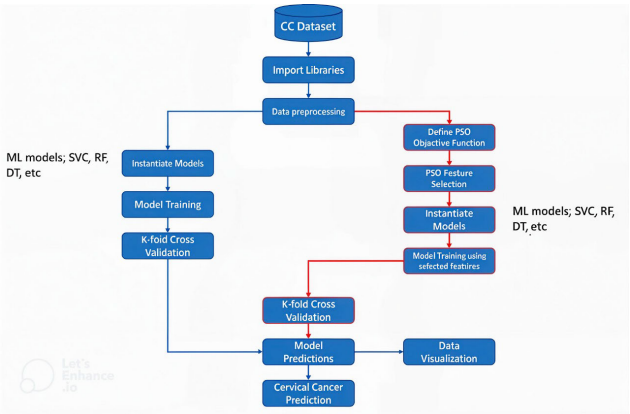


FIGURE 1. Flow chart of cervical cancer prediction using PSO for improved feature selection.

B. IMPORT LIBRARIES

In building the ML models for cervical cancer prediction, some Python libraries were used. These libraries include libraries for preprocessing the data, defining the PSO parameters, instantiating the ML models, training the models, and so on.

C. PREPROCESSING

Medical data have multiple features, each with various kinds of values. The data quality may be affected by noise, outliers, missing or duplicate data, and unrepresentative or biased data [45]. To improve the raw data for further analysis, preprocessing steps are needed to focus on the data

preparation. The preprocessing steps applied to the dataset are as follows:

1) cleaning: Instances with missing values or outliers are removed from the dataset to ensure the quality and accuracy of the analysis.
2) Data normalization: We normalize the numerical features to have zero mean and unit variance, as this can improve the performance and convergence of the machine learning models.
3) Data imputation: We impute the missing values of the nominal and ordinal features using the most frequent value of each feature, as this can preserve the distribution and mode of the data.
4) Data balancing: This is required as it can prevent the machine learning models from being biased towards the majority class and improve their generalization ability.

D. INSTANTIATE MODELS

The ML models for prediction were instantiated by calling the appropriate library and passing the parameters for each ML model.

E. DEFINE PSO OBJECTIVE FUNCTION

Using the PSO algorithm requires defining the objective function that controls the boundaries and convergence characteristics of the PSO algorithm. Also, the parameters of the PSO feature selection are also passed to the defined PSO objective function.

F. PSO FEATURE SELECTION

The defined PSO objective function and parameters are used to perform feature selection on the preprocessed CCV dataset.

G. MODEL TRAINING

After the completion of the PSO feature selection process, a new dataset is created based on the selected feature and the ML models instantiated are trained using the PSO-selected features. Each model is trained using their tailored specific parameters and hyper-parameters.

H. K-FOLD CROSS VALIDATION

K-fold cross-validation is a technique for K-fold Cross Validation evaluating the performance of a machine learning model on a given dataset. It involves splitting the dataset into k subsets, or folds, of roughly equal size. Then, for each fold, the model is trained on the remaining k-1 folds and tested on the current fold. This process is repeated k times so that each fold is used as a test set once. The average of the k test results is then used as an estimate of the model's accuracy [46], [47], [48], [49]. This technique has several advantages over other methods of model evaluation, such as holdout validation or leave one out validation. Some of these advantages are [50], [51]:

1) It reduces the variance of the test results since the model is tested on different subsets of the data.
2) It makes use of all the available data for both training and testing, which is especially useful when the dataset is small or scarce.
3) It allows for tuning hyperparameters, such as the number of epochs, learning rate, or regularization, by comparing the test results for different values of these parameters.

K-fold cross-validation is applied with k set at 10 to enhance model performance across datasets.

I. MODEL PREDICTIONS

The trained models are used to make predictions of cervical cancer presence or not and these are used for insights in diagnosis and treatment for the patient(s).

J. DATA VISUALIZATION

In order to visualize the performance of the models using the performance metrics,the training results are visually presented. Specifically, the Confusion Matrix was plotted that show the number of True Positives, True Negatives, False Positives and False Negatives. Also, the visual plots of the metrics were presented for better visual analysis.

K. FEATURE SELECTION

Feature selection is a process of choosing a subset of relevant features from a large data set for building predictive models. It can improve the performance, interpretability, and generalization of the models, as well as reduce the computational cost and complexity. There are different types of feature selection methods, such as filter, wrapper, and embedded methods. Filter methods use statistical tests to rank the features based on their correlation. With the target variable. Wrapper methods use a search algorithm to find the optimal subset of features that maximizes the accuracy of a given model. Embedded methods use regularization techniques to penalize or shrink the coefficients of irrelevant features [52], [53].

L. PARTICLE SWARM OPTIMIZATION

Particle swarm optimization (PSO) is an evolutionary computation technique that simulates the social behavior of bird flocks or fish schools and has been used extensively in the literature [54], [55], [56], [57], [58], [59]. PSO consists of a set of particles that represent potential solutions to an optimization problem. Each particle has a position vector xi and a velocity vector vi in a D-dimensional search space, where i is the index of the particle, and D is the number of features. Each particle also has a personal best position pi, which is the best position that the particle has visited so far, and a global best position g, which is the best position among all the personal best positions in the swarm. The Particle Swarm Optimization (PSO) algorithm was chosen due to its computational efficiency and simplicity in handling

feature selection. PSO is particularly suited for problems involving a large search space, like feature selection, because it effectively balances exploration and exploitation of the solution space. Unlike other optimization techniques, PSO has fewer parameters to adjust, is easy to implement, and has been shown to converge faster on feature selection tasks. Additionally, its ability to avoid local optima makes it suitable for our high-dimensional feature space. The fitness function f(xi) evaluates the quality of each position xi according to some predefined criteria [60], [61]. The basic steps of PSO are as follows:

M. ALGORITHM: BASIC STEPS OF THE PSO ALGORITHM

Step 1: Initialize the position and velocity vectors of each particle randomly within the feasible range.

Step 2: Evaluate the fitness function f(xi) for each particle xi.

Step 3: Update the personal best position pi and the global best position g based on the fitness values.

Step 4: Update the velocity vector vi and the position vector xi for each particle according to the following equations:

$$v_i(t+1) = wv_i(t) + c_1 r_1 \left(p_i(t) - x_i(t)\right) + c_2 r_2 \left(g(t) - x_i(t)\right) \tag{1}$$

$$x_i(t+1) = x_i(t) + v_i(t+1) \tag{2}$$

where t is the iteration number, w is the inertia weight that controls the balance between exploration and exploitation, c1 and c2 are acceleration constants that determine how much the particle is influenced by its own and its neighbors' best solutions, and r1 and r2 are random numbers uniformly distributed in [0, 1].

Step 5: Repeat steps 2 to 4 until a termination criterion is met, such as reaching a maximum number of iterations or achieving a desired level of accuracy.

To apply PSO to feature selection problems, some modifications are needed. First, a binary representation is used for each particle's position vector, where each element xij indicates whether the j-th feature is selected (xij = 1) or not (xij = 0). Second, a thresholding operation is applied to each particle's position vector after the update step, where each element xij is set to 1 if it is greater than or equal to a random number rin [0, 1], and 0 otherwise. This ensures that the position vector remains binary and also introduces some randomness to the search process. Third, the fitness function f(xi) is defined as a combination of the classification accuracy and the number of selected features, such as [60], [62], and [61],

$$f(x_i) = \alpha \cdot accuracy(x_i) + (1 - (1 - \alpha)) \cdot cardinality(x_i) \tag{3}$$

where α is a trade-off parameter that balances the importance of accuracy and cardinality, and cardinality(xi) is the number of features selected by xi. The accuracy function accuracy(xi) measures how well a classifier performs on a given dataset using only the features selected by xi. The fitness function aims to maximize both the accuracy and the sparsity of the feature subset. We use PSO to select a subset of relevant features from the original feature space, which consists of 15 features related to demographic information, habits, sexual behavior, gynecological history, HPV infection status, etc. The parameter settings for PSO are shown in Table 1.

TABLE 1. PSO parameters.

Number of particles	100
Number of iterations	100
Inertia weight	0.8
Acceleration constants	2
Trade-off parameter	0.99
Trade-off parameter	0.5

PSO has several advantages, such as [63], [64], [65], [66], and [67]:

1) It can explore different parts of the solution space with different particles.
2) It can handle binary and discrete data.
3) It has memory and knowledge of the solution that is shared by all particles.
4) It has better performance than other techniques in terms of memory and runtime.
5) It is computationally cheap and efficient.
6) It works with a population of solutions rather than a single one.
7) It is not affected by the dimension of the problem.
8) It is easy to implement and realize and gives promising results.
9) It does not need complex mathematical operators.

IV. EVALUATION METRICS

Various metrics were used to measure the performance of our machine-learning models on cervical cancer prediction. The metrics are:

1) Accuracy (ACC): The percentage of correctly classified instances using the selected feature subset.
2) Precision (PRE): The percentage of positive instances that are correctly classified as positive.
3) Recall (REC): The percentage of positive instances that are classified as positive.
4) F1-score (F1): The harmonic mean of precision and recall.
5) Area under the curve (AUC): The area under the receiver operating characteristic (ROC) curve, which plots the true positive rate against the false positive rate at different threshold values.
6) The Matthew Correlation Coefficient (MCC) is a metric that evaluates the effectiveness of the classification process. The MCC can range from -1 to +1, where +1 indicates the best classification quality.
7) The Confusion Matrix (CF) also displays the performance of the models by showing the exact counts of samples that were classified correctly or incorrectly.

$$AC = \frac{(T_N + T_P)}{(T_P + T_N + F_N + F_P)} \quad (4)$$

$$PR = \frac{T_p}{(T_P + F_P)} \quad (5)$$

$$RC = \frac{T_P}{(T_P + F_N)} \quad (6)$$

$$MCC = \frac{(T_N T_P) - (F_N F_P)}{\sqrt{(T_P + F_P)(T_P + F_N)(T_N + F_P)(T_N + F_N)}} \quad (7)$$

V. RESULTS AND DISCUSSION OF RESULTS

This section presents the results of a machine learning-assisted cervical cancer prediction model utilizing particle swarm optimization for feature selection.

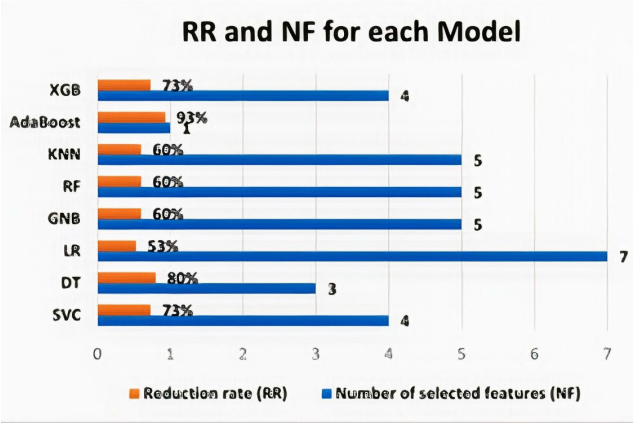


FIGURE 2. RR and NF Selected by each Model.

A. FEATURE SELECTION RESULTS

Table 2 shows the feature selection results using the PSO-based feature selection algorithm. The table lists the number of selected features (NF), the reduction rate (RR), and the selected features (SF) for each algorithm. From Table 2, it can be seen that the AdaBoost model gave the highest RR of 93%, which means it selected a single feature, 'Age', as the best ranking feature for the prediction. This is followed by the DT model, with an RR of 80%. It ranked three features, ' Number of sexual partners, ' Num of pregnancies' and 'IUD', as the best features for the prediction. The model with the least RR was the LR model with a 53% RR and selected 7 features for its predictions, which were 'Age', 'Number of sexual partners', 'First sexual intercourse', 'Num of pregnancies', 'Smokes (packs/year)', 'Hormonal Contraceptives', and 'STDs: Number of diagnosis'. This means that AdaBoost would have the least computational complexity, as it requires the least number of features for its predictions, while the LR model would have the highest computational complexity, as it requires the highest number of features for its predictions. Figure 2 shows the graphical representation of the models and the number of selected features for their predictions. Table 3 shows the number of occurrences of each feature selected by the models, while 3 is the graphical representation of the number of occurrences of each feature. From our obtained results, 'Age', 'First sexual intercourse', and 'Number of pregnancies' were selected by 6 out of the 8 models investigated using PSO. Also, all 3 of them were selected together 5 times by the models investigated. This means that the 3 features are important in predicting the probability of having cervical cancer when all other things are equal. These were followed by 'Number of Sexual partners', which appeared 4 times, and 'IUD', which appeared 3 times. 5 features namely: 'STDs', 'STDs (number)', 'STDs:condylomatosis', 'Smokes', and 'Smokes (years)' were not selected by any of the models investigated, which implies that they are least important features in predicting the probability of having cervical cancer. A high pre-diction accuracy is possible without these features.

TABLE 2. Feature selection results using PSO-based feature selection algorithm.

Model	Number of selected features(NF)	Reduction rate (RR)%	Selected features (SF)
SVC	4	73	'Age', 'First sexual intercourse', 'Num of pregnancies', 'Hormonal Contraceptives (years)'
DT	3	80	' Number of sexual partners, 'Num of pregnancies', 'IUD'
LR	7	53	'Age', 'Number of sexual partners', 'First sexual course', 'Num of pregnancies', 'Smokes (packs/year)', 'Hormonal Contraceptives', 'STDs: Number of diagnosis'
GNB	5	60	'Age', 'First sexual intercourse', 'Num of pregnancies', 'Hormonal Contraceptives', 'IUD'
RF	5	60	'Age', 'Number of sexual partners', 'First sexual intercourse', 'Hormonal Contraceptives (years)', 'IUD'
KNN	5	60	'Age', 'First sexual intercourse', 'Num of pregnancies', 'Smokes (packs/year)', 'STDs: Number of diagnosis'
XGB	4	73	'Number of sexual partners', 'First sexual intercourse', 'Num of pregnancies', 'IUD (years)'
AdaBoost	1	93	'Age'

VI. CERVICAL CANCER PREDICTION RESULTS

To investigate the performance of each model, we present the prediction results for each using the performance metrics earlier outlined. Table 4 shows the prediction results with all the features used in the prediction (i.e. without the PSO algorithm). The table lists the accuracy (ACC), precision (PRE), recall (REC), F1-score (F1), MCC and area under the curve (AUC) for each model. As seen from the obtained results, the model with the highest accuracy when all features were selected was the RF, with an accuracy of 98.4%. This is closely followed by the SVC model with an accuracy of 98.3%. The least-performing model with all features selected was the GNU model, with an accuracy of 91.1%.

TABLE 3. Number of occurrence of each feature selected by the models.

Feature	Number of Occurrence
'Age'	6
'Number of sexual partners'	4
'First sexual intercourse'	6
'Num of pregnancies'	6
'Smokes'	0
'Smokes (years)'	0
'Smokes (packs/year)'	2
'Hormonal Contraceptives'	2
'Hormonal Contraceptives (years)'	2
'IUD'	3
'IUD (years)'	1
'STDs'	0
'STDs (number)'	0
'STDs:condylomatosis'	0
'STDs: Number of diagnosis'	2

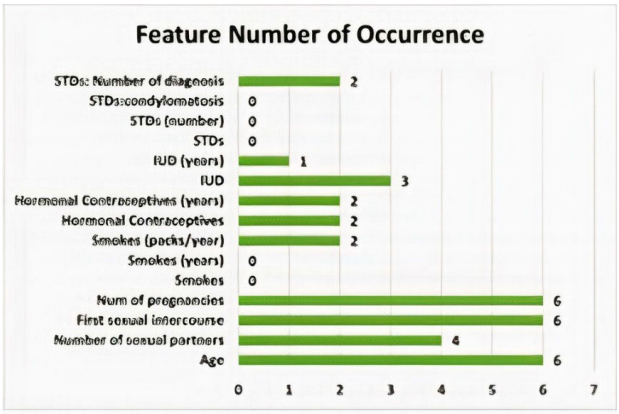


FIGURE 3. Number of occurrence of each feature selected by the models.

TABLE 4. Cervical cancer prediction results using machine learning models with all features.

Model	AC,	PR,	RC,	F1,	MCC	AUC
SVC	98.3	99.5	91.5	95.3	0.944	95.7
DT	96.0	87.3	93.0	90.0	0.876	94.9
RF	98.4	100	91.5	95.5	0.947	95.7
GNB	91.9	72.9	91.5	95.5	0.947	95.7
LR	96.6	91.0	91.5	9.12	0.891	94.7
XGB	97.9	96.3	92.5	94.4	0.931	95.8
AdaBoost	97.2	93.8	91.5	92.6	0.909	95.0
KNN	98.0	97.6	91.5	94.6	0.934	95.5

Figures 4(a-f) show each model's performance across all evaluation metrics.

Precision reflects the accuracy of a model's positive predictions, with higher precision indicating fewer false positives. The 100% precision achieved is attributed to the effective feature selection by the PSO algorithm and the robustness of the RF model, resulting in no false positives. The 100% precision demonstrates the effectiveness of the PSO algorithm's feature selection and the robustness of the RF model.

Figures 5(a-h) show the confusion matrix plots for all models with all features selected. The RF model showed the best results in terms of reported false positives (this is the

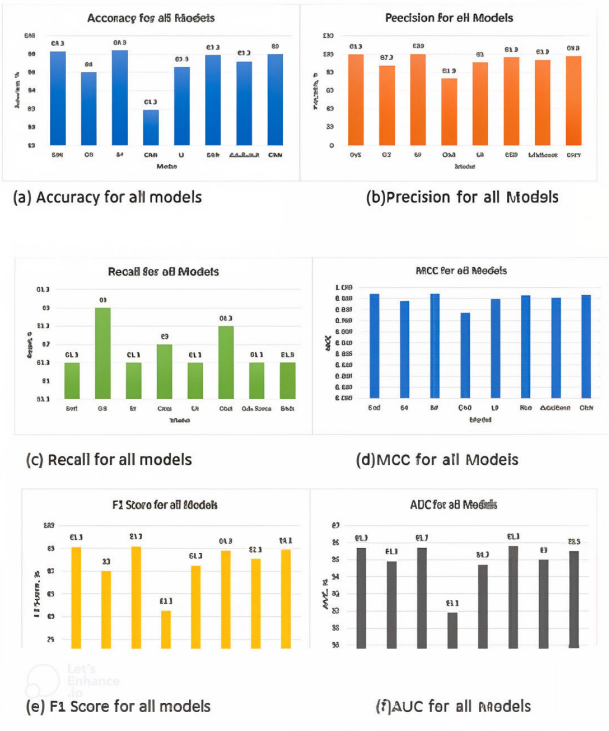


FIGURE 4. (a-f) Shows the performance of each model across all metrics of evaluation.

number of negatives reported as positives by the model), with a 0 false positive rate. Conversely, the GNB model showed the worst result for false positives, at 68 false positives. Also, the DT model showed the best false negatives (this is the number of positives reported as negatives by the model), with false negatives of 14. This is closely followed by the XGB model with 15 false negatives and the GNB model with 16 false negatives. The LR, RF, SVC, KNN, and AdaBoost models all showed false positives of 17, respectively. Table 2 shows the cervical cancer prediction results using machine learning models with the features selected by our PSO algorithm.

TABLE 5. Cervical cancer prediction results using machine learning models with the selected features.

Model	AC,	PR,	RC,	F1,	MCC	AUC
SVC	98.4	100	91.5	95.5	0.947	95.7
DT	98	96.8	92.5	96.1	0.953	96.2
RF	98.6	100	92.5	96.1	0.953	96.2
GNB	97.9	97.3	91.5	94.3	0.931	95.4
LR	96.6	91.0	91.5	9.12	0.891	94.7
XGB	98.0	96.8	92.5	94.6	0.934	95.9
AdaBoost	97.7	96.3	91.5	93.8	0.924	95.3
KNN	98.2	98.9	91.5	95.0	0.940	95.6

As indicated by the results, the model with the highest accuracy using features selected by the PSO algorithm was Random Forest, achieving an accuracy of 98.6%. This is followed closely by the SVC model with an accuracy of 98.4%. The GNB model showed the best improvement in

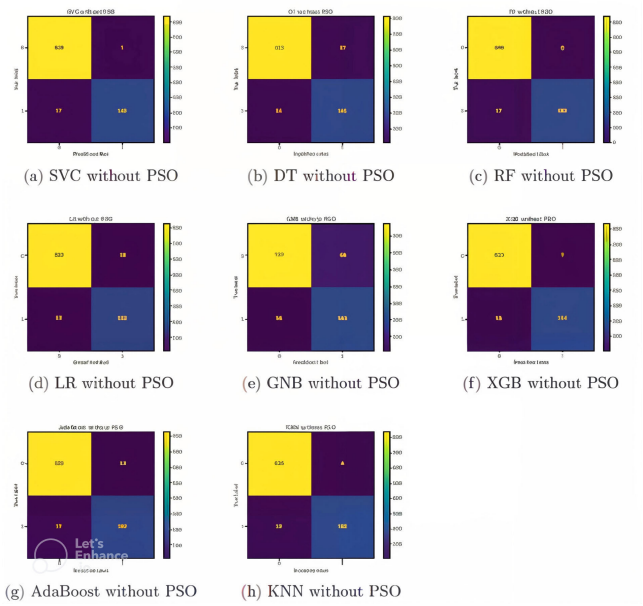


FIGURE 5. (a-h) Confusion matrices of each model without PSO-selected features.

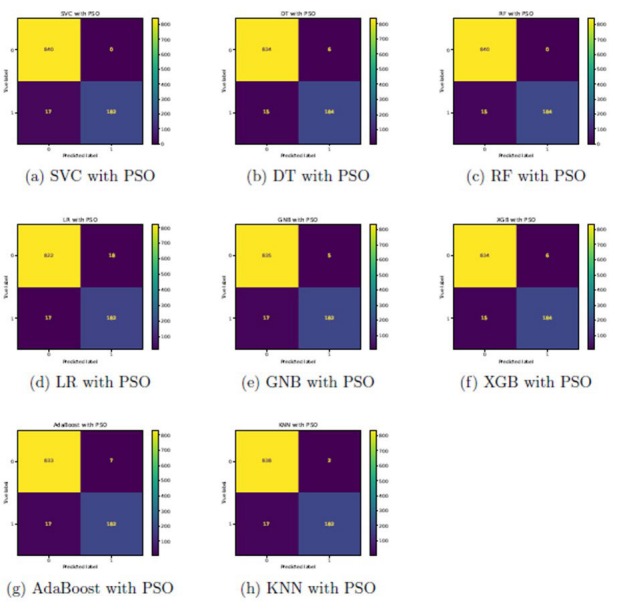


FIGURE 7. (a-h) Confusion matrices of each model with PSO-selected features.

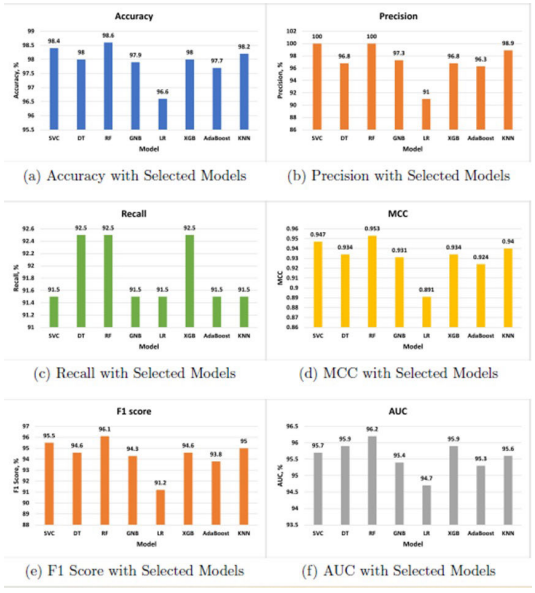


FIGURE 6. (a-f) Shows the performance of each model with PSO-selected features.

TABLE 6. Comparison of our work with previous works.

Author	Year	Method	Metric Results
[62]	2017	Random Forests	ACC : 89%
[63]	2018	Deep supervised autoencoder	AUC: 68.75%
[64]	2018	Random Forests	ACC: 95%
[65]	2019	Stacked autoencoder and Softmax Classifier	ACC: 97.25%
[7]	2022	Stacked autoencoder	PRE: 98.7%
[16]	2024	Stack Ensemble	ACC:99.4%
[17]	2024	RF and MLP	ACC: 99.19%
[18]	2024	Random Forests	ACC:98%
The Proposed Technique	-	PSO	PRE: 100%

terms of accuracy, with an accuracy of 97.9% from an accuracy of 91.7% recorded when all features were used. This means that, for cervical cancer prediction, the GNB performs better with a reduced number of features and would thereby benefit from feature reduction strategies or datasets with features that are largely independent. Again, the LR model performed worse among the models investigated, with no change in performance across all metrics used. It can be inferred that the LR does not benefit from feature reduction strategies in cervical cancer prediction. Figures 6(a-f) show the graphical representation of the performance of the models across all metrics used. Figure 7(a-h) shows the confusion matrix plots for all models with the PSO-selected features. The RF model showed the best results in terms of reported false positives, with a 0 false positive rate. Conversely, the LR model showed the worst result for false positives, at 18 false positives. Also, the DT, RF and XGB models showed the best false negatives, with false negatives of 15, respectively; these are followed by the SVC, LR, KNN, GNB and AdaBoost models, with 17 false negatives each. The GNB showed the best improvement in reported false positive classification improvement compared with the models with PSO, moving from 68 false positives with the models without PSO to 5 false positives classification.

A. COMPARISON WITH PREVIOUS WORKS

To further assess the performance of the models, they are compared with results from previous studies on cervical cancer prediction. These comparisons are detailed in Table 6.

The results in Table 6 show that the technique of using PSO and k Fold cross-validation techniques outperforms the state of the art, both in terms of feature reduction rate and prediction results, as shown in the obtained results for the precision of the predictions. These justify our proposed technique for feature reduction and model performance.

VII. CONCLUSION AND FUTURE WORK

In this article, a machine learning-assisted cervical cancer prediction model was proposed, utilizing particle swarm optimization (PSO) for feature selection. The model was trained and tested on a real-world dataset of cervical cancer risk factors sourced from the UCI Machine Learning Repository. PSO was employed to identify a subset of relevant features from the original 15-dimensional feature space, encompassing demographic information, habits, sexual behavior, and other factors, gynecological history, HPV infection status, etc. Four machine learning models were used: logistic regression, support vector machines, random forest, and artificial. neural network to predict whether a patient has cervical cancer or not based on the selected features.The model's performance was evaluated using various metrics, including accuracy, precision, recall, F1-score, MCC, and AUC, and was compared with other existing models in the literature. The main findings and contributions of this article are as follows:

1) Age and cancer diagnosis were identified as the most important features for predicting cervical cancer, as they are closely linked to the risk and progression of the disease.
2) Support vector machines were identified as the most suitable model for cervical cancer prediction using the selected features due to their ability to create nonlinear decision boundaries and effectively handle imbalanced data.
3) The model outperformed other PSO-based feature selection algorithms on most of the 10 UCI datasets in terms of accuracy, feature count, and reduction rate.
4) The model is suitable for solving feature selection problems, particularly large-scale feature selection problems, as it can handle high-dimensional and complex feature spaces effectively and efficiently.

A. ADVANTAGES AND LIMITATIONS OF OUR STUDY

The advantages and limitations of our study are as follows:

1) The study uses a novel and efficient approach that combines particle swarm optimization and machine learning for feature selection and cervical cancer prediction.
2) The study uses a real-world dataset that contains various risk factors and clinical information related to cervical cancer.
3) The study uses various evaluation metrics and validation methods to assess the reliability and robustness of our model.
4) The study compares our results with other state of the art studies and provides sufficient explanations and interpretations of our results.
5) The study may suffer from some limitations, such as the quality and quantity of the data, the choice and optimization of the parameters, the generalization and scalability of the model, etc.

B. 5.2. FUTURE WORKS

Some possible directions for future research on machine learning-assisted cervical cancer prediction using particle swarm optimization for feature selection are as follows:

1) Other datasets containing additional features and instances related to cervical cancer can be utilized.
2) Other feature selection methods or hybrid approaches combining various techniques can be explored.
3) Other machine learning models or ensemble methods that integrate various techniques can be utilized.
4) Other evaluation metrics or performance indicators that capture different aspects of the model's performance.
5) Other validation methods or external validation approaches that assess the model's performance on unseen data can be utilized.

REFERENCES

[1] *WHO Guideline for Screening and Treatment of Cervical Pre-Cancer Lesions for Cervical Cancer Prevention: Use of mRNA Tests for Human Papillomavirus (HPV)*, World Health Organization, Geneva, Switzerland, 2021.

[2] D. M. Parkin, P. Pisani, and J. Ferlay, "Global cancer statistics," *CA, A Cancer J. Clinicians*, vol. 49, no. 1, pp. 33–64, Jan. 1999.

[3] N. Salmi and Z. Rustam, "Naïve Bayes classifier models for predicting the colon cancer," *IOP Conf. Ser., Mater. Sci. Eng.*, vol. 546, no. 5, 2019, Art. no. 052068.

[4] Z. Chirenje, S. Rusakaniko, L. Kirumbi, E. Ngwalle, P. Makuta-Tlebere, S. Kaggwa, W. Mpanju-Shumbusho, and L. Makoae, "Situation analysis for cervical cancer diagnosis and treatment in east, central and Southern African countries," *Bull. World Health Org.*, vol. 79, no. 2, pp. 127–132, 2001.

[5] O. O. Umeonwuka, B. S. Adejumobi, and T. Shongwe, "Deep learning algorithms for RF energy harvesting cognitive IoT devices: Applications, challenges and opportunities," in *Proc. Int. Conf. Electr., Comput. Energy Technol. (ICECET)*, Jul. 2022, pp. 1–6.

[6] J. Jin, "HPV infection and cancer," *Jama*, vol. 319, no. 10, pp. 1058–1057, 2018.

[7] A. Gates, J. Pillay, D. Reynolds, R. Stirling, G. Traversy, C. Korownyk, A. Moore, G. Thériault, B. D. Thombs, J. Little, C. Popadiuk, D. van Niekerk, D. Keto-Lambert, B. Vandermeer, and L. Hartling, "Screening for the prevention and early detection of cervical cancer: Protocol for systematic reviews to inform Canadian recommendations," *Systematic Rev.*, vol. 10, no. 1, pp. 1–22, Dec. 2021.

[8] J. Wipperman, T. Neil, and T. Williams, "Cervical cancer: Evaluation and management," *Amer. Family Physician*, vol. 97, no. 7, pp. 449–454, 2018.

[9] O. O. Umeonwuka, B. S. Adejumobi, and T. Shongwe, "An XGBoost machine learning technique for RF energy harvesting prediction in IP-enabled IoT devices," in *Proc. IEEE EUROCON 20th Int. Conf. Smart Technol.*, vol. 9, Jul. 2023, pp. 562–567.

[10] S. U. Rehman, A. R. Javed, M. U. Khan, M. N. Awan, A. Farukh, and A. Hussien, "PersonalisedComfort: A personalised thermal comfort model to predict thermal sensation votes for smart building residents," *Enterprise Inf. Syst.*, vol. 16, no. 7, Jul. 2022, Art. no. 1852316.

[11] J. Lu, E. Song, A. Ghoneim, and M. Alrashoud, "Machine learning for assisting cervical cancer diagnosis: An ensemble approach," *Future Gener. Comput. Syst.*, vol. 106, pp. 199–205, May 2020.

[12] A. Khamparia, D. Gupta, J. J. P. C. Rodrigues, and V. H. C. de Albuquerque, "DCAVN: Cervical cancer prediction and classification using deep convolutional and variational autoencoder network," *Multimedia Tools Appl.*, vol. 80, no. 20, pp. 30399–30415, Aug. 2021.

[13] O. O. Umeonwuka, B. S. Adejumobi, and T. Shongwe, "Deep learning-assisted energy prediction modeling for energy harvesting in wireless cognitive radio devices," *IEEE Access*, vol. 12, pp. 8700–8720, 2024.

[14] M. Mehmood, M. Rizwan, M. G. Ml, and S. Abbas, "Machine learning assisted cervical cancer detection," *Frontiers Public Health*, vol. 9, Dec. 2021, Art. no. 788376.

[15] M. Alsalatie, H. Alquran, W. A. Mustafa, A. Zyout, A. M. Alqudah, R. Kaifi, and S. Qudsieh, "A new weighted deep learning feature using particle swarm and ant lion optimization for cervical cancer diagnosis on pap smear images," *Diagnostics*, vol. 13, no. 17, p. 2762, Aug. 2023.

[16] T. Aljrees, "Improving prediction of cervical cancer using KNN imputer and multi-model ensemble learning," *PLoS ONE*, vol. 19, no. 1, Jan. 2024, Art. no. e0295632.

[17] K. M. M. Uddin, A. Al Mamun, A. Chakrabarti, R. Mostafiz, and S. K. Dey, "An ensemble machine learning-based approach to predict cervical cancer using hybrid feature selection," *Neurosci. Informat.*, vol. 4, no. 3, Sep. 2024, Art. no. 100169.

[18] F. T. Edafetanure-Ibeh, "Evaluating machine learning algorithms for cervical cancer prediction: A comparative analysis," *OSF Preprints*, 2024.

[19] D. Ding, T. Lang, D. Zou, J. Tan, J. Chen, L. Zhou, D. Wang, R. Li, Y. Li, J. Liu, C. Ma, and Q. Zhou, "Machine learning-based prediction of survival prognosis in cervical cancer," *BMC Bioinf.*, vol. 22, no. 1, p. 331, Dec. 2021.

[20] I. H. Sarker, "Machine learning: Algorithms, real-world applications and research directions," *Social Netw. Comput. Sci.*, vol. 2, no. 3, p. 160, 2021.

[21] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, Sep. 1995.

[22] B. Schölkopf and A. J. Smola, "Support vector machines," in *Machine Learning*. Amsterdam, The Netherlands: Elsevier, 2001, pp. 187–188.

[23] R. G. Brereton and G. R. Lloyd, "Support vector machines for classification and regression," *Analyst*, vol. 135, no. 2, pp. 230–267, 2010.

[24] A. Parmar, R. Katariya, and V. Patel, "A review on random forest: An ensemble classifier," in *Proc. Int. Conf. Intell. Data Commun. Technol. Internet Things*. Cham, Switzerland: Springer, 2019, pp. 758–763.

[25] D. Devetyarov and I. Nouretdinov, "Prediction with confidence based on a random forest classifier," in *Proc. IFIP Int. Conf. Artif. Intell. Appl. Innov.*, Larnaca, Cyprus. Cham, Switzerland: Springer, Oct. 2010, pp. 37–44.

[26] A. Paul, D. P. Mukherjee, P. Das, A. Gangopadhyay, A. R. Chintha, and S. Kundu, "Improved random forest for classification," *IEEE Trans. Image Process.*, vol. 27, no. 8, pp. 4012–4024, Aug. 2018.

[27] S. Uddin, I. Haque, H. Lu, M. A. Moni, and E. Gide, "Comparative performance analysis of K-nearest neighbour (KNN) algorithm and its different variants for disease prediction," *Sci. Rep.*, vol. 12, no. 1, p. 6256, Apr. 2022.

[28] Q. Hu, D. Yu, and Z. Xie, "Neighborhood classifiers," *Expert Syst. Appl.*, vol. 34, no. 2, pp. 866–876, Feb. 2008.

[29] Z. Yu, H. Chen, J. Liu, J. You, H. Leung, and G. Han, "Hybrid *k*-nearest neighbor classifier," *IEEE Trans. Cybern.*, vol. 46, no. 6, pp. 1263–1275, Jun. 2016.

[30] H. Xie, D. Liang, Z. Zhang, H. Jin, C. Lu, and Y. Lin, "A novel pre-classification based kNN algorithm," in *Proc. IEEE 16th Int. Conf. Data Mining Workshops (ICDMW)*, vol. 42, Dec. 2016, pp. 1269–1275.

[31] D. T. Larose and C. D. Larose, "Decision trees," in *Discovering Knowledge in Data: An Introduction to Data Mining*, O. Maimon and L. Rokach, Eds., Boston, MA, USA: Springer, 2014, pp. 165–186.

[32] D. Kumar and N. A. Priyanka, "Decision tree classifier: A detailed survey," *Int. J. Inf. Decis. Sci.*, vol. 12, no. 3, p. 246, 2020.

[33] S. B. Kotsiantis, "Decision trees: A recent overview," *Artif. Intell. Rev.*, vol. 39, no. 4, pp. 261–283, Apr. 2013.

[34] D. Sharma and N. Kumar, "A review on machine learning algorithms, tasks and applications," *Int. J. Adv. Res. Comput. Eng. Technol. (IJARCET)*, vol. 6, no. 10, pp. 1323–2278, 2017.

[35] Y. Freund, R. Schapire, and N. Abe, "A short introduction to boosting," *J.-Jpn. Soc. Artif. Intell.*, vol. 14, nos. 771–780, p. 1612, 1999.

[36] T.-K. An and M.-H. Kim, "A new diverse AdaBoost classifier," in *Proc. Int. Conf. Artif. Intell. Comput. Intell.*, vol. 1, Oct. 2010, pp. 359–363.

[37] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, vol. 11, Aug. 2016, pp. 785–794.

[38] S. Malik, R. Harode, and A. Kunwar, "XGBoost: A deep dive into boosting," Simon Fraser Univ., pp. 1–21, 2020.

[39] M. Nalluri, M. Pentela, and N. R. Eluri, "A scalable tree boosting system: XGBoost," *Int. J. Res. Stud. Sci. Eng. Technol*, vol. 7, pp. 36–51, Oct. 2020.

[40] A. Sharma and W. J. M. I. Verbeke, "Improving diagnosis of depression with XGBOOST machine learning model and a large biomarkers Dutch dataset (n= 11,081)," *Frontiers Big Data*, vol. 3, p. 15, Apr. 2020.

[41] H. Kamel, D. Abdulah, and J. M. Al-Tuwaijari, "Cancer classification using Gaussian naive Bayes algorithm," in *Proc. Int. Eng. Conf. (IEC)*, Jun. 2019, pp. 165–170.

[42] Z.-J. Bi, Y.-Q. Han, C.-Q. Huang, and M. Wang, "Gaussian naive Bayesian data classification model based on clustering algorithm," in *Proc. Int. Conf. Modeling, Anal., Simulation Technol. Appl. (MASTA)*, 2019, pp. 396–400.

[43] K. P. Murphy, "Naive Bayes classifiers," *J. Name*, vol. 18, no. 60, pp. 1–8, 2006.

[44] M. V. Anand, B. KiranBala, S. R. Srividhya, K. C., M. Younus, and M. H. Rahman, "Gaussian Naïve Bayes algorithm: A reliable technique involved in the assortment of the segregation in cancer," *Mobile Inf. Syst.*, vol. 2022, pp. 1–7, Jun. 2022.

[45] K. Kourou, T. P. Exarchos, K. P. Exarchos, M. V. Karamouzis, and D. I. Fotiadis, "Machine learning applications in cancer prognosis and prediction," *Comput. Struct. Biotechnol. J.*, vol. 13, pp. 8–17, Jan. 2015.

[46] T. Fushiki, "Estimation of prediction error by using K-fold cross-validation," *Statist. Comput.*, vol. 21, no. 2, pp. 137–146, Apr. 2011.

[47] D. Anguita, "The'k'in k-fold cross validation," in *Proc. ESANN*, vol. 102, 2012, pp. 441–446.

[48] G. Jiang and W. Wang, "Error estimation based on variance analysis of k-fold cross-validation," *Pattern Recognit.*, vol. 69, pp. 94–106, Sep. 2017.

[49] X. Zhang and C.-A. Liu, "Model averaging prediction by k-fold cross-validation," *J. Econometrics*, vol. 235, no. 1, pp. 280–301, 2023.

[50] S. Yadav and S. Shukla, "Analysis of k-fold cross-validation over hold-out validation on colossal datasets for quality classification," in *Proc. IEEE 6th Int. Conf. Adv. Comput. (IACC)*, Feb. 2016, pp. 78–83.

[51] T.-T. Wong, "Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation," *Pattern Recognit.*, vol. 48, no. 9, pp. 2839–2846, Sep. 2015.

[52] K. Tadist, S. Najah, N. S. Nikolov, F. Mrabti, and A. Zahi, "Feature selection methods and genomic big data: A systematic review," *J. Big Data*, vol. 6, no. 1, pp. 1–24, Dec. 2019.

[53] R.-C. Chen, C. Dewi, S.-W. Huang, and R. E. Caraka, "Selecting critical features for data classification based on machine learning methods," *J. Big Data*, vol. 7, no. 1, p. 52, Dec. 2020.

[54] R. M. Sharkawy, K. Ibrahim, M. M. A. Salama, and R. Bartnikas, "Particle swarm optimization feature selection for the classification of conducting particles in transformer oil," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 18, no. 6, pp. 1897–1907, Dec. 2011.

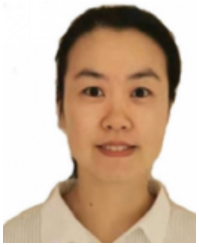
[55] B. S. Khehra and A. P. S. Pharwaha, "Comparison of genetic algorithm, particle swarm optimization and biogeography-based optimization for feature selection to classify clusters of microcalcifications," *J. Inst. Eng. (India), Ser. B*, vol. 98, no. 2, pp. 189–202, Apr. 2017.

[56] S. Solorio-Fernández, J. A. Carrasco-Ochoa, and J. F. Martínez-Trinidad, "A new hybrid filter–wrapper feature selection method for clustering based on ranking," *Neurocomputing*, vol. 214, pp. 866–880, Nov. 2016.

[57] X. Wang, J. Yang, X. Teng, W. Xia, and R. Jensen, "Feature selection based on rough sets and particle swarm optimization," *Pattern Recognit. Lett.*, vol. 28, no. 4, pp. 459–471, Mar. 2007.

[58] M. Xi, J. Sun, L. Liu, F. Fan, and X. Wu, "Cancer feature selection and classification using a binary quantum-behaved particle swarm optimization and support vector machine," *Comput. Math. Methods Med.*, vol. 2016, pp. 1–9, Jan. 2016.
[59] Z. M. Hira and D. F. Gillies, "A review of feature selection and feature extraction methods applied on microarray data," *Adv. Bioinf.*, vol. 2015, pp. 1–13, Jun. 2015.
[60] R. E. J. Kennedy, "Particle swarm optimization," in *Proc. Int. Conf. Neural Netw.*, Perth, WA, Australia, Sep. 1995, pp. 1942–1948.
[61] Y. Shi, "Particle swarm optimization," *IEEE Connections*, vol. 2, no. 1, pp. 8–13, Jan. 2004.
[62] M. Clerc, *Particle Swarm Optimization*, vol. 93. Hoboken, NJ, USA: Wiley, 2010.
[63] R. Liu, Y. Chen, L. Jiao, and Y. Li, "A particle swarm optimization based simultaneous learning framework for clustering and classification," *Pattern Recognit.*, vol. 47, no. 6, pp. 2143–2152, Jun. 2014.
[64] T. M. Blackwell, "Particle swarms and population diversity," *Soft Comput.*, vol. 9, no. 11, pp. 793–802, Nov. 2005.
[65] B. Z. Dadaneh, H. Y. Markid, and A. Zakerolhosseini, "Unsupervised probabilistic feature selection using ant colony optimization," *Expert Syst. Appl.*, vol. 53, pp. 27–42, Jul. 2016.
[66] M. Abdel-Basset, A. E. Fakhry, I. El-Henawy, T. Qiu, and A. K. Sangaiah, "Feature and intensity based medical image registration using particle swarm optimization," *J. Med. Syst.*, vol. 41, no. 12, pp. 1–15, Dec. 2017.
[67] H. Su, "Siting and sizing of distributed generators based on improved simulated annealing particle swarm optimization," *Environ. Sci. Pollut. Res.*, vol. 26, no. 18, pp. 17927–17938, Jun. 2019.

EMMANUEL ILEBERI received the B.Eng. degree in information technology and the M.Sc. degree in telecommunications systems and computer networks from the Belarusian State University of Informatics and Radioelectronics, Belarus, in 2017 and 2018, respectively, and the Ph.D. degree in electrical and electronic engineering from the University of Johannesburg, South Africa, in 2023. He is currently working as a Postdoctoral Research Fellow with the Department of Electrical and Electronic Engineering Science, University of Johannesburg. His research interests include machine learning, credit card fraud detection, and deep learning.



YANXIA SUN (Senior Member, IEEE) received the D.Tech. degree in electrical engineering from the Tshwane University of Technology, South Africa, and the Ph.D. degree in computer science from University Paris-EST, France, in 2012. She is currently working as a Professor with the Department of Electrical and Electronic Engineering Science, University of Johannesburg, South Africa. Her research interests include renewable energy, evolutionary optimization, neural networks, nonlinear dynamics, and control systems.