




# Multi-modal prediction of breast cancer using particle swarm optimization with non-dominating sorting

International Journal of Distributed  
Sensor Networks  
2020, Vol. 16(11)  
© The Author(s) 2020  
DOI: 10.1177/1550147720971505  
journals.sagepub.com/home/dsn  


Vijayalakshmi S<sup>1</sup>, John A<sup>2</sup>, Sunder R<sup>3</sup>, Senthilkumar Mohan<sup>4</sup> , Sweta  
Bhattacharya<sup>4</sup>, Rajesh Kaluri<sup>4</sup>, Guang Feng<sup>5</sup> and Usman Tariq<sup>6</sup> 

## Abstract

Cancer is enlisted as the second leading reason for death across the world wherein almost one person out of six dies of cancer. Breast cancer is one of the most common forms of cancer predominant in women having the second highest mortality rate in the world. Various scientific studies have been conducted to combat this disease, and machine learning approaches have been an extremely popular choice. Particle swarm optimization has been identified as one of the most powerful and efficient technique for the diagnosis of breast cancer guiding physicians towards timely and accurate treatment. It is also pertinent to mention that multi-modal prediction methods are used to make decisions depending upon different scenarios and aspects whereas the non-dominating sorting feature is useful to sort different objects based on differing requirements. The main novelty of this work is multi-modal prediction algorithm for breast cancer prediction is proposed. The work encompasses the use of particle swarm optimization, non-dominating sorting and multi-classifier techniques, namely, *k*-nearest neighbour method, fast decision tree and kernel density estimation. Finally, Bayes' theorem is implemented for revising the results to achieve optimum accuracy in the breast cancer prediction. The proposed particle swarm optimization and non-domination sorting with classifier technique model helps to select the most significant features relevant to breast cancer predictions. The selected features design the objective of the problem model. The proposed model is implemented on the WBCD and WDBC breast cancer data sets publicly available from the UCI machine learning data repository. The metrics considered are sensitivity, specificity, accuracy and time complexity. The experimental results of the study using measures such as sensitivity, specificity, accuracy and time complexity. The experimental results of the study are evaluated against the state-of-the-art algorithms, namely, genetic algorithm kernel density estimation and particle swarm optimization kernel density estimation wherein the results justify the superiority of the proposed model.

## Keywords

Swarm intelligence, breast cancer, feature selection, multi-classification, Bayes' theorem, particle swarm optimization

Date received: 4 May 2020; accepted: 2 October 2020

Handling Editor: Benny Lo

<sup>1</sup>School of Computer Science and Engineering, Rajiv Gandhi College of Engineering & Technology, Puducherry, India

<sup>2</sup>School of Computer Science and Engineering, Galgotias University, Greater Noida, India

<sup>3</sup>Department of Computer Science and Engineering, Sahrdaya College of Engineering and Technology, Thrissur, India

<sup>4</sup>School of Information Technology and Engineering, Vellore Institute of Technology, Vellore, India

<sup>5</sup>Center of Network Information & Modern Education Technology, Guangdong University of Technology, Guangzhou, China

<sup>6</sup>College of Computer Engineering and Sciences, Prince Sattam Bin Abdulaziz University, Al-Kharj, Saudi Arabia

## Corresponding author:

Guang Feng, Center of Network Information & Modern Education Technology, Guangdong University of Technology, No. 100 Waihuan Xi Road, Guangzhou Higher Education Mega Center Pany, Guangzhou 510006, China.

Email: von@gdut.edu.cn



Creative Commons CC BY: This article is distributed under the terms of the Creative Commons Attribution 4.0 License (<https://creativecommons.org/licenses/by/4.0/>) which permits any use, reproduction and distribution of the work

without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

## Introduction

Breast cancer is one of the leading causes of death among women in the world today.<sup>1-4</sup> It is one of the most invasive types of cancer, and statistics reveal that every year approximately 40,000 women die from breast cancer in the United States alone. In India, as per the reports of 2012, around 1,44,937 women had breast cancer and out of which 70,218 of them succumbed to death.<sup>4</sup> It is thus an enormously serious health concern for women worldwide. Breast cancer develops from breast tissue results in lump in the breast, changes in the shape of the breast, breast skin abnormalities, fluids from the nipple and many other repercussions. The accuracy in the prediction of breast cancer plays a vital role as it leads to timely and conclusive decision-making providing opportunity to avail the sophisticated modern day healthcare treatments. With the ever increasing number of techniques introduced for the cancer diagnosis in the present day and age, data mining and artificial intelligence (AI)-based diagnosis hold a popular and significant choice among researchers for the prediction and discovery of tissues affected by cancer. The data mining techniques are used to detect large, heterogeneous, time series and complex information from data sets aiding various diagnostic and therapeutic services in the health care industry. The data sets in these approaches are fragmented, distributed and then analysed to yield predictive results for accurate diagnosis.

AI is similarly a widely used technique for the detection and diagnosis of various types of cancer. AI tools are mostly used for classification and clustering of gene data of malignant cells, mutated cells and cancer-affected tissues and such approaches play significant role in cancer treatment as indicated by Agrawal and Agarwal.<sup>5</sup> Beni and Wang<sup>6</sup> mentioned swarm intelligence algorithm to be used for forecasting scenarios as well as monitoring and diagnosis of cancer. Various machine learning techniques and optimization methods integrated with swarm intelligence have been successfully used for predicting and diagnosing the cancer-affected tissues. This study proposes a new breast cancer determination and diagnosing method based on swarm intelligence for the detection of breast cancer.

Jele et al.<sup>7</sup> and Teague et al.<sup>8</sup> proposed a fine needle technique for aspiration biopsy (FNA) which is a simple and quick procedure for removing breast lesion by a fine needle. Chen et al.<sup>9</sup> mentioned the fact that machine learning characteristics and techniques are needed for diagnosis and also for finding any possible errors made by physicians due to time-constrained examinations. The study also informs that classification and machine learning methods work correctly when probability densities are known. In case density of data is not available estimations could be performed with

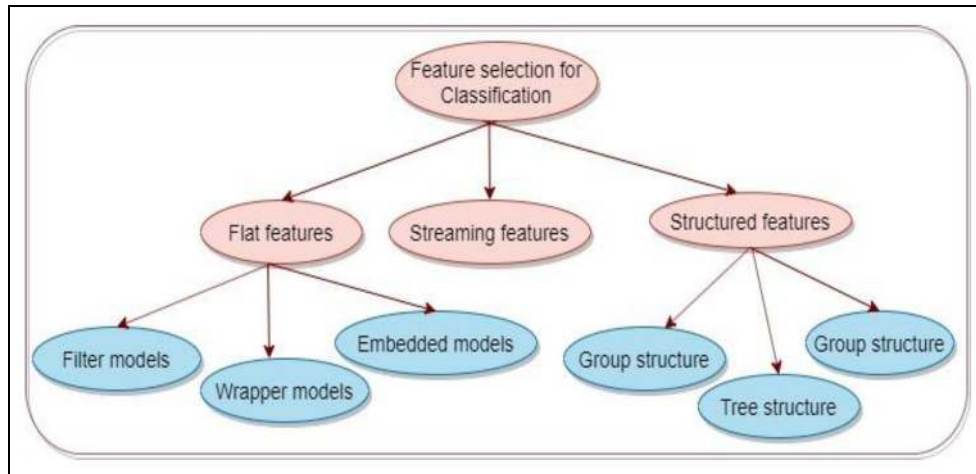
the help of kernel density estimation (KDE). Sheikhpour et al.<sup>13</sup> cited kernel density as a widely useful technique for data estimation. Kernel estimation depends on selection of the most relevant features which further improves the performance indicator. The main advantage of KDE used to increase the prediction accuracy with the help of class conditional densities of data using the Naïve Bayes (NB) classifier.

It is acknowledged that feature selection is crucial yet has associated challenges of consuming fair amount of space and search time. Computation thus plays a very important part in the performance of a feature selection method. Bolasn-Canedo et al.<sup>10</sup> classified feature selection into three types, namely, wrapper, embedded and filters.<sup>11-13</sup> The analysis of these three feature selection techniques identifies wrapper capable of providing the highest accuracy, as mentioned by Lal et al.<sup>14</sup> In consistence with the same computational aspect, Wang et al.<sup>15</sup> proposed particle swarm optimization (PSO) intelligence as a powerful computation method being applied in large number of applications. Similarly, Jia et al.<sup>16</sup> proposed the use of particle swarm intelligence approach, statistical and discrete high-quality feature subsets for estimations establishing the same as one of the best computing techniques predominantly used in various applications today.<sup>17</sup> The combination of PSO and feature selection algorithms has immense capability to produce optimized prediction results in cancer diagnosis.

This study introduces a new particle swarm optimization integrated with non-dominating object-based feature selection method (PSO-NDS). The selected features are then subjected to three classification algorithms such as  $k$ -nearest method, fast decision tree and KDE which predict breast cancer. The data sets used in the study are WBCD and WDBC data sets from publicly available from the UCI machine learning data repository. The evaluation metrics used are predictions rate, accuracy, sensitivity, specificity and time complexity. The main novelty of the present work is as follows:

- The user of multi-classifier which reduces the possibility of error in the results generated.
- The implementation of PSO algorithm in association with non-dominating sorting (NDS) and multi-classifier such as KNN, decision tree and KDE for prediction of breast cancer.
- The use of Bayes' theorem to further revise the results and obtain the best predictions.

The subsequent research work has been arranged as follows. Section 'Feature selection methods' elaborates on the feature selection method, and section 'PSO using non-dominating method' describes classification techniques, and proposed PSO-based multi-classification



**Figure 1.** Feature selection classification.

model is elaborated in section ‘Working and processing of PSO-NDS preprocessing’. The experimental setups of data and result comparisons are presented in section ‘Results and Discussion’. This article concludes with conclusion and future directions discussed in section ‘Conclusion’.

## Feature selection methods

Feature selection method is used to remove irrelevant features from the data set in order to infer accurate information from the data analysis performed on the data set. For example, the biological platform for feature selection is based on gathering similar gene expressions. This section presents a brief overview on feature selections based on biological and cancer-related research works conducted. Hira and Gillies mentioned that the objective of feature selection and extraction is to avoid over-fitting of the data to continue further analysis.<sup>18</sup> As mentioned in the study, feature selection method is divided into three sub-methods, namely, filter, wrapper and embedded classification as shown in Figure 1.

The filter method extracts data without any prior learning. The wrapper classifier uses prior learning with evaluation and embedded method is a combination of filter feature selection and embedded classification technique. It is thus prominent that decision-making based on the most appropriate feature selection method can be a difficult task. Various feature selection methods and types of classification are shown in Table 1.

Jafari et al.<sup>19</sup> mentioned in their study that the *T*-test feature selection technique finds the maximal difference of mean with minimal variable. Hall<sup>20</sup> proposed correlation-based feature selection method to be used to find highly correlated data wherein each classification should be uncorrelated. Rau et al.<sup>21</sup> identified

Bayesian network feature selection to be used in determining causal relationship with each class ensuring each class may not have any relationship. Yang et al.<sup>22</sup> highlighted that information gain could be used to measure common features performing comparisons with all the classes. The authors Ooi and Tan<sup>23</sup> proposed genetic algorithms (GAs) to be used to measure smaller set of features in order to produce the highest accuracy. Guyon et al.<sup>24</sup> proposed support-vector machine (SVM) technique for feature elimination and also pointed the fact that SVM classifiers could be a good choice and it omits irrelevant features using weighted approach.

Jiang et al.<sup>25</sup> proposed random forests for the creation of decision trees, the use of diverse samples of the original data and the use of different average algorithm for improving the accuracy. The study in Ma et al.<sup>26</sup> identified that least absolute shrinkage and selection operator (LASSO) method features are classified based on zero and nonzero. Anushaa and Sathiaseela<sup>27</sup> proposed the NLMOGA feature selection method which is based on constraint selection from a sub-population. This method involves finding the most suitable or closest set of objects from the group and uses Pareto’s front method to minimize the inner classes. In this method, computation cost and time complexity are quite high. The authors Khan and Baig<sup>28</sup> proposed multi-object GA (non-dominated sorting genetic algorithm (NSGA-II)) to be used for resolving multi-objective feature and subset feature-based problems wherein data used for large attributes and non-relevant features are eliminated. The proposed algorithm NSGA-II features involve measurement using ID3. The experiment is conducted with the help of NDS. Two classification algorithms are used such as ID3 and Pareto’s front. The NSGA-II method is applied to different applications such as salary prediction and DNA

**Table 1.** Various feature selection method and types of classification.

Method	Types of classifier	Linear	Non-linear
Feature selection using <i>T</i> -test (2006)	Filter	Yes	No
Feature selection with correlation (1998)	Filter	Yes	No
Bayesian networks (2010)	Filter	Yes	No
Genetic algorithms (2010)	Wrapper	No	Yes
Recursive feature elimination using SVM (2003)	Embedded	Yes	No
Random forests (2004)	Embedded	Yes	No
Selection operator and least absolute shrinkage (2007)	Embedded	Yes	No
<i>k</i> -means genetic algorithm for neighbourhood learning (2015)	Pareto front solution	Yes	
Feature subset selection using multi-objective optimization (2015)	Wrapper		Yes
Genetic algorithms for multi-objective (2015)	Wrapper, Pareto's front solution	Yes	No
Feature selection using genetic algorithm (2017)	STuMs	Yes	No
Feature subset selection using ant colony and hybrid approach (2018)	Multi-classifier	Yes	No
Particle swarm optimization (PSO) (2018)	Naïve Bayes, IBK and REP-Tree	Yes	No
PSO-KDE model (2018)	Kernel density estimation	Yes	No

SVM: support-vector machine; STuMs: support tracker machines.

sequence, yielding immensely satisfactory accuracy (95.2) of prediction of data.

The authors Zeng et al.<sup>29</sup> proposed a method for optimization that uses genetic algorithm kernel density estimation (GA-KDE) named support tracker machines (STuMs). The method sweeps out the irrelevant information yielding better accuracy. Naseer et al.<sup>30</sup> proposed a hybrid approach using ant colony optimization (ACO) and multi-classification techniques. This hybrid approach used a filter-based classifier to enhance the accuracy prediction. The experiment was conducted using four different classifiers over 11 data sets and then the hybrid method results were compared to the PSO and GA results. The evaluation results appear to be similar to the PSO method with 95.27% and 95.99% accuracy, respectively. Sakri et al.<sup>31</sup> proposed a PSO-feature selection method which is based on three classifiers such as *k*-nearest neighbour (KNN), NB and fast decision tree. This method had four processing steps such as data acquisition, data preprocessing, classification with and without feature selection, and finally the comparative analysis. The method yielded lesser accuracy in its prediction compared to the other methods. Sheikhpour et al.<sup>13</sup> proposed a particle swarm optimization kernel density estimation (PSO-KDE) for breast cancer detection. The main intent of this method was to increase the accuracy, thereby reducing errors. The method produced optimal accuracy level using only one classifier which had possibilities of generating erroneous predictions in comparison to other popular methods. It is thus observed from the related research work that majority of the studies conducted using the PSO feature selection method produced optimal results compared to the other methods yet had associated challenges pertaining to its application on only linear methods. On the contrary, it is

expected that prediction methods consider all dimensions of predictions. The KDE is a readily available package which can be used easily for classification and hence quite popular. Also the KDE works extremely accurately in case of bimodal or highly skewed distributions, especially used for estimates in discriminant analysis. The predictions are also more accurate as the misclassification rates are reduced.

The authors Reddy et al.<sup>32</sup> proposed a novel approach – deep neural network and support value (DNNS) for the prediction of breast cancer using large-scale data sets from a reputed hospital in India. The accuracy, precision and recall values resulting from the proposed methodology were compared with the state of the art techniques. The accuracy (97.21%) of the DNNS-based approach, although was better than the other traditional approaches but was not extremely promising. Ramadan et al.<sup>33</sup> proposed a computer-aided diagnosis (CAD) system for the detection of breast cancer. The framework involved the use of mammogram data in which were classified for the purpose of disease prediction. The study highlighted the various features and factors contributing towards detection of breast cancer using CAD systems. The comparisons of the various CAD methods are conducted and the receiver operating characteristic (ROC) is calculated. However, the CAD results have not been found reliable enough to be confidently considered as a standalone technique for breast cancer diagnosis. The study finally indicates the need of deep learning and similar approaches for enhancing the performances of CAD systems in order to generate more accurate detection results.

The study by Mohammed et al.<sup>34</sup> analysed breast cancer data using machine learning techniques, namely, KNN, SVMs, NB and various other classification methods. The results of the classifier were validated

**Table 2.** Classification accuracy with various techniques.

Data set	Classifier	PSO	GA	ACO	ACO-CE
WBCD and WDBC	Ripper	94.56	94.56	94.56	94.84
	KNN	95.27	94.84	95.27	95.99
	Naive Bayes	97.28	96.85	97.28	97.28
	Kernel density estimation	98.2	97	–	–

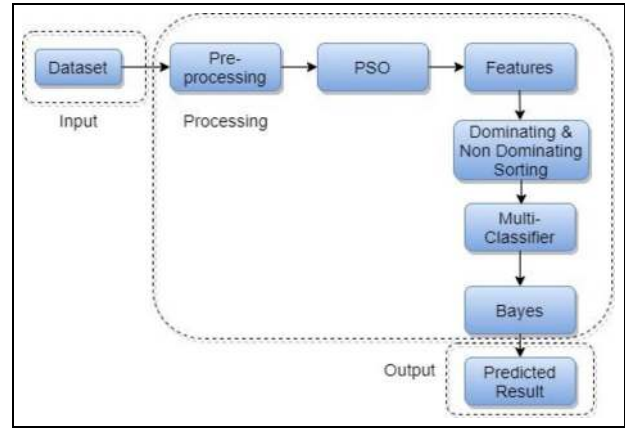
PSO: particle swarm optimization; GA: genetic algorithm; ACO: ant colony optimization; ACO-CE: ant colony optimization classifier ensemble; KNN: *k*-nearest neighbour.

and compared using two popular data sets – Wisconsin Breast Cancer (WBC) and the publicly available Breast Cancer data set. The study primarily emphasized on issues pertaining to handling of imbalanced data sets and resampling of the data was performed to resolve the issues. To evaluate the approach, 10-fold cross-validation was also performed and the efficiency of the classifiers was gauged using the true positive and false positive values, ROC, standard deviation and accuracy metrics. The comparative analysis identified sequential minimal optimization (SMO) as a better classifier after conducting data resampling on the WBC data set. In case of Breast cancer data set, the J48 algorithm generated better results after resampling was conducted. The study by Hou et al.<sup>35</sup> performed the evaluation of four machine learning algorithm in predicting breast cancer among Chinese women. The data set included breast cancer cases and healthy patient data, considered as control for the modelling, training and testing of the machine learning models. The metrics used for evaluation were namely – area under the curve (AUC), sensitivity, specificity and accuracy. The results justified the superiority of XGBoost algorithm in comparison to the other approaches. The various classification techniques, namely, PSO, GA, ACO and ant colony optimization classifier ensemble (ACO-CE) techniques and the relevant classifier accuracy results are depicted in Table 2.<sup>13,30</sup>

### PSO using non-dominating method

The main drawbacks of the previous methods as shown in Table 2 are that, it does not use multi-classifiers which enhances the possibility of high error and does not include re-verification of predicted results. The study consists of PSO, NDS and multi-classifier techniques such as *k*-nearest method, fast decision tree and KDE, and finally Bayes' theorem for revising the predicted results with the help of degree of belief for accuracy. The main objective of the study is to increase the accuracy, reducing the errors and revising or rechecking of the predicted results.

The multi-classification technique is used to increase the accuracy in different directions. The

**Figure 2.** Diagrammatic representation of PSO-NDS method.

non-dominating method is used to rank the selection features based on good points. Bayes' theorem is used to verify the results with a high degree of belief. The proposed model of PSO-NDS is depicted in Figure 2. The PSO method has three parts to it, such as the input, processing and the output.

The proposed method utilizes a non-linear method, and hence training samples  $((x_i, y_i), i = 1)$  from the input are considered high-dimensional feature space and the mapping function  $(\Phi)$  is called the kernel function  $(k)$ . The inner products use the kernel function. The non-linear approach uses the dual Lagrangian LD  $(\alpha)$

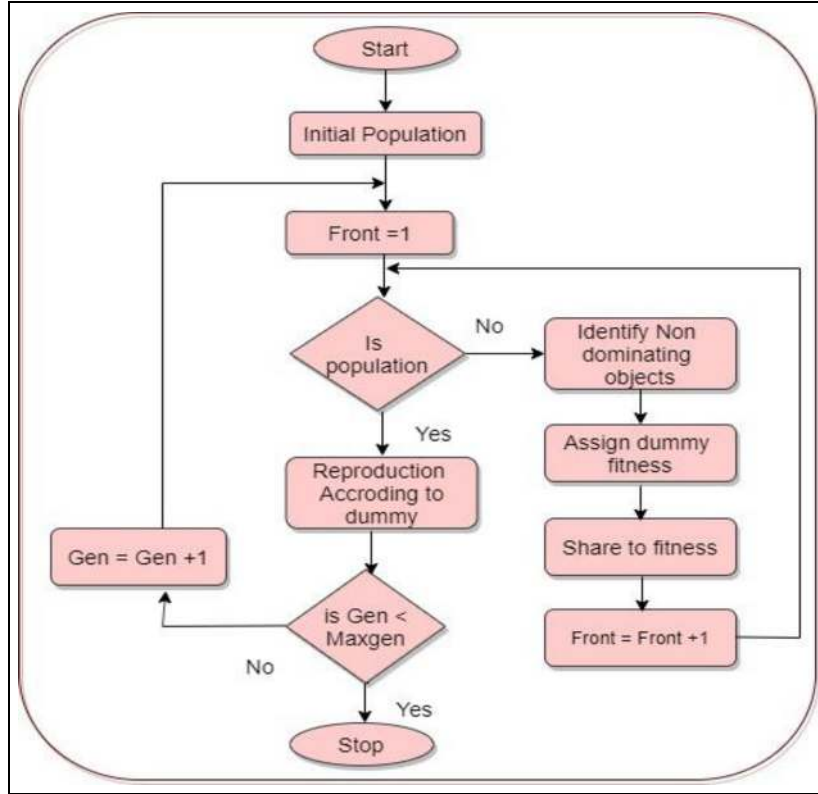
$$\Phi(x)_i, \Phi(y)_i = k(x_i, y_i) \quad (1)$$

$$LD(\alpha) = a_0 + \sum_{i=1}^m \alpha - \frac{1}{2} \alpha_j y_i y_j k(x_i, y_i) \quad (2)$$

Equation (2) is used to train the data in multi-directional ways and is used in feature selection and multi-classifications.

### PSO

PSO is used to find the objectives, position and velocity based on the timing  $t$ . The precious objects (cancer cell)



**Figure 3.** Flow representation of non-dominating sorting.

are in a different directional or multi-dimensional space as per the velocity ( $v$ ) and space. Each object position and velocity are calculated wherein  $x_i = \{x_1, x_2, x_3, \dots, x_n\}$  and  $v_i = \{v_1, v_2, v_3, \dots, v_n\}$ . The optimal solution searches for objects or particles and changes the position based on two factors—starting position of the particle and best position of the particle. The velocity and position of the particle are calculated using the following equation

$$V_{i+1} = \omega v_i + C_1 * rand() * PB_i - X_i + C_2 * rand() * (GB_i - X_i) \quad (3)$$

$$X_{i+1} = X_i + V_{i+1} \quad (4)$$

where  $\omega$  is the initial weight,  $v$  is the velocity,  $X$  is the position,  $C_1$  and  $C_2$  are the learning factors,  $PB$  is the personal best performance – best performance of the group. The basic processing steps of PSO algorithm are initialization, evaluation, finding the position of the particle, finding the best, and updating the velocity and position stopping and initialization the evaluation.

### NDS

Population and objective of the function are calculated based on non-dominating objects.<sup>28</sup> The NDS technique is used to evaluate each subset of the data. Based on the features, if the subset of data is tidy, it is passed

to evaluate and each subset data is assigned fitness values. In addition, an initial distance is used to calculate the feature subset. The crowding distance is calculated to find out how close the objects are to their neighbours. The large average data or the resultant cording distance value produces enhanced diversity. The population selection is based on ranking and crowding distances. Hence, the decision is taken based on the crowding distance. The selected distance generates (gen) crossover and mutation operators. The objective of the function with current cancer objects are sorted based on non-domination with  $N$  individuals' selection.  $N$  is the population size or objective function size. The final population size depends on the crowd distance pertaining to the feature subset in the cancer cell. The overall operation and flow representation of NDS techniques are shown in Figure 3. The sharing of distance and non-dominating object sharing are derived in equation (5)

$$sh(d_{ij}) = \left\{ 1 - \left( \frac{d_{ij}}{\alpha \text{ share}} \right)^2, \text{ if } d_{ij} < \alpha \text{ share}, \text{ otherwise } 0 \right\} \quad (5)$$

where  $d_{i,j}$  is the distance between two individual objects, and  $\alpha \text{ share}$  is the distance allowed between two objects.

### Multi-classification

The multi-classification method, also termed as multinomial classification is used to test the accuracy on a given data set with different identifications and labels. The multi-classifier uses different classifiers wherein each classifier uses different features for its prediction. Each feature of the data sets usually has different instances or sub-features. When multi-classifiers are used, these sub-features are also analysed which enhances the prediction results and also the prediction rate. The advantage of using a multi-classifier thus lies in the improvement of prediction rate and prediction results. However, it also has its associated challenges pertaining to increase in time complexity involved in the process of analysing the various features at different levels.

The classification is based on various instances and steps involved in multi-classification are as follows:

- Step 1: loading the test data.
- Step 2: labelling the data set into test and training.
- Step 3: training KNN classifier, fast decision classifier and kernel density classifier.
- Step 4: using the classification to predict the test data.
- Step 5: measuring the accuracy.

**KNN classifier.** KNN algorithm is a non-parametric classification and regression technique mentioned by Sakri et al.<sup>30</sup> In this study,  $k$ -nearest training feature space is used as input. The output is the most common nearest  $k$ -positive integer. If  $k = 1$ , it is assigned to nearest neighbours.

**Fast decision tree.** Fast decision tree classification is used for decision-making from large data sets as mentioned by Manapragada et al.<sup>36</sup> The algorithms help in decision-making<sup>37-46</sup> without compromising on the accuracy and to top it up increases the space complexity. This study implements fast decision tree with conditional independence. The conditional independency information gaining (IG) is defined as

$$GA(S, X) = E(S) - \frac{\Sigma(S_x)}{S(S_x)} \quad (6)$$

where  $S$  is a set of training instance,  $X$  is the attributes and  $x$  is the value,  $E$  is an entropy,  $S_x$  is a subset of instance and similarity entropy id defined as  $P_s$ ,

$$E(S) = \sum_{i=1}^{|C|} (P_s)(C_i) \log P_s(C_i) \quad (7)$$

$P_s(C_i)$  percentage of instance belongs to  $C_i$  and  $|C|$ .

**Kernel density classifier.** Sheikhpour et al.<sup>13</sup> proposed the kernel density method which focuses on identifying past conditions similar to the prediction time. This method directly estimates the density of data without any assumption. Considering  $\{x^t\}$  as independent having  $d$ -distributed training data with unknown distance  $P(x)$ , wherein  $x$  is the closer point. The density is defined as

$$P(x) = 1/h[N(xt \leq x' + h) - N(xt \leq x')]/N \quad (8)$$

where  $x^t$  is the training instance,  $x'$  is the new arrival data,  $N$  is the number of instances, and  $h$  is the length of interval.

This multi-classification technique increases the accuracy of the prediction as the prediction rate is measured in different scenarios. The first type of multi-classifier is KNN classifier and it classifies the affected cells or particles near to the adjacent cells. If the cells are affected, then it is considered as 1 and the surrounding cells are scanned. Similarly, all the affected cells are measured in each iteration of the scanning process. The second type of multi-classifier is the fast decision tree that increases the prediction rate in large data sets. When the data set or prediction range of the surface increases, the prediction accuracy decreases automatically. The fast decision tree increases space complexity of the prediction and gains information from the large space based on conditional independency of objects. Each positive cell is predicted using trained data and subsets of affected cells. Simultaneously when space of the prediction is increases, thereby false rate of the predictions using various instances reduces. The third classifier is the density classifier and this classifier classifies the cancer cells based on the past training experience. Once the affected cells are predicted, the corresponding density of the cells is measured without any assumptions. These three types of multi-classifiers thus contribute immensely in increasing the prediction rate and accuracy. The most important advantage in using the multi-classifier lies in the increase of prediction rate and prediction accuracy in comparison to the traditional methods. On the contrary, the disadvantages include the increase in time complexity in the prediction and analysis of the features.

### Bayes' theorem

The main usage of Bayes' theorem is updating the prediction probability and increasing the belief rate using predicted results of affected cells. Mathematically, Bayes' theorem is represented as follows

$$(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (9)$$

For proposition  $A$  and evidence  $B$ ,  $P(A)$ , the prior, is the initial degree of belief in  $A$ . The quotient  $P(B|A)/P(B)$  represents the support  $B$  provides for  $A$ .  $P(A|B)$ , the posterior, is the degree of belief having accounted for  $B$ .

In this work,  $A$  and  $B$  are correlated incorporating accuracy generated due to the implementation of multi-classifier and PSO. The error is also calculated with help of the following formula<sup>13</sup>

$$Error\ rate = \frac{\sum_{i=1}^{10} Error\ rate_i}{10} \quad (10)$$

$Error\ rate$  is the error rate for each data set.

In the proposed work,  $A$  is considered as the prediction evidence of breast cancer, and it also considered as the high-level belief.  $B$  is the corresponding evidence of the prediction.

## Working and processing of PSO-NDS preprocessing

WBCD<sup>47</sup> and WDBC<sup>48,49</sup> data sets are used in this study for learning and analysis of the data set. WDBC presents 569 instances with 30 features and WBCD presents 699 instances. The data are collected from various cases, images of human breast tissues which are digitalized. The various parameters are analysed using predefined parameters used such as velocity, starting point and objective function. The analysis and preprocessing technique followed is based on PSO-KDE approach<sup>13</sup> described in the following section.

### Processing of PSO-NDS

The proposed multi-model consists of non-dominated sorting, PSO and multi-classification. Bayes' theorem is used for analysing the various factors. The processing of PSO-NDS are mentioned as follows:

- Step 1: parameters of PSO and instance of particles are first conserved.
- Step 2: particle position and velocity of search space are initiated.
- Step 3: object function of particle is calculated using PSO-NDS.
- Step 4: using PSO-NDS, the various objective function is updated and individual performance of objects is updated.
- Step 5: velocity and position of the particles are updated.
- Step 6: if the desired number of iterations is not reached, the return to Step 3.
- Step 7: the features and accuracy are presented.

The algorithm of PSO using NDS with multi features is shown in Algorithm 1.

**Algorithm 1.** Particle swarm optimization using non-dominating sorting.

---

**Input:** Data sets with various labels  
**Output:** Feature subset  
**Method:**  
 Initial parameter: size, iteration, number of particles, search space, velocity with population  
 Iteration (t) = 1  
 If t ≤ maximum iteration  
 Calculate the objective function using NDS  
 For i = 1 to population size do  
 Objective of function in various stages updated  
 End if  
 If objective function updated < maximized function  
 Update the function up to reach the objective function  
 End if  
 End for  
 Update the velocity and population  
 t = t + 1;  
 Provide the best objective function

---

NDS: non-dominating sorting.

The PSO-NDS with multi-classification simulation determines the accuracy and Bayes' theorem is used to make the relevant factors to maximize the objective function and features subset. This helps to maximize the accuracy of the prediction enhancing the accuracy rate. In PSO using NDS, the particles  $pbest$  are not compared with its potential offsprings, rather the  $pbest$  of the entire population of  $N$  particles and the  $N$  of these particles are combined to form a population of  $2N$  particles temporarily. On these  $2N$  particles, the non-dominated sorting method is applied which sorts the entire population into non-domination fronts. Here, the first front is a non-dominant set in the current population and the second front is dominated by individuals in the first one and process continues in the similar pattern. The individuals in each front are assigned a fitness value based on this source front. As an example, the individuals of the first front are assigned a fitness value of 1, individuals of the second front are assigned a value of 2 and this trend continues. Apart from the fitness value, the parameters – crowding distance and niche count – are computed for each individual to obtain the best distribution of non-dominant solutions. The overall summarizing step of the proposed method is mentioned as follows:

- Step 1: initialize the parameter.
- Step 2: initialize the population.
- Step 3: objective function is calculated using PSO-NDS.
- Step 4: selection, mutation and crossover selected over particle.



**Table 3.** Accuracy comparison between various methods (GA-KDE, PSO-KDE, PSO-NDS).

Data sets	PSO-KDE		GA-KDE		PSO-NDS (1) (proposed)		PSO-NDS (2) (proposed)	
	Accuracy	Feature number	Accuracy %	Feature number	Accuracy	Feature number	Accuracy	Feature number
WBCD	96.14	5.5	96.12	5.5	98.4	5.5	98.28	5.6
WDBC	97.21	15.1	97.01	14.5	98.88	15.15	98.8	15.4

PSO: particle swarm optimization; NDS: non-dominating sorting; KDE: kernel density estimation.

Step 5: accuracy is calculated using multi-classification techniques.

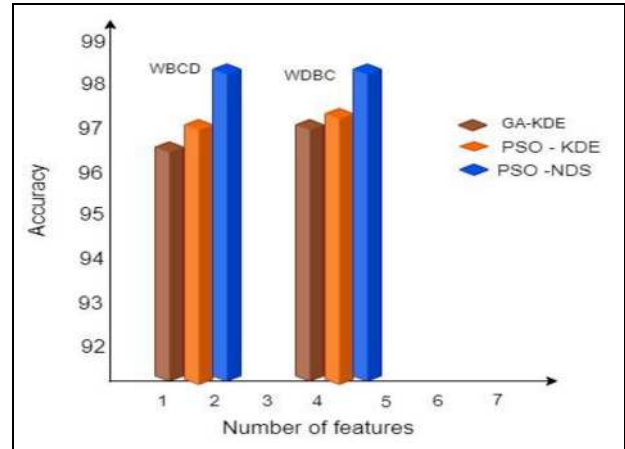
Step 6: using Bayes' theorem, supporting factors are calculated.

Step 7: satisfied objective function is generated.

## Results and discussion

The proposed PSO-NDS model was used for the prediction of breast cancer. For experimentation purposes, WBCD and WDBC data sets were used. For the performance evaluation purpose, various parameters such as number of iterations, velocity and initial position of particles are used and considered for experimentation. The WBCD and WDCD data sets are randomly divided into subsets and used for training. The number of iterations is set as 20. The experiment is conducted on the above parameters with 20 iterations. The justification on the superiority of PSO-non-dominating method against the traditional methods is prominent from the generated results. The PSO-non-dominating method produces better results in identifying the benign and malignant features or sub-features. The predictions are also more accurate and computed considering each feature as per the dominant properties. Finally, the computation of the study results of PSO-NDS is compared with PSO-KDE<sup>13</sup> and GA-KDE.<sup>29</sup> The experiment is validated using accuracy, specificity and sensitivity were validated and performed.

The performance of PSO-NDS, PSO-KDE and GA-KDE experiment is analysed on training data sets to gain the ideal feature subset and classification. The accuracy of PSO-KDE and GA-KDE were almost similar but yielded inferior accuracies when compared to the proposed methods. Razieh et al.<sup>13</sup> proposed a PSO-KDE for breast cancer detection. The main intent of this method was to increase the accuracy, thereby reducing errors. The method produced optimal accuracy level using only one classifier which had possibilities of generating erroneous predictions in comparison to other popular methods. It is thus observed from the related research work that majority of the studies conducted using the PSO feature selection method produced optimal results compared to the other methods.

**Figure 4.** Accuracy of WBCD and WDBC.

However, there exist associated challenges pertaining to its application on linear methods. In case of GA-KDE, the GA and non-parametric KDE-based classifier are hybridized to compute the optimal bandwidth and also the subset of features.

The experiment is conducted using various trails and average accuracy is calculated. Table 3 and Figure 4 show the accuracy of PSO-NDS in comparison to other stated of the art methods with relevant features mentioned.

The PSO-NDS achieved 98.28% and 98.8% accuracy for features between 5–6 and 15–20, respectively, as shown in the figure for WBCD and WDCD data sets. The proposed work obviously improved the accuracy compared to other studies conducted. The accuracy of the prediction of WBCD and WDCD data sets is generally high. But, in a real world scenario, when data sets with huge amounts of data are subjected to scanning and prediction, the prediction rate and accuracy of the prediction decrease as a natural effect. However, in the present work, the Bayes theorem is implemented to elevate the accuracy. The supporting factors and evidence are verified using revised mechanism, and the false positive prediction is easily reduced using the Bayes theorem. The increase in prediction accuracy is achieved by reducing the number of features, considering the most relevant ones in the analysis as shown in Table 4.

**Table 4.** Performance comparison between various methods (GA-KDE, PSO-KDE, PSO-NDS).

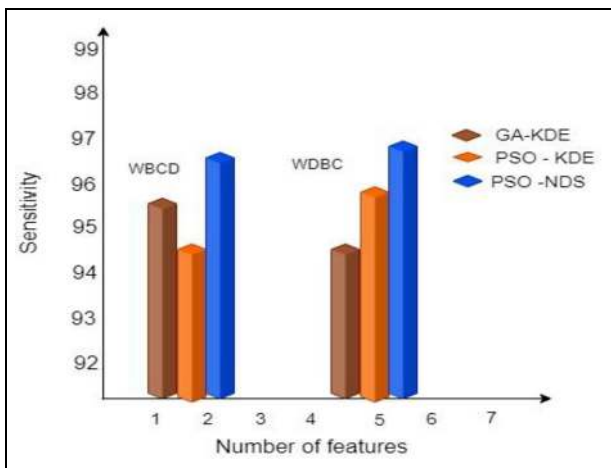
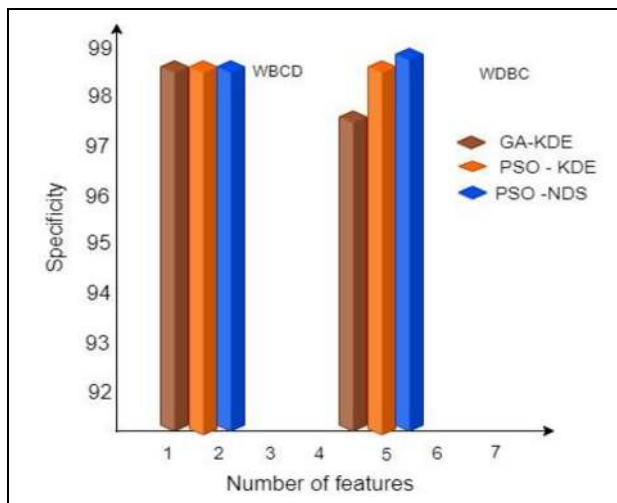
Data sets	Measurements	GA-KDE	PSO-KDE	PSO-NDS(1)	PSO-NDS (2)	PSO-NDS (3)
WBCD	Accuracy	96.12	96.14	98.4	98.28	98.88
	Sensitivity	95.1	94.84	96.02	95.5	96.8
	Specificity	99.86	99.86	99.8	99.7	99.8
WDBC	Accuracy	97.01	97.21	98.88	98.8	98.6
	Sensitivity	94.32	96.14	97.12	96.12	97
	Specificity	97.62	98.96	98.12	98.2	98.3

PSO: particle swarm optimization; NDS: non-dominating sorting; KDE: kernel density estimation.

**Table 5.** Error range.

Data sets	Measurements	PSO-NDS (1)	PSO-NDS (2)	PSO-NDS (3)
WBCD	Error range	0.5	0.8	0.4
WDBC	Error range	0.6	0.5	0.3

PSO: particle swarm optimization; NDS: non-dominating sorting.

**Figure 5.** Sensitivity of WBCD and WDBC.**Figure 6.** Specificity of WBCD and WDBC.

The average sensitivity of PSO-NDS is also compared with PSO-KDE and GA-KDE. The performance evaluations are shown in Table 4 and Figure 5. Similarly, Table 4 and Figure 6 show the comparative values of specificity wherein the performance of the proposed method proves to be better.

In Figure 5 and Table 4, the sensitivity of prediction is increased in both WBCD and WDBC data sets. In these data sets, minimum eight features are considered for predictions. But in the 12, 18 and 24 features, the false positive will increase. In Figures 4 and 5, if the features are not selected, that particular feature not used for prediction and that particular features are not affected.

In the data set used, as well as in case of real-time data, the number of the features decreases

automatically and positive test cases increase with prediction rate decreasing otherwise. Figure 6 shows similar decrease in specificity (correctly generated negative).

Table 5 shows the error rate of WBCD and WDBC of data set computation. The error ranges are less than 1.0. So for the predicted results, both sensitivity and specificity are achieved in the maximum 12 features of proposed work. But the number of prediction factors or features is more in the WDBC data sets, and in case of huge real data sets, the error rate may be increased.

The other factors that help to achieve the accuracy are the supporting factors such as objective function of particles, multi-classification prediction parameter and accuracy, all of which are considered for prediction of results. For the supporting rate and evidence of belief

**Table 6.** Supporting factor rate and evidence of belief (10).

Data Sets	Measurements	PSO-NDS (1)	PSO-NDS (2)	PSO-NDS (3)
WBCD	Belief rate	10	9.9	10
WDCD	Belief rate	10	10	10

around 10 factors are considered and based on the 10 factors, the results and prediction are supported to achieve the prediction results for breast cancer as shown in Table 6. In Bayes' theorem, the probability measures the 'degree of belief'. Also the prediction is based either on a single features, collected evidence or conditional probability which updates the belief evidence. It basically shows how a 'degree of belief' when expressed as a probability changes realistically based on the account of available evidence.

## Conclusion

The proposed work emphasizes on predicting breast cancer with optimal level of performance measures using a multi-modal model. The data set used in the study is WBCD and WDCD which are popularly available for the purpose of conducting research. The proposed multi-modal classification model consists of various techniques, namely, NDS method, multi-classifier and Bayes' theorem to accurately classify the breast cancer data sets. The proposed multi-modal classification model is called PSO-NDS. The error factors were reduced using supporting factors with evidence of various factors. The proposed work is applied into an  $n$ -dimensional space. The proposed PSO-NDS model when implemented on the WBCD and WDCD data sets generated optimum (98.8%, 98.6%) level of accuracy. The sensitivity and specificity achieved were 98.8%, 97.12% and 99.8% and 98.3% which are quite promising. The future direction of research would involve prediction and detection of cancer cells using Internet of thing (IOT) devices.


## Declaration of conflicting interests


The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

## ORCID iDs

Senthilkumar Mohan  <https://orcid.org/0000-0002-8114-3147>

Usman Tariq  <https://orcid.org/0000-0001-7672-1187>

## References

1. Sheikhpour R, Ghassemi N, Yaghmaei P, et al. Immuno-histochemical assessment of P53 protein and its correlation with clinicopathological characteristics in breast cancer patients. *Indian J Sci Technol* 2014; 7(4): 472–479.
2. Kaya Y. A new intelligent classifier for breast cancer diagnosis based on a rough set and extreme learning machine: RS + ELM. *Turk J Electr Eng Comput Sci* 2013; 21: 2079–2091.
3. <http://news.mit.edu/2017/artificial-intelligence-early-breast-cancer-detection-1017>
4. [http://www.breastcancerindia.net/statistics/stat\\_global.html](http://www.breastcancerindia.net/statistics/stat_global.html)
5. Agrawal S and Agrawal J. Neural network techniques for cancer prediction: a survey. *Procedia Comput Sci* 2015; 60: 769–774.
6. Beni G and Wang J. Swarm intelligence in cellular robotic systems. In: *Proceedings of the NATO advanced workshop on robots and biological systems*, Tuscany, Italy, 26–30 June 1993. Berlin: Springer.
7. Jeleń Ł, Fevens T and Krzyżak A. Classification of breast cancer malignancy using cytological images needle aspiration biopsies. *Int J Appl Math Comput Sci* 2008; 18(1): 75–83.
8. Teague MW, Wolberg WH, Street WN, et al. Fine needle aspiration cytology of lymph nodes in breast cancer follow-up is a feasible alternative to watchful waiting and to histology. *BMC Womens Health* 1997; 81(2): 129–135.
9. Chen HL, Yang B, Liu J, et al. A support vector machine classifier with rough set-based feature selection for breast cancer diagnosis. *Expert Syst Appl* 2011; 38(7): 9014–9022.
10. Bolasn-Canedo V, Sanchez-Marño N and Alonso-Betanzos A. A review of feature selection methods on synthetic data. *Knowl Inf Syst* 2013; 34(3): 483–519.
11. Maldonado S, Weber R and Basak J. Simultaneous feature selection and classification using kernel-penalized support vector machines. *Inf Sci* 2011; 181(1): 115–128.
12. Guyon I and Elisseev A. An introduction to variable and feature selection. *J Mach Learn Res* 2003; 3: 1157–1182.
13. Sheikhpour R, AghaSarram M and Sheikhpour R. Particle swarm optimization for bandwidth determination and feature selection of kernel density estimation based classifiers in diagnosis of breast cancer. *Appl Soft Comput* 2016; 40: 113–131.
14. Lal TN, Chapelle O, Weston J, et al. Embedded methods. In: Guyon I, Nikravesh M, Gunn S, et al. (eds) *Feature extraction*. Berlin: Springer Heidelberg, 2006, pp.137–165.
15. Wang X, Yang J, Teng X, et al. Feature selection based on rough sets and particle swarm optimization. *Pattern Recognit Lett* 2007; 28(4): 459–471.

16. Jia L, Cheng D and Chiu MS. Pareto-optimal solutions based multi-objective particle swarm optimization control for batch processes. *Neural Comput Appl* 2012; 21(6): 1107–1116.
17. Ouyang A, Tang Z, Zhou X, et al. Parallel hybrid PSO with CUDA for 1D heat conduction equation. *Comput Fluids* 2014; 110: 198–210.
18. Hira ZH and Gillies DF. A review of feature selection and feature extraction methods applied on microarray data. *Adv Bioinf* 2015; 2015: 198363.
19. Jafari P and Azuaje F. An assessment of recently published gene expression data analyses: reporting experimental design and statistical factors. *BMC Med Inf Decis Making* 2006; 6(1): 27.
20. Hal MA. Correlation-based feature selection for machine learning. In: *Proceedings of the 2018 3<sup>rd</sup> international conference on communication and electronics systems (ICCES)*, Coimbatore, India, 15-16 October 1998. New York: IEEE.
21. Rau A, Jaffrézic F, Foulley J-L, et al. An empirical Bayesian method for estimating biological networks from temporal microarray data. *Stat Appl Genet Mol Biol* 2010; 9: 9.
22. Yang P, Zhou BB, Zhang Z, et al. Multifilter enhanced genetic ensemble system for gene selection and sample classification of microarray data. *BMC Bioinformatics* 2010; 11(suppl1): S5.
23. Ooi CH and Tan P. Genetic algorithms applied to multi-class prediction for the analysis of gene expression data. *Bioinformatics* 2003; 19(1): 37–44.
24. Guyon I, Weston J, Barnhill S, et al. Gene selection for cancer classification using support vector machines. *Mach Learn* 2002; 46(1–3): 389–422.
25. Jiang H, Deng Y, Chen HS, et al. Joint analysis of two microarray gene-expression data sets to select lung adenocarcinoma marker genes. *BMC Bioinformatics* 2004; 5: 81.
26. Ma S, Song X and Huang J. Supervised group Lasso with applications to microarray data analysis. *Bioinformatics* 2007; 8: 60.
27. Anushaa M and Sathiaseela JGR. Feature selection using K-means genetic algorithm for multi-objective optimization. *Procedia Comput Sci* 2015; 57: 1074–1080.
28. Khan A and Baig AR. Multi-objective feature subset selection using non-dominated sorting genetic algorithm. *J Appl Res Technol* 2015; 13(1): 145–159.
29. Zeng D, Wang S, Shen Y, et al. A GA-based feature selection and parameter optimization for support tucker machine. *Procedia Comput Sci* 2017; 111: 17–23.
30. Naseer A, Shahzad W and Ellahi A. A hybrid approach for feature subset selection using ant colony optimization and multi-classifier ensemble. *Int J Adv Comput Sci Appl* 2018; 9(1): 090142.
31. Sakri SB, Rashid NBA and Zain ZM. Particle swarm optimization feature selection for breast cancer recurrence prediction. *Special Section on Big Data Learning and Discovery*. Epub ahead of print 4 June 2018. DOI: 10.1109/ACCESS.2018.2843443.
32. Reddy GT, Reddy MPK, Lakshmana K, et al. Hybrid genetic algorithm and a fuzzy logic classifier for heart disease diagnosis. *Evol Intel* 2019; 13: 1-12.
33. Ramadan SZ. Methods used in computer-aided diagnosis for breast cancer detection using mammograms: a review. *J Healthc Eng* 2020; 2020: 9162464.
34. Mohammed SA, Darrab S, Noaman SA, et al. Analysis of breast cancer detection using different machine learning techniques. In: Tan Y, Shi Y and Tuba M (eds) *Data mining and big data. DMBD 2020. Communications in computer and Information Science*, Vol 1234, Singapore: Springer, 2020.
35. Hou C, Zhong X, He P, et al. Predicting breast cancer in Chinese women using machine learning techniques: algorithm development. *JMIR Med Inform* 2020; 8: e17364
36. Manapragada C, Webb G and Salehi M. Extremely fast decision tree. Arxiv Preprint arXiv: 1802.08780, 2018.
37. Tang J, Alelyani S and Liu H. Feature selection for classification: a review. In: Aggarwal C (ed.) *Data classification: algorithms and applications*. Boca Raton, FL: CRC Press, 2013.
38. Reddy GT, Reddy MPK, Lakshmana K, et al. Analysis of dimensionality reduction techniques on big data. *IEEE Access* 2020; 8: 54776–54788.
39. Iwendi C, Maddikunta PKR, Gadekallu TR, et al. A metaheuristic optimization approach for energy efficiency in the IoT networks. *Software: Pract Exper*. Epub ahead of print 11 February 2020. DOI: 10.1002/spe.2797.
40. Gadekallu TR, Khare N, Bhattacharya S, et al. Early detection of diabetic retinopathy using PCA-firefly based deep learning model. *Electronics* 2020; 9(2): 274.
41. Moreira MW, Rodrigues JJ, Kumar N, et al. Postpartum depression prediction through pregnancy data analysis for emotion-aware smart systems. *Inf Fusion* 2019; 47: 23–31.
42. Garg S, Kaur K, Kumar N, et al. Hybrid deep-learning-based anomaly detection scheme for suspicious flow detection in SDN: a social multimedia perspective. *IEEE Trans on Multimed* 2019; 21(3): 566–578.
43. Moreira MW, Rodrigues JJ, Korotaev V, et al. A comprehensive review on smart decision support systems for health care. *IEEE Syst J* 2019; 13(3): 3536–3545.
44. Reddy GT and Khare N. Heart disease classification system using optimised fuzzy rule based algorithm. *Int J Biomed Eng Technol* 2018; 27(3): 183–202.
45. Patel H, Singh Rajput D, Thippa Reddy G, et al. A review on classification of imbalanced data for wireless sensor networks. *Int J Distr Sensor Netw* 2020; 16(4): 1550147720916404.
46. Mangasarian OL, Setiono R and Wolberg WH. Pattern recognition via linear programming: Theory and application to medical diagnosis. *Large-scale Num Optim* 1990; 878: 22-31.
47. Mangasarian OL, Street WN and Wolberg WH. Breast cancer diagnosis and prognosis via linear programming. *Oper Res* 1995; 43(4): 570–577.
48. Wolberg WH, Street WN, Heisey DM, et al. Computer-derived nuclear features distinguish malignant from benign breast cytology. *Human Pathol* 1995; 26(7): 792–796.
49. Yeh WC, Chang W-W, Chung YY, et al. A new hybrid approach for mining breast cancer pattern using discrete particle swarm optimization and statistical method. *Expert Syst Appl* 2009; 36(4): 8204–8211.