IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 6 April 2024, accepted 12 April 2024, date of publication 16 April 2024, date of current version 24 April 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3390008

RESEARCH ARTICLE

Health-Related Data Analysis Using Metaheuristic Optimization and Machine Learning

ANNISA DARMAWAHYUNI[1,2], SITI NURMAINI[2], (Member, IEEE), BAMBANG TUTUKO[2], MUHAMMAD NAUFAL RACHMATULLAH[2], FIRDAUS FIRDAUS[2], ADE IRIANI SAPITRI[2], ANGGUN ISLAMI[2], JORDAN MARCELINO[2], RENDY ISDWANTA[2], AND MUHAMMAD IRFAN KARIM[2]

[1]Faculty of Engineering, Universitas Sriwijaya, Palembang 30139, Indonesia
[2]Intelligent System Research Group, Universitas Sriwijaya, Palembang 30139, Indonesia

Corresponding author: Siti Nurmaini (sitinurmaini@gmail.com)

This work was supported by Intelligent Systems Research Group (ISysRG), Faculty of Computer Science, Universitas Sriwijaya, Palembang, Indonesia.

ABSTRACT Health-related data has a decisive role in disease diagnosis. Collecting relevant information from health-related data in medical records has been facilitated by evaluating the features of the data. Relevant research has shown that outcomes are significantly impacted by the use of feature selection (FS) in different medical domain data. FS provides an analysis of the most significant features to improve classification accuracy. The FS technique aims at minimizing the number of input variables and computational overload to maximize classification performance results. However, identifying the optimal features poses issues due to the high dimensionality of large features and the small sample size of health-related data. The metaheuristics optimization algorithm (MOA) plays an important role in generating the best subset features with exploration and exploitation phases. This study experiments with well-known MOAs and supervised learning from the UC Irvine Machine Learning Repository, PhysioNet, Kent Ridge Bio-Medical Dataset, and MIMIC-III v1.4 Repository with varying feature dimensions. To increase the quality of health-related data, this study proposes missing data imputation based on a deep learning approach, an autoencoder (AE). With AE imputation, the performance results obtain 0.0167 mean squared error (MSE) and 0.129 root mean squared error (RMSE). As a result, MOA shows its excellence in achieving minimal features, but still outstanding performance in low- and high-dimensional data. MOA is successfully applied to varying diverse health-related datasets with low- and high-dimensional data.

INDEX TERMS Autoencoder, classification, data imputation, feature selection, health-related dataset, metaheuristic algorithms.

I. INTRODUCTION

Accurate assessment of information in health-related data will become an increasingly important challenge for research and access to large amounts of data. In addition, health-related data in databases related to clinics is growing at a much faster rate [1]. Therefore, it is crucial to retrieve health-related data from a huge amount of data so that the information that is retrieved might aid in the diagnosis and treatment of a variety of patient conditions. Collecting relevant information from health-related data has been facilitated by evaluating the features of data. The relevant features of health-related data are crucial to diagnose diseases [2]. The relevant attribute in the health-related data is extensive and heavily weighted. The redundant features affect the performance of the algorithm and add to the computational costs [3]. Relevant literature has shown that using feature selection (FS) to analyze data from different medical domains has a significant impact on the results [2], [3], [4]. Since most health-related data have a huge number of features, it is important and difficult to effectively extract possible risk factors utilizing FS approaches [2].

FS is the process of selecting the optimum subset of features, in which the relevant features are selected and

The associate editor coordinating the review of this manuscript and approving it for publication was R. K. Tripathy.

irrelevant features are removed [4], [5]. FS aims to maximize the classification accuracy and minimize the number of selected features [6]. FS plays a decisive role as an important preprocessing step for several machine learning tasks. The selection of essential features can also reduce the computational cost and improve the understanding of the problem. However, FS poses a challenge with health-related data with a massive number of features close to or larger than the number samples size (high-dimensional data), such as microarray and biomolecules data (i.e., deoxyribonucleic acid (DNA), ribonucleic acid (RNA), proteins, and metabolites) [7], [8]. Small sample size sets make machine learning algorithms not have enough space to learn the training samples, so it has a high risk of overfitting [9].

To overcome such a problem, this study proposes the FS method to obtain the optimum subset of features from varying dimensional. A wrapper approach as the FS method draws attention to its excellence in improving classification performance. Jovic et al. [10] have categorized the wrapper method based on search strategy, i.e., exponential, sequential, and randomized selection strategy. Unfortunately, the exponential strategy is not practically possible due to the number of evaluated features increasing exponentially with the size of the features. It is hard to handle high-dimensional data problems. In addition, sequential search tends to lead to local optima, because it includes or removes the features sequentially [6]. Hence, addressing the generation of subset features (feature search) through randomized selection strategies such as metaheuristic optimization algorithms (MOA) has been proposed in one decade of research [6], [11], [12], [13], [14], [15], [16], [17], [18], [19], [20]. MOA is a derivative-free technique and can avoid local optima and prevent the algorithms from premature convergence [21]. MOA has two main components for the generation of subset features, i.e., exploration and exploitation [22]. In the exploration process, MOA explores the entire search space to find a promising search space, then exploits the essential information found in the local search space of promising areas that are found in the exploration process.

To increase health-related data quality, this study also proposes an algorithm for handling missing data in medical records. Due to the high diversity and volume of medical data, the resulting medical records are highly susceptible to quality issues, such as missing information and errors in data entry [23], [24], [25]. For example, the issues of data acquired from high-throughput omics, such as low sensitivity in protein and peptide detection can affect the biological sample analysis [8]. Incomplete evaluations of a patient's status might result from missing information, which can have a negative impact on clinical decision-making and patient outcomes. Thus, accurate imputation of missing data is necessary to diagnose patient's conditions [23]. Statistical methods have been applied to data imputation, which replaces the missing observations with the most similar ones among the training data (mean or mode imputation) [26]. However, such a method does not preserve the relationship among variables, in addition, in high-dimensional data, mean imputation cannot account for dependence structure among features [27]. Conventional data imputation methods are prone to adding biases [8]. Deep learning (DL) has received attention for solving data imputation. Autoencoder (AE), as the DL approach has the capability of learning from corrupted data, which is a natural extension to the field of missing data [28]. It attempts to replicate a representation of the data at the output layer after learning it from the input layer. The model allows the algorithm to provide precise values for imputation while learning from incomplete input.

Therefore, this study proposes a data imputation algorithm using AE to handle missing information and MOA for FS to find an optimal subset of features in health-related data. To the best of our knowledge, there is limited research addressing two key preprocessing methods of end-to-end methodology: (i) missing data imputation with DL, and (ii) FS to obtain relevant and optimal features in health-related datasets. Well-known MOAs with machine learning classifiers have been experimented with. For experimental analysis of various dimensions of data, this study is highly concerned with experimenting the health-related data with low- (small features) and high-dimensional (large features). The main contributions of this paper are as follows:

- Developing an end-to-end methodology, which includes data imputation, FS, and classification, customized for health-related data of diverse sizes and dimensions for producing high accuracy;
- Identifying significant features in health-related data through an exploration and exploitation approach based on MOA;
- Proposing a missing data imputation approach to improve classification performance using machine learning methods, and
- Evaluating the proposed methodology across seven datasets to ensure the robustness of the model.

II. LITERATURE REVIEW

The use of wrapper-MOA to solve the FS problems in medical datasets has shown promising results. In a promising study, Singh and Singh [2] explored a hybrid ensemble-filter wrapper FS approach for medical datasets. They proposed ensemble-filter-based hybrid FS (EFHFS), with fifteen experimented filter and wrapper methods by using four classifiers. For datasets with low dimensions, the EFHFS method typically picks between 9 and 13 features. In medium-dimensional datasets, it selects between 23 and 28 features. For high-dimensional datasets, the range extends from 28 to 36 features chosen by EFHFS. Canayaz [29] have experimented with the Binary Bat Algorithm (BBA), Equilibrium Optimizer (EO), Gravity Search Algorithm (GSA), and Gray Wolf Optimizer (GWO) with support vector machine (SVM), and random forest (RF) for diabetic retinopathy classification. They achieved a high accuracy with a minimum of 250 features. Bashir et al. [30] proposed a Genetic Algorithm (GA) and SVM for microarray and the

Cleveland Heart Disease datasets. An accuracy of 94.45% and 91% is attained on each respective dataset.

In recent years, Talpur et al. [31] have also presented SCSO-KNN to find optimum features from ten benchmark medical datasets. They resulted in an average classification accuracy of 93.96% by selecting 14.2 features. Vommi and Battula [32] proposed ReliefF and Fuzzy Entropy - Binary Enhanced Equilibrium Optimizer (RFE – BEE) for medical dataset classification. The suggested RFE-BEE method employs a minimal number of features, averaging 665.29 across four datasets, particularly suitable for very high-dimensional datasets. Compared to other existing methods, it selects fewer features for three out of the four datasets. Qtaish et al., [46] proposed Binary Memory-based SCSO (BMSCSO) and K-nearest neighbors (KNN) with twenty-one benchmark disease datasets as an experimental study. BMSCSO has incorporated a memory-oriented approach into the updating mechanism of the Sand Cat Swarm Optimizer (SCSO) to utilize and safeguard the optimal solutions more effectively. They obtained an average classification accuracy of 88.62%.

The main contribution of this research lies in the development of a wrapper-MOA approach to achieve optimal FS for various health-related datasets across different dimensionalities. The results presented in previous studies are commendable. However, to the best of our knowledge, there is limited research addressing two key preprocessing methods: (i) missing data imputation with DL, and (ii) FS to obtain relevant and optimal features in health-related datasets. In more detail, the primary contribution of this study is to propose an end-to-end methodology, which includes data imputation, FS, and classification, customized for health-related data of diverse sizes and dimensions for producing high accuracy.

III. MATERIAL AND METHOD

The research methodology of the study is required to describe in detail the experimental procedures. The workflow of this study can be presented in Figure 1, which consisted of: (i) the varying features and samples of raw data (low- and high-dimensional data) experimented, (ii) handling missing data from the dataset, data imputation using AE is required to reduce the significant degree of bias so that analyzing the data more efficiency, (iii) the process of FS using MOA and supervised learning, and (iv) analysis the performance results of MOA and supervised learning.

A. DATA PREPARATION

In this study, the experimented health-related datasets are composed of n as rows and f as columns (structured data). There are two categorical experimented datasets:

1. Low- dimensional datasets: It refers to the number of features (f) are lower than the number of sample sizes (n); $f < n$.
   - With small features and small sample sizes to obtain the best MOA with the machine learning classifiers, three medical datasets are experimented with and explored, i.e., Pima Indians Diabetes, Breast Cancer Wisconsin (Diagnostic), and Chronic Kidney Disease.
   - With small features and a large sample size to generate the missing data imputation model using AE based on the best MOA and classifier, this study explored PhysioNet: MIMIC-III v1.4 [33].
2. High-dimensional datasets: It refers to the number of features that are close to or higher than the number of samples; $f \geq n$. To validate the proposed framework from missing data imputation and FS, we explored the high-dimensional data, such as microarray dataset, i.e., Breast Cancer, Ovarian Cancer and Central Nervous System from Kent Ridge Bio-Medical Dataset Repository. The experimented high-dimensional data has thousands of features with small sample sizes. The detailed seven medical datasets can be seen in Table 1.

B. DATA IMPUTATION

Missing data in health-related data leading missing observations in disease diagnosis. It is a problem often found in health-related datasets and it can degrade the performance of classification tasks. To handle the missing data problems, plausible values are generated to replace the missing values based on two methods, i.e., statistical-based and machine learning-based. For the statistical-based method, the missing observations can be replaced by the most similar values among training data (mean or mode imputation). For machine learning-based methods, this study proposes a DL approach for missing data imputation. Among the DL approach, AE has received much attention for data imputation. AE has the capability of learning from corrupted data, which is a natural extension to the field of missing data [28].

AE is composed of the input, hidden, and output layers which can be divided into encoder (from the input layer to the hidden layer) and decoder (from the hidden layer to the output layer) [34]. The encoder part maps an input vector x to hidden representation y, through a nonlinear transformation $f_\theta(x) = s\left(xW^T + b\right)$ where θ represents the weight matrix W and bias vector b. For the resulting of y representation, it is the mapped back to vector z which has the same shape of x, where z is equal to $g'_\theta(y) = s\left(W'y + b'\right)$ [26]. There are two main steps in how AE is used for missing data imputation. It can be described as follows [35]:

AE was trained on a dataset that contains missing data; for each variable, the average of the known values is used to fill in the missing values. All components of the input vector x_n that contain missing values are masked out when forming the error. Trained AE reconstructs an incomplete input; imputing missing values for an input vector is relatively straightforward. When entered into the encoder, all missing values are replaced with mean values, similar to those in the training process. The imputed values are then shown in the decoder's relevant output.

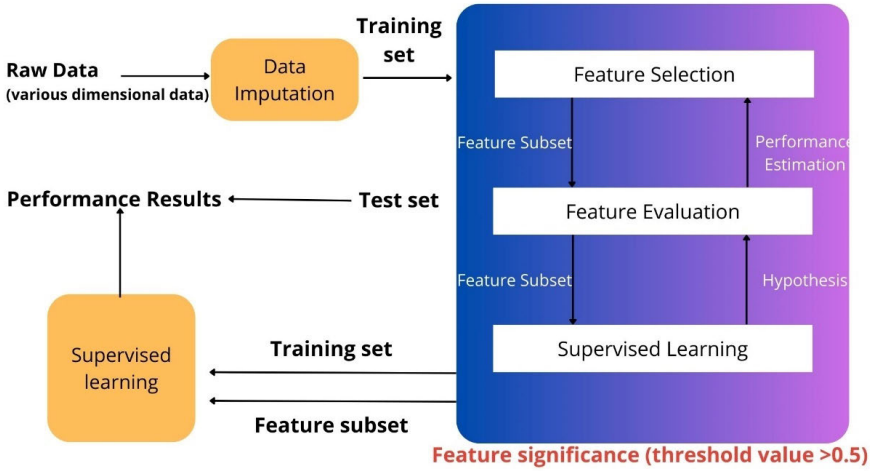


FIGURE 1. The research methodology of FS.

TABLE 1. The experimented seven medical datasets.

Data	Name of Dataset	Abbreviations	Features	Instance	Classes	Missing Values
Low-dimensional	Pima Indians Diabetes	LD1	8	768	2	0
	Breast Cancer Wisconsin (Diagnostic)	LD2	30	569	2	0
	Chronic Kidney Disease	LD3	25	400	2	1,009
	MIMIC-III v1.4	MIMIC-III v1.4	8	50,000	2	32,588,876
High-dimensional	Breast Cancer	HD1	24,481	96	2	0
	Ovarian Cancer	HD2	15,154	252	2	0
	Central Nervous System	HD3	7,129	60	2	0

To generate the missing data imputation model using AE, we conducted the PhysioNet: MIMIC-III Clinical Database v1.4. MIMIC-III v1.4 is a health-related database associated with critical care unit patients, which consists of information such as patient demographics, vital sign measurements, laboratory results, and medications. From such information, we are concerned patient's vital sign measurements for cardiac arrest (CA) classification. MIMIC-III v1.4 has 26 tables, to require the information on patient vital signs, the CHARTEVENTS and ITEMID tables are connected. There are nine features of a patient vital signs, i.e., *heart rate, sysbp, diasbp, meanbp, resprate, tempc, spO2, glucose,* and *label.* Due to the MIMIC-III v1.4 having no label for the classification task, we added the *label* feature for the MIMIC-III v1.4 dataset for CA interpretation based on medical rules. The medical rules of CA and non-CA can be represented in Table 2. From Table 2, the medical rules for each feature are validated to interpret CA and non-CA.

The proposed AE architecture for generating a missing data imputation model can be presented in Table 3. Table 3 lists the used hyperparameters of the proposed AE architecture. We constructed the encoder to decoder parts with 72 – 36 – 18 – 18 – 36 – 72 nodes. The used hyperparameters to generate the AE model for missing data imputation are 50 epochs, 32 batch size, 10^{-3} learning rate, mean squared error (MSE) as loss function, and stochastic gradient descent (SGD) as optimizer.

C. FEATURE SELECTION

The reduction of original features to find an optimal subset of features by preserving correlated information and removing the uncorrelated ones is one of the most challenging tasks in machine learning [36]. Hence, various methods have been proposed to overcome FS problems. Generally, based on the dependencies of any learning method, the methods are classified into two categories; filter and wrapper methods [6], [37]. The filter method is independent of any learning method, so it is suitable for the low computational task [4], [38]. The wrapper method is classifier-dependent as it requires a learning method or a classifier for its processes, which in turn makes it more computationally expensive than the filter approach [2], [39], [40], [41]. However, a wrapper outperformed the filter method to present better results in classification performance [6], [7]. Additionally, the main disadvantage of the filter method is that it ignores feature dependencies, potentially causing features that convey similar information to be selected, resulting in redundancy [7].

The wrapper approach as the FS method draws attention to its excellence in improving classification performance. The wrapper method focused on generated and evaluated subset features. Unfortunately, the generation of subset features is challenging. The exhaustive search is not practically possible due to the number of evaluated features increasing exponentially with the size of the features. It is hard to handle high-dimensional data problems. In addition, sequential

TABLE 2. Medical rules of CA and non-CA interpretation [32].

Vital Sign	CA	Non-CA
heartrate	68–110 beats/minutes	71–93 beats/minutes
sysbp	56–103 mmHg	104–134 mmHg
diasbp	31–56 mmHg	50–71 mmHg
meanbp	39–71 mmHg	67–88 mmHg
resprate	14–24 breaths/min	16–23 breaths/min
spO2	78–99 %	95–99 %

TABLE 3. The hyperparameters of proposed AE architecture.

Layer	Output	Parameter
dense_1 (Dense)	(None, 72)	648
dense_2 (Dense)	(None, 36)	2592
dense_3 (Dense)	(None, 18)	648
dense_4 (Dense)	(None, 18)	324
dense_5 (Dense)	(None, 36)	648
dense_6 (Dense)	(None, 72)	2592

search tends to lead to local optima, because it includes or removes the features sequentially [6]. Hence, addressing the generation of subset features (feature search) through MOA has been proposed in one decade of research [6], [11], [12], [13], [14], [15], [16], [17], [18], [19], [20].

MOA is becoming a modern optimization for the FS domain. Several MOAs have been developed over the past three decades to address various optimization issues [6]. MOA is classified into the following two main categories based on the search process; single and population (multiple) solutions [6]. A single solution is used from the start of the optimization process, and it is updated during the iterations. It could result in trapping into local optima and also only partially exploring the search space. On the other hand, a population solution produces a population of solutions and begins the optimization process. The number of generations or iterations updates the population of solutions and finally, the one with good fitness is selected as the optimal solution. The algorithms perform adequately at avoiding local optima [6], [42], [43].

MOA has two main components; exploration and exploitation [22]. Exploration search or explore the entire search space for new better diverse solutions, while exploitation processes the information found in the local search region [22]. Based on the behavior, MOA can be divided into four categories [6];

- Evolution-based algorithm: search method that imitates the metaphor of biological evolution in the wild and/or the social behavior of different species. The behavior of such species is guided by learning, adaptation, and evolution [19],
- Physics-based algorithm: search method inspired by the rules of physics in the universe [18],
- Swarm intelligence-based algorithm: search method mimics the social behavior of swarms, birds, insects, and animal groups [44],
- Human behavior-related algorithm: search method inspired by human behavior [45].

In this experiment study, ten well-known MOAs among four categories, i.e., Artificial Bee Colony (ABC), Particle Swarm Optimization (PSO), Genetic Algorithm (GA), Cuckoo Search Algorithm (CSA), Harmony Search (HS), Simulated Annealing (SA), Differential Evolution (DE), Teaching Learning Based Optimization (TLO), Biogeography Based Optimization (BBO), dan Firefly Algorithm (FFA) have explored. A detailed description of ten MOAs popular can be seen in Table 4.

D. CLASSIFIER

Supervised learning is a popular machine learning type that involves training a predictive model that includes the target outputs. One of the supervised methods is classification, which means to group the output inside a class. In this study, a wrapper approach (MOA) as a FS method is proposed. The wrapper method interacts with any learning method (classifier) to evaluate the candidate's subset of features. To evaluate MOAs with the learning algorithm, we have experimented with four classifiers, i.e., decision tree (DT), random forest (RF), KNN, and support vector machine (SVM) with the default parameter of each classifier (refer to Table 5). The process of FS for the classification task can be presented in Figure 2. As shown in Figure 2, the training stage of classification is mostly impacted by FS. Following feature generation, FS for classification will first do FS to select a subset of features, and then process the data with the selected features to the learning algorithm (supervised learning).

TABLE 4. The detailed information of ten well-known MOAs.

MOAs	Year	Category	Information
ABC [16]	2005	Swarm-intelligence	Simulates the foraging behavior of honey bees
PSO [43]	1995	Swarm-intelligence	Simulation of the social behavior of birds within a flock
GA [44]	1992	Evolution	Evolutionary concepts
CSA [45]	2009	Swarm-intelligence	The parasitic nature of some cuckoo species, along with Levy flights random walks
HS [15]	2001	Physics	A musician searches for the perfect notes to develop a perfect harmony
SA [46]	1983	Physics	Metal annealing process
DE [47]	1997	Evolution	Inspired by Darwin's theory of evolution
TLO [48]	2011	Human behavior	Teaching and learning in a classroom
BBO [49]	2008	Human behavior	Inspired by the migration of species between habitats.
FFA [17]	2008	Swarm-intelligence	Social behavior of fireflies

TABLE 5. The detailed parameter of classifiers.

DT		RF		KNN		SVM	
Parameters	Value	Parameters	Value	Parameters	Value	Parameters	Value
criterion	Gini	number of estimators	100	number of nearest neighbors (k)	5	regularization parameter (C)	1
splitter	Best	criterion	Gini	weights	uniform	kernel	radial basis function (RBF)
maximum (max)_depth	None	max_depth	None	algorithm	auto	degree	3
minimum (min)_samples_split	2	min_samples_split	2	leaf_size	30	gamma	scale
min_sample_leaf	1	min_sample_leaf	1	power parameter for the Minkowski metric (p)	2	tolerance	1e-3
min_weight_fraction_leaf	0	max_features	sqrt	metric	minkowski	maximum iteration	-1 *(no limit)*
max_features	None	max_leaf_nodes	None	distance	Euclidean	decision_function_shape	ovr *(one vs rest)*
random_state	None	bootstrap	True				
max_leaf_nodes	None						
min_impurity_decrease	0.0						
class_weight	None						
ccp_alpha	0.0						

E. EVALUATION OF FITNESS FUNCTION

Fitness Function evaluates how close a given solution is to the optimum solution of the desired problem. It determines how fit a given solution is in solving the problem. Due to the FS aims to maximize the classification accuracy and minimize the number of selected features, this study calculates a fitness function (Z) in high-dimensional data analysis as follows [6];

$$\max Z = (\xi_1 \text{Accuracy} + \xi_2 \text{Sensitivity} + \xi_3 \text{Specificity} + \xi_4 \text{Precision}) + \xi_5 \frac{\text{Total number of features-Number of selected features}}{\text{(Total number of features)-1}} \quad (1)$$

where ξ_1 - ξ_5 are each criteria coefficient, which ξ_1 - $\xi_4 = (1-\xi_5)/4$ and $\xi_5 \epsilon [0.01]$. The value of ξ_1 is 0.99 [46], [47], so that $\xi_5 \epsilon [0.01]$, wherein $1 - 0.99 = 0.01$.

Accuracy, sensitivity, specificity, and precision values are generated from the evaluation of the confusion matrix (CM). CM is used to measure the performance of a classification model, which consists of four main components, i.e., True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN).

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \quad (2)$$

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP+FN}} \quad (3)$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN+FP}} \quad (4)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \quad (5)$$

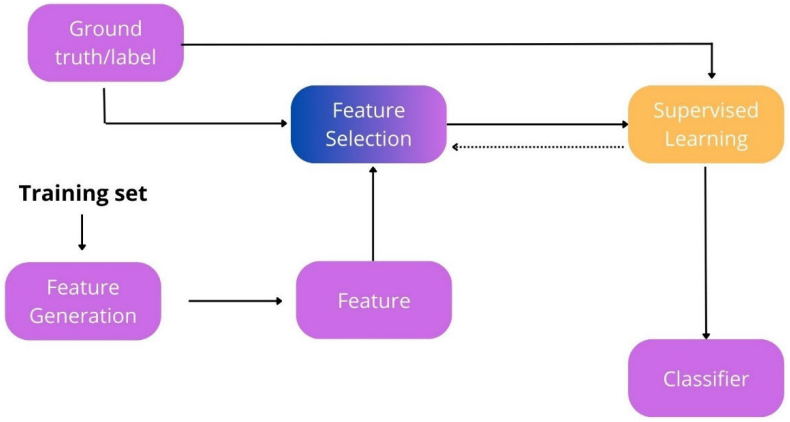


FIGURE 2. FS process for classification task.

TABLE 6. The total number of selected features by ten well-known MOAs that conducted the classifiers.

Classifier	Dataset	Initial features	Number of Selected Features									
			ABC	PSO	GA	CSA	HS	SA	DE	TLO	BBO	FFA
DT	LD1	8	5	5	5	5	4	5	5	5	5	5
	LD2	30	11	15	17	11	15	10	13	10	16	13
	LD3	25	13	15	16	13	18	11	13	3	13	13
RF	LD1	8	6	3	5	4	5	6	3	3	5	5
	LD2	30	10	6	13	10	15	20	11	6	13	10
	LD3	25	11	16	17	9	10	15	11	6	11	9
KNN	LD1	8	3	3	7	3	5	6	5	4	7	5
	LD2	30	8	13	16	16	13	18	10	9	13	3
	LD3	25	12	13	15	10	13	10	13	4	10	5
SVM	LD1	8	4	4	4	4	4	5	4	4	4	4
	LD2	30	13	15	16	16	15	16	15	12	18	10
	LD3	25	17	17	10	14	12	14	10	4	13	12

where TP is the number of correct positive predictions, TN is the number of correct negative predictions, FP is the number of observations that belong to the negative class but are predicted by the model as the positive class, and FN is the number of observations that belong to the positive class but are predicted by the model as the negative class.

F. PLATFORM

The selected features by MOAs become the input features for the classifiers. This study has split each experimented medical dataset into 90% training and the rest for testing set. The classification metrics calculate accuracy, sensitivity, specificity, and precision to evaluate the performance. Ten well-known MOAs experiment with seven health-related datasets on a workstation with one Intel(R) Core(TM) I9-9900K CPU @ 3.60 GHz (16 CPUs) ~3.6GHz, 32GB RAM, and one NVIDIA GeForce RTX 2080 Ti 27GB GPU (11 GB Dedicated, 16 GB Shared) is conducted. All experiments were run on Windows 10 Pro 64 Bit. Python codes in Spyder 4.1.5 with libraries, i.e., VS Code, TensorFlow, NumPy, pandas, scikit-learn, SciPy, matplotlib, seaborn, and mealpy [48] were used.

IV. RESULTS AND DISCUSSION

The performance results of MOA experiments to low- and high-dimensional data can be discussed as follows;

A. LOW-DIMENSIONAL DATA

1) LD1, LD2, LD3 DATASETS

More datasets have a large number of observations than features, e.g., Pima Indians Diabetes (LD1), Breast Cancer Wisconsin (Diagnostic) (LD2), and Chronic Kidney Disease (LD3) dataset. In our experiment for low-dimensional data, ten well-known MOAs (refer to Table 6) conducted to DT, RF, KNN, and SVM are compared. For LD1, LD2, and LD3 datasets, there is an initial number of features. The features

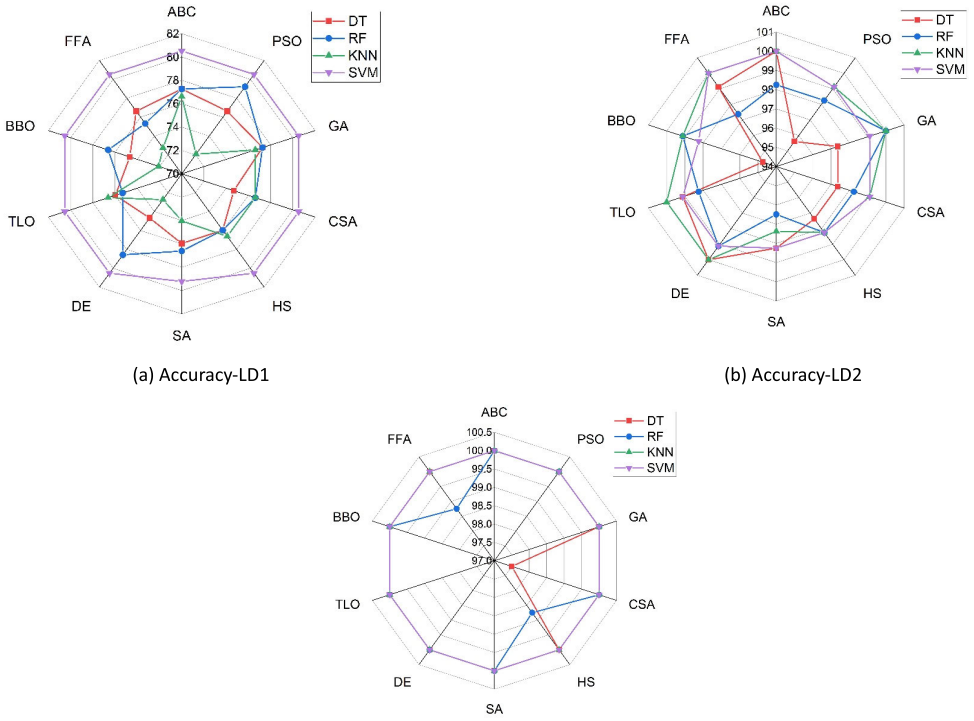


FIGURE 3. Radar chart of ten well-known MOAs based on DT, RF, KNN, and SVM classifiers in low-dimensional datasets.

TABLE 7. The feature significance of TLO+SVM in MIMIC-III v1.4 dataset.

Case of MIMIC-III v1.4	Feature Significance							
	heartrate	*sysbp*	*diasbp*	*meanbp*	*resprate*	*tempc*	*spo2*	*glucose*
Mean Imputation	0.503	0.504	0.000	0.000	0.000	0.494	0.397	0.000
AE Imputation	0.523	0.568	0.543	0.072	0.504	0.586	0.467	0.523

TABLE 8. The performance results of MIMIC-III v1.4 with mean and AE imputation.

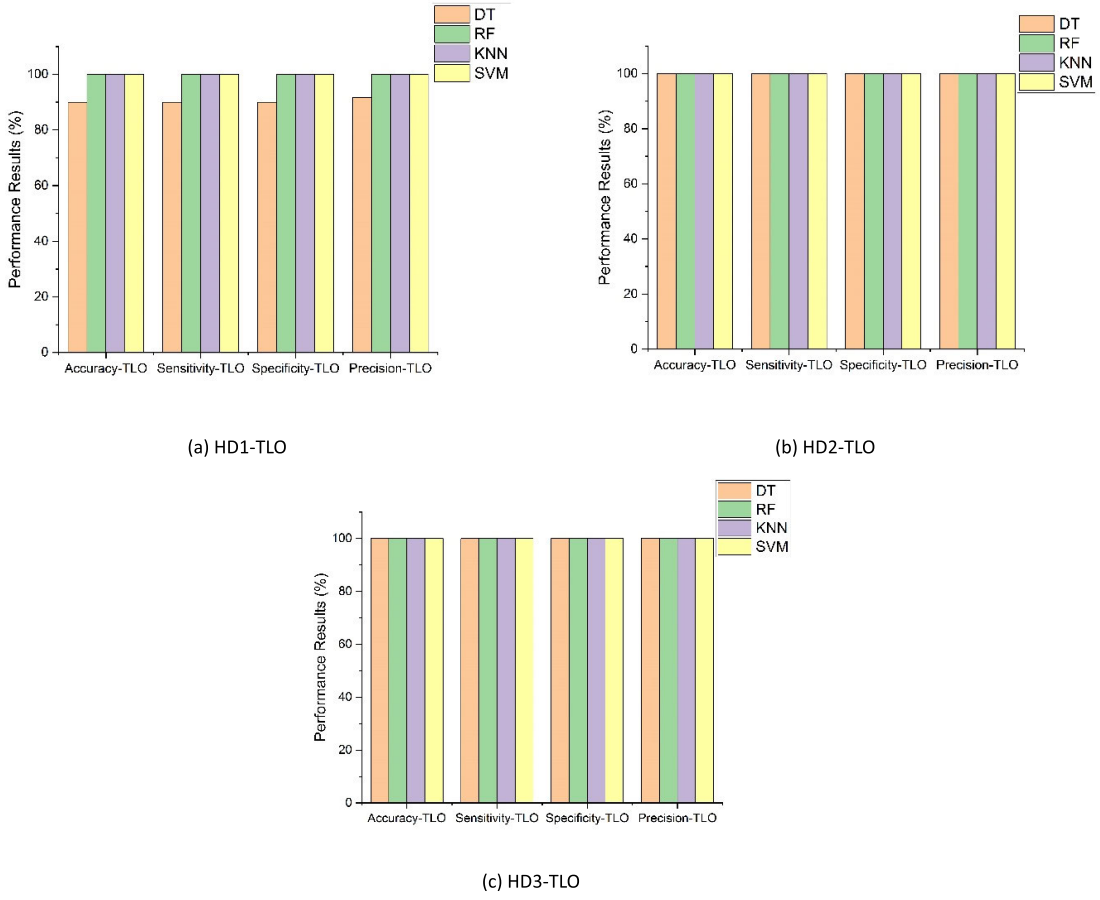
Case of MIMIC-III v1.4	Performance results (%)			
	Accuracy	Sensitivity	Specificity	Precision
Mean Imputation	99.5	73.6	73.6	78.9
AE Imputation	99.8	80.4	80.4	99.9

are selected, and the initial number of features is reduced. The number of selected features by MOA that conducted the classifiers can be listed in Table 6. Table 6 shows the results of MOAs that successfully reduced the features with varying ranges. For the LD1 dataset, MOAs on average selected a maximum of seven features (GA-KNN) and a minimum of three features (PSO-RF, DE-RF, TLO-RF, ABC-KNN, PSO-KNN, and CSA-KNN). For the LD2 dataset, MOAs on average selected a maximum of 20 features (SA-RF) and a minimum of three features (FFA-KNN). For the LD3 dataset, MOAs on average selected a maximum of 18 features (HS-DT) and a minimum of three features (TLO-DT). Among all the MOAs, TLO selected a relatively smaller subset of features in the LD1, LD2, and LD3 datasets. This was because, TLO was designed to work on the philosophy of teaching and learning (teacher and learned phases). The algorithm is based on the effect of the teacher's influence on the learner output in a class. It is easily implemented, has high consistency, and requires less computational memory due to there being no specific parameters that should be needed in TLO.

For the comparison of accuracy for ten well-known MOAs on low-dimensional datasets, the radar chart based on DT, RF, KNN, and SVM classifiers is depicted in Figure 3. The radar chart presents the multivariate observations with an arbitrary

TABLE 9. The execution time and number of features selected by TLO in high-dimensional data based on DT, RF, KNN, and SVM classifiers.

MOA	Dataset	Initial Features	Classifier	Execution Time (s)	Number of selected features
TLO	HD1	24,481	DT	2459.80	12
			RF	3658.24	2
			KNN	221.45	2
			SVM	346.89	2
	HD2	15,154	DT	566.78	2
			RF	3153.24	2
			KNN	151.56	2
			SVM	155.32	2
	HD3	7,129	DT	96.06	2
			RF	1629.51	2
			KNN	59.39	2
			SVM	53.26	2



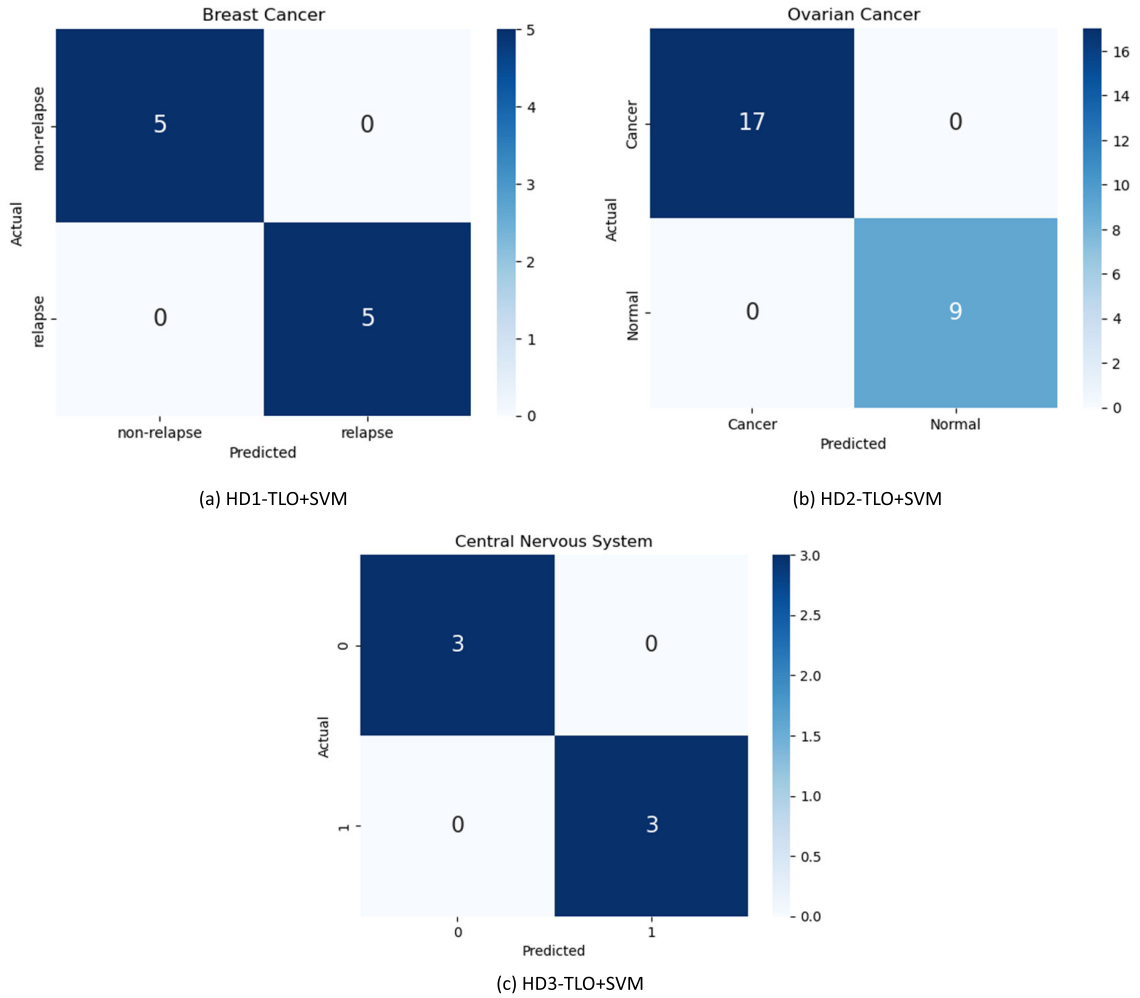
(a) HD1-TLO

(b) HD2-TLO

(c) HD3-TLO

FIGURE 4. The performance results of TLO based on DT, RF, KNN, and SVM classifiers in HD1, HD2, and HD3 datasets.

number of variables. Each classifier (DT, RF, KNN, and SVM) makes a polygon shape that shows performance results based on accuracy. The radar chart having a symmetrical shape and larger area demonstrates a better performance. Among the classifiers, SVM mostly has higher symmetry and maximum area for accuracy in all low-dimensional datasets. In this study, SVM outperformed other experimented classifiers in all low-dimensional datasets. It can be observed that other experimented classifiers (DT, RF, and KNN) have asymmetrical shapes since their performance on LD1 and LD2 wasn't satisfactory enough, but better performance results in LD3.



(a) HD1-TLO+SVM

(b) HD2-TLO+SVM

(c) HD3-TLO+SVM

FIGURE 5. The heatmap CM of TLO+SVM in HD1, HD2, and HD3 datasets.

2) MIMIC-III V1.4

The results of the feature significance of MIMIC-III v1.4 can be listed in Table 7. In this study, the threshold value for feature significance in all dimensional datasets is > 0.5. As a result, we attempt the proposed TLO+SVM with two cases; (i) mean imputation, and (ii) AE imputation. Table 8 presented the performance results of MIMIC-III v1.4 with mean and AE imputation. With mean imputation and two selected features (*heartrate* and *sysbp*), TLO+SVM obtained 99.5% accuracy, 73.6% sensitivity and specificity, and 78.9% precision. In addition, with the AE imputation, the prediction of missing values achieves 0.0167 MSE and 0.129 root mean squared error (RMSE). The results show the average squared difference between the target and predicted values has minim error due to the value achieved below zero. Based on the results of AE imputation, with six selected features (*heartrate*, *sysbp, diasbp, resprate, tempc,* and *glucose*). TLO+SVM has successfully achieved 99.8% accuracy, 80.4% sensitivity and specificity, and 99.9% precision. AE imputation outperformed mean imputation in performance results, though the minimum features were obtained by mean imputation (only two features). TLO+SVM with AE imputation can learn from incomplete data and generate new plausible values for imputation. The extremely imbalanced MIMIC-III v1.4 dataset has affected the classification performance. It brings challenges to feature correlation, class separation, and evaluation. However, TLO+SVM is well-performed; still has high performance in accuracy, sensitivity, specificity, and precision with minimum features.

B. HIGH-DIMENSIONAL DATA

More datasets have larger features than observations, e.g., Breast Cancer (HD1), Ovarian Cancer (HD2), and Central Nervous System (HD3) datasets. HD1, H2, and HD3 are samples of microarray data. Microarray data mostly consists of complex and high-dimensional features, and the number of features is much larger than the number of sample sizes. Moreover, most of these attributes are irrelevant to the classification task. Microarray data hold the expression of features extracted from tissues.

To analyze high-dimensional data, this study is also concerned with conducting TLO on each classifier. High-

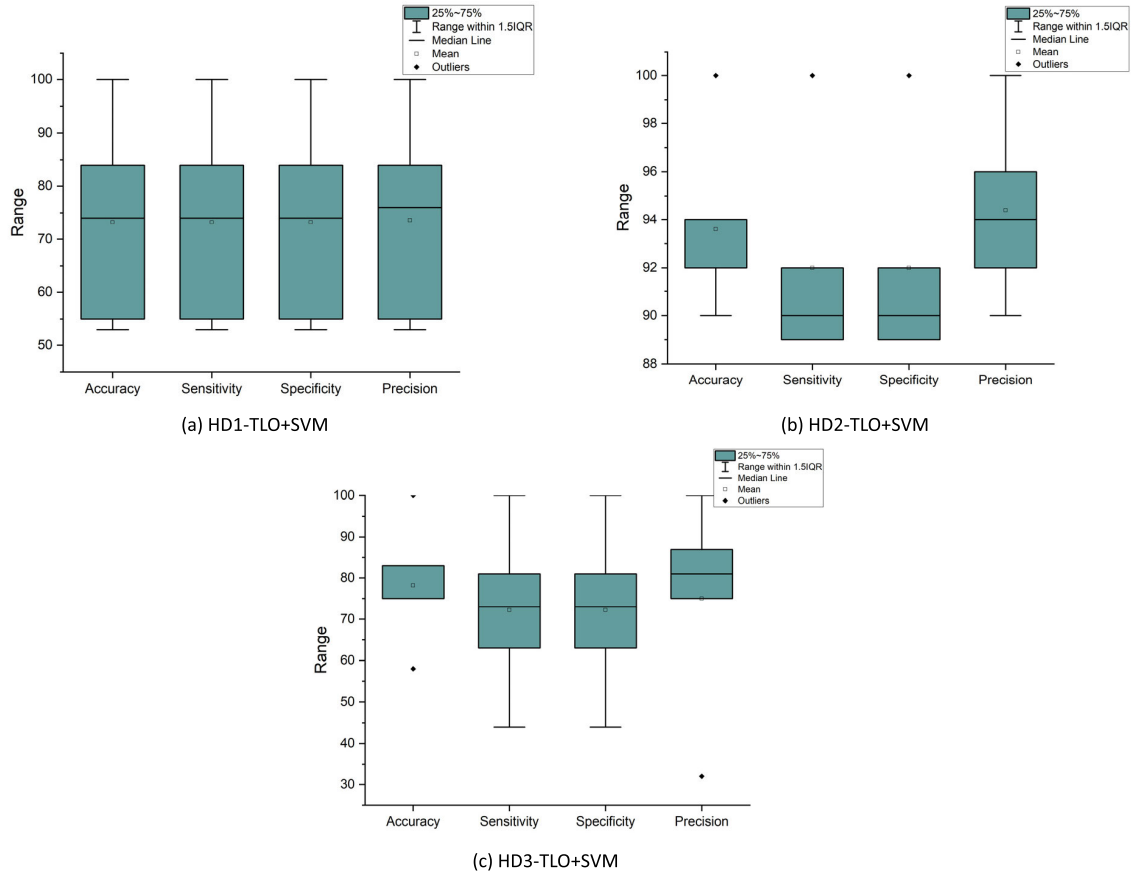


FIGURE 6. The boxplot of TLO+SVM in HD1, HD2, and HD3 datasets with Stratified 5-cross-validation.

dimensional data analysis poses challenges; it is hard to visualize and difficult to identify a single response variable, making standard data exploration and analysis techniques less useful. However, this study was not concerned with the problems above. This study is extensively focused on the number of selected features with a massive of features. Table 9 lists the differences number of initial features and selected features. The number of initial features abruptly decreased, which is only around two, and 12 features selected. Numerous redundant and irrelevant features frequently reduce the classification accuracy of high-dimensional datasets. However, this study has successfully obtained 100% accuracy, sensitivity, specificity, and precision in HD1 with only two selected features (RF, KNN, and SVM) (refer to Figure 4). For HD2 and HD3, all performance metrics had also achieved 100%, with only two selected features in all experimented classifiers. TLO+SVM are still showing outstanding performance in all experimented high-dimensional datasets of this study. As evidence results, the heatmap CM is presented in Figure 5. As presented in Figure 5, TLO+SVM has made no mistakes in its predictions (FP and FN are zero).

To assess the generalization ability of TLO+SVM and provide a more robust estimate of model performance, we have experimented with Stratified K-cross validation (K=5) in HD1, HD2, and HD3 datasets. The high-dimensional datasets have been divided into 5 approximately equal-sized folds. The results of TLO+SVM using K-cross validation can be performed in Figure 6. Figure 6 shows a boxplot of 5-cross validation of accuracy, sensitivity, specificity, and precision in HD1, HD2, and HD3 datasets. Boxplot visualization provides insights into the distribution, variability, and outliers within HD1, HD2, and HD3 datasets. As represented in Figure 6, there are varying results of accuracy, sensitivity, specificity, and precision of HD1, HD2, and HD3 datasets in each fold with Stratified 5-cross-validation. Fold 2, 3, and 4 yields 100% accuracy, sensitivity, specificity, and precision in HD1, HD2, and HD3, respectively.

In this study, we benchmark previous research that provides the challenges of various dimensions of data (refer to Tables 10 and 11). Table 10 listed the benchmarking studies for missing data imputation method DL-based from previous research. Yoon et al. [49] proposed a multi-directional recurrent neural network (M-RNN) to deal with missing data in the MIMIC-III database. They obtained 0.0312 RMSE for the imputation block in their proposed architecture. Qian et al. [50] presented DEep Attention Recurrent Imputation (DEARI) to estimate missing values in heterogeneous multivariate time series. They extracted the MIMIC-III dataset for 21,128 samples with 59 variables.

TABLE 10. Benchmarking studies of missing data imputation DL-based in MIMIC III datasets.

Author	Data imputation method	Performance results	
		MSE	RMSE
Yoon et al., [51]	M-RNN	-	0.0312
Qian et al., [52]	DEARI	0.09165	-
This study	AE	0.0167	0.129

TABLE 11. Benchmarking studies of the FS method in various dimensions of data (the case of health-related datasets).

Authors	FS method	Dataset	Classifier	Number of selected features	Classification Metrics (%)			
					Accuracy	Sensitivity	Specificity	Precision
Singh *et al.* [2]	EFHFS	LD3	SVM	11	100	100	100	-
		HD3		32	86.48	86.3	74.2	-
Gauthama Raman *et al.*, [53]	RSHGT	HD1	RBF-SVM	±10,500	-	83.64	-	84.70
		HD2		±2,700	-	94.92	-	96.58
		HD3		±1,600	-	88.65	-	89.92
This study	A wrapper method with MOA (TLO)	LD2	SVM	12	100	100	100	100
		LD3		4	100	100	100	100
		HD1		2	100	100	100	100
		HD2		2	100	100	100	100
		HD3		2	100	100	100	100

They achieved 0.09165 MSE. As a result, our proposed AE architecture for missing data imputation was well-performed, and it presented good results when compared to other state-of-the-art methods.

As listed in Table 11, there are two concerned parameters; the number of selected features and performance results of the FS with classification metrics. Singh and Singh [2] explored a hybrid ensemble-filter wrapper FS approach for medical datasets. They proposed ensemble-filter-based hybrid FS (EFHFS), with 15 experimented filter and wrapper methods by using Naïve Bayes (NB), SVM, RF, and KNN classifiers. Gauthama Raman et al. [51] proposed Rough Set Theory and Hypergraph (RSHGT)-based FS to identify the informative feature subset in high-dimensional datasets (HD1, H2, and HD3). RSHGT combines the benefit of rough set theory and hypergraph properties to identify the informative feature subset in minimal time. Overall, a wrapper method with TLO+SVM outperformed EFHFS+SVM and RSHGT+RBF has proven its excellence in minimizing a subset of features and maximizing the classification performance. The performance results in this study have high consistency with 100% accuracy, sensitivity, and specificity in the experimented classifiers using TLO+SVM in low- and high-dimensional data. Nevertheless, many open challenges have also been identified that need future research.

The performance results in low- and high-dimensional data look promising, however, there are limitations of this study:

- The experimented datasets are structured data, which consist of rows and columns.
- The analysis is limited to features and sample size in binary classification.

V. CONCLUSION

Feature subset selection is an important technique to find optimal and informative features for machine learning tasks. It poses a challenge due to the number of features and sample size being extensively imbalanced. The problem of FS tends to be related to a high-dimensionality problem, in which the number of features is much larger than the number of sample sizes. It can reduce the classification accuracy due to the sample set of the training set being limited to learning. For experimental analysis of various dimensions of data, this study is highly concerned with experimenting the health-related data due to the relevant attribute in it having a lot of weight and significance for the classification task. Additionally, the health-related data acquired from various medical sources has extremely high feature dimensions. The problem of health-related data is not limited to feature dimensions, due to the high diversity and volume of medical data, the resulting medical records are highly susceptible to missing information in disease diagnosis.

To address those problems, this study explores the FS technique using ten MOAs and four classifiers in low-dimensional datasets and applies the best MOA and

classifier in high-dimensional data from diverse binary medical datasets. As a result, MOA shows its excellence in achieving minimal features, but still outstanding performance with above 90% in low- and high-dimensional data for all performance metrics. TLO + SVM is the best fusion that outperformed other experimented MOAs with DT, RF, and KNN classifiers. A wrapper method with MOA is successfully applied to varying diverse health-related datasets with low- and high-dimensional data. To obtain comprehensive results, this study also proposes a missing data imputation method using AE. With the AE imputation, the prediction of missing values achieves 0.0167 MSE and 0.129 RMSE. The results are well-performed to handle missing values problems in large data.

ACKNOWLEDGMENT

The authors would like to thank the Intelligent System Research Group (ISysRG), Universitas Sriwijaya, Indonesia, for its full support of their research infrastructure.

DATA AVAILABILITY

The datasets generated and/or analyzed during the current study are available in the PhysioNet repository (https://physionet.org/), UC Irvine Machine Learning Repository (https://archive.
ics.uci.edu/), and Kent Ridge Bio-Medical Dataset Repository (https://leo.ugr.es/elvira/DBCRepository/).

CONFLICT OF INTEREST

There is no conflict of interest.

REFERENCES

[1] S. Kaur, Y. Kumar, A. Koul, and S. K. Kamboj, "A systematic review on metaheuristic optimization techniques for feature selections in disease diagnosis: Open issues and challenges," *Arch. Comput. Methods Eng.*, vol. 30, no. 3, pp. 1863–1895, Apr. 2023, doi: 10.1007/s11831-022-09853-1.

[2] N. Singh and P. Singh, "A hybrid ensemble-filter wrapper feature selection approach for medical data classification," *Chemometric Intell. Lab. Syst.*, vol. 217, Oct. 2021, Art. no. 104396, doi: 10.1016/j.chemolab.2021.104396.

[3] F. Masood, J. Masood, H. Zahir, K. Driss, N. Mehmood, and H. Farooq, "Novel approach to evaluate classification algorithms and feature selection filter algorithms using medical data," *J. Comput. Cognit. Eng.*, vol. 2, no. 1, pp. 57–67, May 2022, doi: 10.47852/bonviewjcce2202238.

[4] M. Alirezanejad, R. Enayatifar, H. Motameni, and H. Nematzadeh, "Heuristic filter feature selection methods for medical datasets," *Genomics*, vol. 112, no. 2, pp. 1173–1181, Mar. 2020, doi: 10.1016/j.ygeno.2019.07.002.

[5] B. Remeseiro and V. Bolon-Canedo, "A review of feature selection methods in medical applications," *Comput. Biol. Med.*, vol. 112, Sep. 2019, Art. no. 103375, doi: 10.1016/j.compbiomed.2019.103375.

[6] P. Agrawal, H. F. Abutarboush, T. Ganesh, and A. W. Mohamed, "Metaheuristic algorithms on feature selection: A survey of one decade of research (2009–2019)," *IEEE Access*, vol. 9, pp. 26766–26791, 2021, doi: 10.1109/ACCESS.2021.3056407.

[7] C. Yan, J. Ma, H. Luo, and A. Patel, "Hybrid binary coral reefs optimization algorithm with simulated annealing for feature selection in high-dimensional biomedical datasets," *Chemometric Intell. Lab. Syst.*, vol. 184, pp. 102–111, Jan. 2019, doi: 10.1016/j.chemolab.2018.11.010.

[8] B. Mirza, W. Wang, J. Wang, H. Choi, N. C. Chung, and P. Ping, "Machine learning and integrative analysis of biomedical big data," *Genes*, vol. 10, no. 2, p. 87, Jan. 2019, doi: 10.3390/genes10020087.

[9] R. Alanni, J. Hou, H. Azzawi, and Y. Xiang, "Deep gene selection method to select genes from microarray datasets for cancer classification," *BMC Bioinf.*, vol. 20, no. 1, pp. 1–15, Dec. 2019, doi: 10.1186/s12859-019-3161-2.

[10] A. Jovic, K. Brkic, and N. Bogunovic, "A review of feature selection methods with applications," in *Proc. 38th Int. Conv. Inf. Commun. Technol., Electron. Microelectron. (MIPRO)*, May 2015, pp. 1200–1205, doi: 10.1109/MIPRO.2015.7160458.

[11] P. Priyadharshini and B. S. E. Zoraida, "Bat-inspired metaheuristic convolutional neural network algorithms for CAD-based lung cancer prediction," *J. Appl. Sci. Eng.*, vol. 24, no. 1, pp. 65–71, 2021, doi: 10.6180/jase.202102_24(1).0008.

[12] S. Yildirim, Y. Kaya, and F. Kılıç, "A modified feature selection method based on metaheuristic algorithms for speech emotion recognition," *Appl. Acoust.*, vol. 173, Feb. 2021, Art. no. 107721, doi: 10.1016/j.apacoust.2020.107721.

[13] O. M. Alyasiri, Y.-N. Cheah, A. K. Abasi, and O. M. Al-Janabi, "Wrapper and hybrid feature selection methods using metaheuristic algorithms for English text classification: A systematic review," *IEEE Access*, vol. 10, pp. 39833–39852, 2022, doi: 10.1109/ACCESS.2022.3165814.

[14] P. Agrawal, T. Ganesh, and A. W. Mohamed, "Chaotic gaining sharing knowledge-based optimization algorithm: An improved metaheuristic algorithm for feature selection," *Soft Comput.*, vol. 25, no. 14, pp. 9505–9528, Jul. 2021, doi: 10.1007/s00500-021-05874-3.

[15] Z. Woo Geem, J. Hoon Kim, and G. V. Loganathan, "A new heuristic optimization algorithm: Harmony search," *Simulation*, vol. 76, no. 2, pp. 60–68, Feb. 2001, doi: 10.1177/003754970107600201.

[16] D. Karaboga and B. Basturk, "A powerful and efficient algorithm for numerical function optimization: Artificial bee colony (ABC) algorithm," *J. Global Optim.*, vol. 39, no. 3, pp. 459–471, Oct. 2007, doi: 10.1007/s10898-007-9149-x.

[17] X. S. Yang, "Firefly algorithms for multimodal optimization," in *Proc. Int. Symp. Stochastic Algorithms*, in Lecture Notes in Computer Science: Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics, 2009, pp. 169–178, doi: 10.1007/978-3-642-04944-6_14.

[18] F. A. Hashim, E. H. Houssein, M. S. Mabrouk, W. Al-Atabany, and S. Mirjalili, "Henry gas solubility optimization: A novel physics-based algorithm," *Future Gener. Comput. Syst.*, vol. 101, pp. 646–667, Dec. 2019, doi: 10.1016/j.future.2019.07.015.

[19] E. Elbeltagi, T. Hegazy, and D. Grierson, "Comparison among five evolutionary-based optimization algorithms," *Adv. Eng. Informat.*, vol. 19, no. 1, pp. 43–53, Jan. 2005, doi: 10.1016/j.aei.2005.01.004.

[20] R. C. Eberhart and Y. Shi, "Comparison between genetic algorithms and particle swarm optimization," in *Proc. Int. Conf. Evol. Program.*, 1998, pp. 611–616.

[21] S. Gupta and K. Deep, "A novel random walk grey wolf optimizer," *Swarm Evol. Comput.*, vol. 44, pp. 101–112, Feb. 2019, doi: 10.1016/j.swevo.2018.01.001.

[22] D. Devikanniga, K. Vetrivel, and N. Badrinath, "Review of meta-heuristic optimization based artificial neural networks and its applications," *J. Phys., Conf. Ser.*, vol. 1362, no. 1, Nov. 2019, Art. no. 012074, doi: 10.1088/1742-6596/1362/1/012074.

[23] P. Sharma, F. E. Shamout, V. Abrol, and D. A. Clifton, "Data preprocessing using neural processes for modeling personalized vital-sign time-series data," *IEEE J. Biomed. Health Informat.*, vol. 26, no. 4, pp. 1528–1537, Apr. 2022.

[24] A. Kumar and H. Anjomshoa, "A two-stage model to predict surgical patients' lengths of stay from an electronic patient database," *IEEE J. Biomed. Health Informat.*, vol. 23, no. 2, pp. 848–856, Mar. 2019.

[25] L. Cappelletti, T. Fontana, G. W. Di Donato, L. Di Tucci, E. Casiraghi, and G. Valentini, "Complex data imputation by auto-encoders and convolutional neural networks—A case study on genome gap-filling," *Computers*, vol. 9, no. 2, p. 37, May 2020.

[26] R. C. Pereira, M. S. Santos, P. P. Rodrigues, and P. H. Abreu, "Reviewing autoencoders for missing data imputation: Technical trends, applications and outcomes," *J. Artif. Intell. Res.*, vol. 69, pp. 1255–1285, Dec. 2020.

[27] Y. Haitovsky, "Missing data in regression analysis," *J. Roy. Stat. Soc., B, Methodol.*, vol. 30, no. 1, pp. 67–82, 1968.

[28] A. Sánchez-Morales, J.-L. Sancho-Gómez, J.-A. Martínez-García, and A. R. Figueiras-Vidal, "Improving deep learning performance with missing values via deletion and compensation," *Neural Comput. Appl.*, vol. 32, no. 17, pp. 13233–13244, Sep. 2020.

[29] M. Canayaz, "Classification of diabetic retinopathy with feature selection over deep features using nature-inspired wrapper methods," *Appl. Soft Comput.*, vol. 128, Oct. 2022, Art. no. 109462, doi: 10.1016/j.asoc.2022.109462.
[30] S. Bashir, I. U. Khattak, A. Khan, F. H. Khan, A. Gani, and M. Shiraz, "A novel feature selection method for classification of medical data using filters, wrappers, and embedded approaches," *Complexity*, vol. 2022, pp. 1–12, Aug. 2022, doi: 10.1155/2022/8190814.
[31] N. Talpur, S. J. Abdulkadir, M. H. Hasan, H. Alhussian, and A. Alwadain, "A novel wrapper-based optimization algorithm for the feature selection and classification," *Comput., Mater. Continua*, vol. 74, no. 3, pp. 5799–5820, 2023, doi: 10.32604/cmc.2023.034025.
[32] A. M. Vommi and T. K. Battula, "A hybrid filter-wrapper feature selection using fuzzy KNN based on Bonferroni mean for medical datasets classification: A COVID-19 case study," *Expert Syst. Appl.*, vol. 218, May 2023, Art. no. 119612, doi: 10.1016/j.eswa.2023.119612.
[33] A. E. W. Johnson, T. J. Pollard, L. Shen, L.-W. H. Lehman, M. Feng, M. Ghassemi, B. Moody, P. Szolovits, L. A. Celi, and R. G. Mark, "MIMIC-III, a freely accessible critical care database," *Sci. Data*, vol. 3, no. 1, pp. 1–9, 2016.
[34] S. Nurmaini, A. Darmawahyuni, A. N. Sakti Mukti, M. N. Rachmatullah, F. Firdaus, and B. Tutuko, "Deep learning-based stacked denoising and autoencoder for ECG heartbeat classification," *Electronics*, vol. 9, no. 1, p. 135, Jan. 2020, doi: 10.3390/electronics9010135.
[35] N. Abiri, B. Linse, P. Edén, and M. Ohlsson, "Establishing strong imputation performance of a denoising autoencoder in a wide range of missing data problems," *Neurocomputing*, vol. 365, pp. 137–146, Nov. 2019.
[36] 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, pp. 1–26, Dec. 2020, doi: 10.1186/s40537-020-00327-4.
[37] D. Jain and V. Singh, "Feature selection and classification systems for chronic disease prediction: A review," *Egyptian Informat. J.*, vol. 19, no. 3, pp. 179–189, Nov. 2018, doi: 10.1016/j.eij.2018.03.002.
[38] H. Dong, J. Sun, T. Li, R. Ding, and X. Sun, "A multi-objective algorithm for multi-label filter feature selection problem," *Appl. Intell.*, vol. 50, no. 11, pp. 3748–3774, Nov. 2020, doi: 10.1007/s10489-020-01785-2.
[39] T. M. Le, T. M. Vo, T. N. Pham, and S. V. T. Dao, "A novel wrapper–based feature selection for early diabetes prediction enhanced with a metaheuristic," *IEEE Access*, vol. 9, pp. 7869–7884, 2021, doi: 10.1109/ACCESS.2020.3047942.
[40] W. Liu and J. Wang, "Recursive elimination–election algorithms for wrapper feature selection," *Appl. Soft Comput.*, vol. 113, Dec. 2021, Art. no. 107956, doi: 10.1016/j.asoc.2021.107956.
[41] A. Got, A. Moussaoui, and D. Zouache, "Hybrid filter-wrapper feature selection using whale optimization algorithm: A multi-objective approach," *Expert Syst. Appl.*, vol. 183, Nov. 2021, Art. no. 115312, doi: 10.1016/j.eswa.2021.115312.
[42] G. I. Sayed, A. Tharwat, and A. E. Hassanien, "Chaotic dragonfly algorithm: An improved metaheuristic algorithm for feature selection," *Appl. Intell.*, vol. 49, no. 1, pp. 188–205, Jan. 2019, doi: 10.1007/s10489-018-1261-8.
[43] A. C. Cinar, "A comprehensive comparison of accuracy-based fitness functions of metaheuristics for feature selection," *Soft Comput.*, vol. 27, no. 13, pp. 8931–8958, Jul. 2023, doi: 10.1007/s00500-023-08414-3.
[44] J. Tang, G. Liu, and Q. Pan, "A review on representative swarm intelligence algorithms for solving optimization problems: Applications and trends," *IEEE/CAA J. Autom. Sinica*, vol. 8, no. 10, pp. 1627–1643, Oct. 2021, doi: 10.1109/JAS.2021.1004129.
[45] M. Allam and M. Nandhini, "Optimal feature selection using binary teaching learning based optimization algorithm," *J. King Saud Univ.-Comput. Inf. Sci.*, vol. 34, no. 2, pp. 329–341, Feb. 2022, doi: 10.1016/j.jksuci.2018.12.001.
[46] N. Al-Madi, H. Faris, and S. Mirjalili, "Binary multi-verse optimization algorithm for global optimization and discrete problems," *Int. J. Mach. Learn. Cybern.*, vol. 10, no. 12, pp. 3445–3465, Dec. 2019, doi: 10.1007/s13042-019-00931-8.
[47] E. Emary, H. M. Zawbaa, and A. E. Hassanien, "Binary ant lion approaches for feature selection," *Neurocomputing*, vol. 213, pp. 54–65, Nov. 2016, doi: 10.1016/j.neucom.2016.03.101.
[48] N. Van Thieu and S. Mirjalili, "MEALPY: An open-source library for latest meta-heuristic algorithms in Python," *J. Syst. Archit.*, vol. 139, Jun. 2023, Art. no. 102871, doi: 10.1016/j.sysarc.2023.102871.
[49] J. Yoon, W. R. Zame, and M. van der Schaar, "Estimating missing data in temporal data streams using multi-directional recurrent neural networks," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 5, pp. 1477–1490, May 2019.
[50] L. Qian, Z. Ibrahim, and R. Dobson, "Uncertainty-aware deep attention recurrent neural network for heterogeneous time series imputation," 2024, *arXiv:2401.02258*.
[51] M. R. G. Raman, S. Nivethitha, K. Kannan, and V. S. S. Sriram, "A hybrid approach using rough set theory and hypergraph for feature selection on high-dimensional medical datasets," *Soft Comput.*, vol. 23, no. 23, pp. 12655–12672, Dec. 2019.

ANNISA DARMAWAHYUNI received the master's degree in computer science from Universitas Sriwijaya, Indonesia, in 2019, where she is currently pursuing the Ph.D. degree with the Faculty of Engineering. She is a Lecturer and a Researcher with the Intelligent System Research Group, Faculty of Computer Science, Universitas Sriwijaya. Her research interests include biomedical engineering, deep learning, and machine learning.

SITI NURMAINI (Member, IEEE) received the master's degree in control systems from Institut Teknologi Bandung (ITB), Indonesia, in 1998, and the Ph.D. degree in computer science from Universiti Teknologi Malaysia (UTM), in 2011. She is currently a Professor with the Faculty of Computer Science, Universitas Sriwijaya. Her research interests include biomedical engineering, deep learning, machine learning, image processing, control systems, and robotics.

BAMBANG TUTUKO is currently a Professor with the Faculty of Computer Science, Universitas Sriwijaya. His research interests include control systems, robotic, deep learning, and machine learning.

MUHAMMAD NAUFAL RACHMATULLAH is currently a Lecturer and a Research Assistant with the Intelligent System Research Group, Faculty of Computer Science, Universitas Sriwijaya, Indonesia. His research interests include medical imaging, biomedical signal and engineering, deep learning, and machine learning.

FIRDAUS FIRDAUS is currently is a Lecturer and a Researcher with the Intelligent System Research Group, Faculty of Computer Science, Universitas Sriwijaya, Indonesia. His research interests include text processing, deep learning, and machine learning.

ADE IRIANI SAPITRI is currently is a Lecturer and a Researcher with the Intelligent System Research Group, Faculty of Computer Science, Universitas Sriwijaya, Indonesia. Her research interests include medical imaging, deep learning, and machine learning.

ANGGUN ISLAMI is currently is a Lecturer and a Researcher with the Intelligent System Research Group, Faculty of Computer Science, Universitas Sriwijaya, Indonesia. Her research interests include text processing, deep learning, and machine learning.

JORDAN MARCELINO is currently a postgraduate student with the Faculty of Computer Science, Universitas Sriwijaya, Indonesia. His research interests include machine learning, signal/image processing, and deep learning.

RENDY ISDWANTA is currently a postgraduate student with the Faculty of Computer Science, Universitas Sriwijaya, Indonesia. His research interests include machine learning, signal/image processing, and deep learning.

MUHAMMAD IRFAN KARIM is currently a postgraduate student with the Faculty of Computer Science, Universitas Sriwijaya, Indonesia. His research interests include machine learning and deep learning.