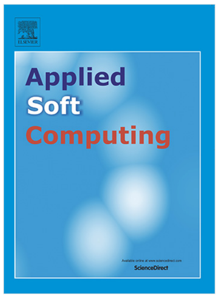
Journal Pre-proof



New feature selection methods based on opposition-based learning and self-adaptive cohort intelligence for predicting patient no-shows

Mohammed Aladeemy, Linda Adwan, Amy Booth, Mohammad T. Khasawneh, Srikanth Poranki

PII: S1568-4946(19)30647-7
DOI: https://doi.org/10.1016/j.asoc.2019.105866
Reference: ASOC 105866

To appear in: *Applied Soft Computing Journal*

Received date : 25 June 2018
Revised date : 7 September 2019
Accepted date : 10 October 2019

Please cite this article as: M. Aladeemy, L. Adwan, A. Booth et al., New feature selection methods based on opposition-based learning and self-adaptive cohort intelligence for predicting patient no-shows, *Applied Soft Computing Journal* (2019), doi: https://doi.org/10.1016/j.asoc.2019.105866.

New feature selection methods based on opposition-based learning and self-adaptive cohort intelligence for predicting patient no-shows

Mohammed Aladeemy[a], Linda Adwan[a], Amy Booth[b], Mohammad T. Khasawneh[a], Srikanth Poranki[b]

[a] Department of Systems Science and Industrial Engineering
State University of New York at Binghamton
Binghamton, New York 13902, USA

[b] United Health Services Hospitals
Binghamton, New York 13903, USA

Abstract

Patient no-shows have significant adverse effects on healthcare systems. Therefore, predicting patients' no-shows is necessary to use their appointment slots effectively. In the literature, filter feature selection methods have been prominently used for patient no-show prediction. However, filter methods are less effective than wrapper methods. To the best of the authors' knowledge, wrapper methods based on metaheuristics have not been applied to patient no-show prediction. This paper presents new wrapper methods based on three proposed variants of Opposition-based Self-Adaptive Cohort Intelligence (OSACI) algorithm. The proposed variants of OSACI are: OSACI-Init, OSACI-Update, and OSACI-Init_Update, which are formed by the integration of Self-Adaptive Cohort Intelligence (SACI) with three Opposition-based Learning (OBL) strategies; namely: OBL initialization, OBL update, and OBL initialization and update, respectively. The performance of the proposed algorithms was examined and compared with that of Genetic Algorithms (GA), Particle Swarm Optimization (PSO), Differential Evolution (DE), and SACI in terms of AUC, sensitivity, specificity, dimensionality reduction, and convergence speed. Patient no-show data of a primary care clinic in upstate New York was used in the numerical experiments. The results showed that the proposed algorithms outperformed the other compared algorithms by achieving higher dimensionality reduction and better convergence speed while achieving comparable AUC, sensitivity, and specificity scores.

Keywords: Opposition-Based Learning; Metaheuristic; Cohort Intelligence; Feature Selection; Patient No-show

1. Introduction

Centers for Medicare and Medicaid Services (CMS) reported an increase in the U.S. healthcare expenditures from \$3 trillion in 2014 (CMS, 2014) to \$3.20 trillion in 2015 (CMS, 2015). This increasing healthcare expenditure is expected to continue and create more pressure on healthcare providers and stakeholders (CMS, 2016). In addition, healthcare services are shifting from inpatient setting to outpatient setting (primary and specialty care clinics) since 2000 according to the Medicare Payment Advisory Commission (MPAC) statistics (MPAC, 2016). Therefore, increasing the operational efficiency of primary care clinics has become crucial because of the increasing demand for primary care services and shrinkage in the supply side (Schwartz, 2012). One of the significant factors that can reduce the operational efficiency of primary care clinics is patients' no-shows. Previous research works showed that patients' no-shows cause significant revenue loss and expensive physician underutilization (Kheirkhah, 2010) and negatively affect continuity of care (Perron et al., 2010; Huang & Zuniga, 2012; Huang & Hanauer, 2014).

In the literature, numerous research works used feature selection methods to identify the significant factors (features) that can be used to predict patient no-show (Lagerlund et al., 2000; Kheirkhah, 2010; Rinder, 2012; Elvira et al., 2017). Feature selection can be considered as an optimization problem that aims to find the optimal feature subset that improves the accuracy of prediction models. Feature selection is a challenging combinatorial problem especially with large datasets (Xue et al., 2014) and it can be stated as in (Katrutsa & Strijov, 2017),

$$\mathcal{S}^* = \underset{\mathcal{S}\subseteq\mathcal{J}}{\operatorname{argmin}}\ Q(\mathcal{S}|\mathbf{X},\mathbf{y}), \tag{1}$$

where $Q: \mathcal{S} \to \mathbb{R}$ is a function that assesses the quality of $\mathcal{S}$ and $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_d] \in \mathbb{R}^{n\times d}$ is a design matrix (i.e., training set) with size of n and a d-dimensional features space, where $\mathbf{x}_j \in \mathbb{R}^n$ is the j^{th} feature of $\mathbf{X}$ and its feature index set is denoted by $\mathcal{J} = \{1, 2, \dots, d\}$.

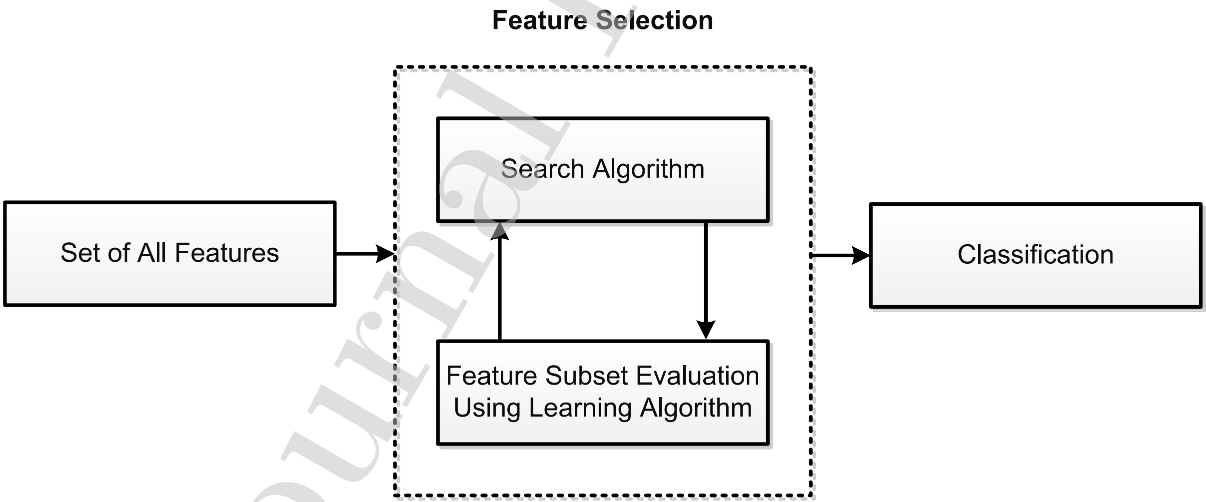
Whereas $\mathbf{y} \in \mathbb{R}^n$ is the class vector. Generally, feature selection methods can be grouped into filter and wrapper methods (Xue et al., 2016) as illustrated in Figure 1.

Filter methods are more efficient than wrapper methods because they do not use prediction models to evaluate the selected feature subsets (Lin et al., 2008). However, wrapper methods are more effective than filter methods because they evaluate every feature subset using prediction models and, thus, they can find optimal feature subsets (Masood et al., 2017). The optimal feature subset, $\mathcal{S}^*$, can be obtained using exhaustive search over all the $2^d - 1$ subsets of $\mathcal{J}$, which makes the problem computationally intractable (NP-hard problem). Therefore, heuristic methods are promising candidates for feature selection as they can find near-to-optimal solutions within a reasonable time (Mafarja & Mirjalili, 2018). In the literature, filter methods, such as statistical analysis (Daggy et al., 2010; Kheirkhah, 2010), have been widely used to predict patient no-shows. Other studies provided a ranking of features importance rather than providing a clear feature subset as in (Al-Mashraie, 2016; Elvira et al., 2017). To the best of the authors' knowledge, wrapper methods based on metaheuristics have not been applied to patient no-show prediction, which is one of the motivations of this research. Metaheuristics are stochastic optimization methods that have been widely applied to feature selection (Xue et al., 2014). Metaheuristics can find promising solutions to optimization problems with incomplete information or limited computation capacity by efficiently exploring and exploiting the search space. Although metaheuristics do not guarantee solutions' optimality, their application becomes inevitable to find near-to-optimal solutions to optimization problems that cannot be solved optimally within reasonable computational efforts (Cortés et al., 2017).

Metaheuristics can be categorized into: nature-inspired algorithms and human-inspired algorithms (Mirjalili & Lewis, 2016). The first category, nature-inspired algorithms, can be grouped into three groups: evolutionary algorithms, swarm-based algorithms, and physics-based algorithms. Evolutionary-based algorithms are bio-inspired algorithms that mimic biological phenomena, in which individuals evolve towards better solutions. Two well-known evolutionary algorithms that are widely used for feature selection are Genetic Algorithms (GA) (Goldberg, 1989) and Differential Evolution (DE) (Storn & Price, 1997). Swarm-based algorithms mimic the cooperative swarm behavior in nature. There are numerous swarm-based algorithms, including Particle Swarm Optimization (PSO) (Eberhart & Kennedy, 1995), Grey Wolf Optimizer (GWO) (Mirjalili et al., 2014), Dragonfly Algorithm (Mirjalili, 2016), and Whale Optimization Algorithm (WOA) (Mirjalili & Lewis, 2016). Recent Swarm-based algorithms applied to feature selection includes: Binary Salp Swarm Algorithm (BSSA) (Faris et al., 2018), Asynchronous Accelerating Salp Swarm Algorithm (Aljarah et al., 2018), Grasshopper Optimization Algorithm (GOA) with Evolutionary Population Dynamics (EPD) (Mafarja et al., 2018a), Binary Dragonfly Algorithm (BDA) (Mafarja et al., 2018b), and WOA (Mafarja & Mirjalili, 2018). The third group of nature-inspired metaheuristics is physics-based algorithms that are inspired from physical rules in the universe (Mirjalili & Lewis, 2016), such as Simulated Annealing (SA) (Kirkpatrick et al., 1983) and Gravitational Search Algorithm (GSA) (Rashedi et al., 2009). The second category of metaheuristics is human-inspired algorithms that mimic human behavior, which includes Tabu Search (TS) (Glover, 1989), Cohort Intelligence (CI) (Kulkarni et al., 2013), Interior Search Algorithm (ISA) (Gandomi, 2014). One of the recent human-based algorithms applied to feature selection is Self-Adaptive Cohort Intelligence (SACI) proposed by Aladeemy et al. (2017). The authors (Aladeemy et al., 2017) showed that SACI outperformed well-known algorithms, including GA and PSO, which are the most two popular metaheuristics applied to feature selection according to an extensive literature survey by Xue et al. (2016). However, SACI's performance can be improved by accelerating its convergence speed to be more suitable for large datasets, such as patient no-show, which is another motivation of this research.

Generally, metaheuristics suffer from different limitations, such as being trapped in local optimum and slow convergence speed (Rojas-Morales et al., 2017). Thus, numerous research works in the literature have proposed different strategies to address to these limitations, such as (Guedria et al., 2016; Amoshahy et al., 2016; Nakisa et al., 2018). Among these strategies is Opposition-based Learning (OBL) (Tizhoosh, 2005a), which has been successfully applied to different metaheuristics to improve their performance as discussed in Section 2.2.

(a) Filter approach



(b) Wrapper approach

Figure 1: Two main feature selection approaches

The primary goal of this research is to develop effective feature selection methods that can identify relevant features to be used to predict patients' no-shows accurately. An effective feature selection method provides insights into the factors that drive the patient no-show behavior. Such insights can inspire preventive interventions to reduce the no-show rate and/or its adverse impact on a clinic's operational performance. This research presents new wrapper methods that use three proposed variants of Opposition-based Self-Adaptive Cohort Intelligence (OSACI) algorithm. The proposed variants of OSACI are formed by integrating SACI with three OBL strategies to achieve better convergence speed. The rationale for proposing wrapper methods for patient no-show prediction is that wrapper methods are more effective than filter methods that are widely used in the patient no-show domain. In addition, the number of original features in this problem domain is relatively not large (usually < 40 features), which mitigates their computational complexity. The major contributions of this research are: first, the proposal of three variants of OSACI: OSACI-Init, OSACI-Update, and OSACI-Init_Update, which are formed by the integration of SACI with three OBL strategies; namely: OBL initialization, OBL update, and OBL initialization and update, respectively. Second, the investigation of the performance of wrapper methods based on the proposed algorithms and other well-known algorithms; namely: GA, PSO, DE, and SACI in patient no-show prediction. To the best of the authors' knowledge, this is the first attempt to apply wrapper methods based on metaheuristics to patient no-show prediction.

The remainder of this paper is organized as follows: Section 2 summarizes the related works on patient no-show prediction and OBL. Section 3 presents a background of SACI, whereas Section 4 presents the data description and proposed methods and an analysis of their computational complexity. Section 5 discusses the experimental results and their practical implications, and Section 6 concludes this paper and presents future work.

2. Related work

This section discusses the relevant research works that addressed the patient no-show problem and related works on OBL.

2.1 Patient no-show review

The patient no-show problem has been studied since the early 1960s. Among the early studies are (Alpert, 1964; Shonick & Klein, 1977; Deyo & Inui, 1980; Dove & Schneider, 1981; Goldman et al., 1982). Huang and Hanauer (2014) defined patient no-show as a patient's failure to arrive for a previously scheduled appointment or appointment cancelation within a very short period of time such that the clinic is unable to fill the appointment slot with another patient. The revenue loss and expensive physician underutilization because of patients' no-shows cannot be neglected (Kheirkhah, 2010). In addition, a patient no-show is considered a major barrier to the continuity and stability of care that interferes with appropriate care of patients (Perron et al., 2010; Huang & Zuniga, 2012; Huang & Hanauer, 2014). For instance, Nuti et. al. (2012) showed that diabetic patients with prior hospital admissions followed by no-shows at their scheduled appointments were at 60% higher risk of hospital readmission. Therefore, identifying the relevant features that drive the patient no-show behavior can help reduce the no-show rate and improve continuity of care.

Statistical analysis has been widely used to identify significant factors that drive the patient no-show behavior. Lee et al. (2005) employed logistic regression to predict patients' no-shows in outpatient clinics in Singapore and included several factors, such as demographic factors, lead time and no-show history and reported model's accuracy of 73%, Area Under the Curve (AUC) of 0.84, sensitivity of 80%, and specificity of 70%. Daggy et al. (2010) used logistic regression to predict the patients' no-shows at a Midwestern Veterans Affairs (VA) Hospital, which achieved an AUC score of 0.82. Kheirkhah (2010) employed logistic regression and Support Vector Machines (SVM) using data from Michael E. DeBakey VA Medical Center where the no-show rate was 18.50%. Stepwise logistic regression selected all 15 original features and achieved an AUC score of 0.72, sensitivity of 97.74%, and specificity of 12.36 % (Kheirkhah, 2010).

Alaeddini et al. (2011) used several features, including date of birth, marital status, ZIP code, clinic, and no-show history, which corresponded to 1,543 patients from a VA medical center. Logistic regression was used with Bayesian inference to predict the patients' no-shows. The predictions from the logistic regression model were considered as initial no-show probability estimates that were used in the Bayesian update as a prior parameter to calculate the posterior no-show probability. Norris et al. (2014) used multinomial logistic regression and DT to predict the patients' no-shows/cancelations. Prior no-show history, age, insurance type, and lead time were found to have the highest association with patients' no-shows based on the split of DT.

Huang and Hanauer (2014) used logistic regression to predict patients' no-shows and concluded that their proposed overbooking strategy based on patient no-show predictions decreased the average waiting time, overtime, and cost (Huang & Hanauer, 2014). Saomorani and Gonfiantini et al. (2015) used Correlation-based Feature Selection (CFS), which is a filter method. GA was used as a search strategy to identify the feature subset that maximizes the correlation between the selected feature subset and the class while minimizing the multicollinearity. Al-Mashraie (2016) used Deep Learning (DL) to predict patients' no-shows at an endocrinology outpatient clinic and considered several factors, including age, no-show history, and portal registration. Elvira et al. (2017) employed gradient boosting algorithm to predict the no-show probability, which achieved AUC of 0.74 (Elvira et al., 2017). Appendix A summarizes the relevant studies in the patient no-show literature.

The following are the major limitations identified in the patient no-show literature. First, the selection of relatively large numbers of features as in (Daggy et al., 2010; Kheirkhah, 2010; Al-Mashraie, 2016), which can reduce the effectiveness of prediction models and make them not easy to operationalize. Second, the use of performance metrics that did not adequately evaluate the effectiveness of the developed prediction models accurately as in (Daggy et al.,2010; Aladeddini et al., 2011; Rinder, 2012; Norris et al., 2014). Third, the selection of features that are not easy to obtain in practice and usually include missing values, such as race, education level, poverty level, and psychiatric diagnosis, as in (Lagerlund et al., 2000; Lehmann et al., 2007; Daggy et al., 2010; Rinder, 2012; Al-Mashraie, 2016). Fourth, the rank of features' importance rather than using a specific subset of selected features as in (Al-Mashraie, 2016; Elvira et al., 2017). The present work addresses these limitations by introducing metaheuristics as artificial intelligence tools to identify the best feature subset for patient no-show prediction.

2.2 Opposition-based learning

OBL was first introduced in Tizhoosh (2005a) to improve the convergence speed and quality of solutions identified by metaheuristics. Its main concept is the simultaneous consideration of a solution and its opposite. The rationale for this concept is that an opposite solution has a higher probability to be closer to the optimal solution than that of another random guess assuming that the original solution is not closer to the optimal solution (Rahnamayan et al., 2008; Ventresca et al., 2010). OBL strategy in 1-dimensional and d-dimensional continuous domain can be presented by Definitions 1-3 as given in Tizhoosh (2005a).

Definition 1. Let $z \in \mathbb{R}$ be a real number on a given interval $z \in [a, b]$. The opposite number $\tilde{z}$ is defined by:

$$\tilde{z} = a + b - z. \quad (2)$$

For a point in a d-dimensional domain, its opposite point can be defined as follows:

Definition 2. Let $\mathbf{z} = (z_1, z_2, \dots, z_d) \in \mathbb{R}^d$ be a d-dimensional point (solution) where $z_j \in [a_j, b_j],\ j = 1,2,\dots,d$. The opposite solution $\tilde{\mathbf{z}}$ is defined by its coordinates $\tilde{z}_1, \tilde{z}_2, \dots, \tilde{z}_d$, i.e.,

$$\tilde{z}_i = a_j + b_j - z_j \qquad j = 1,2,\dots,d. \tag{3}$$

Based on the Definition 2 of an opposite point, the OBL scheme can be defined as follows:

Definition 3. Let $\mathbf{z} = (z_1, z_2, \dots, z_d) \in \mathbb{R}^d$ be a d-dimensional solution and $\boldsymbol{f}(\mathbf{z})$ be a fitness function that evaluates the solution. If the fitness of the opposite solution $\tilde{\mathbf{z}}$, $\boldsymbol{f}(\tilde{\mathbf{z}})$, is better than the fitness of the original solution $\mathbf{z}$, $\boldsymbol{f}(\mathbf{z})$, i.e., $\boldsymbol{f}(\tilde{\mathbf{z}}) \geq \boldsymbol{f}(\mathbf{z})$ (in maximization sense), then $\tilde{\mathbf{z}}$ replaces $\mathbf{z}$. Otherwise, the original solution $\mathbf{z}$ is kept.

There are different types of search using the OBL concept, including Type-I opposition and Type-II opposition. Type-I opposition defines a function to map each solution in the search space to its opposite and continue the search with the higher quality solution. Type-II opposition defines the relationship between a solution and its opposite based on evaluating their qualities. Type-I opposition, adopted in this research, is the most common OBL type used in the literature (Rojas-Morales et al., 2017). In binary domain, the opposite solution can be defined using Type-I opposition concept as follows (Seif & Ahmadi, 2015):

Definition 4. Let $\mathbf{z} = (z_1, z_2, \dots, z_d) \in \{0,1\}^d$ be a d-dimensional vector, where d is the number of features and $z_j \in \{0,1\},\ j = 1,2,\dots,d$. The vector, $\tilde{\mathbf{z}} = (\tilde{z}_1, \tilde{z}_2, \dots, \tilde{z}_d) \in \{0,1\}^d$, is the Type-I opposite of $\mathbf{z}$ that represents the opposite feature subset defined by:

$$\tilde{z}_j = 1 - z_j \qquad j = 1,2,\dots,d. \tag{4}$$

In feature selection the solution is, typically, represented in 1-d vector, where each feature is represented by “1” if selected or “0” otherwise. Based on Definition 4, the opposite of a selected feature subset is the subset of features that were not selected as illustrated in Figure 2. In the continuous domain, if a current solution is far away from the optimal solution, its opposite leads in the opposite direction toward the unknown optimal solution (Mahdavi et al., 2018). Similarly, it is reasonable that if a selected feature subset is far away from the optimal feature subset, its opposite can lead to the unknown optimal feature subset. Ventresca and Tizhoosh (2008) used Hamming distance to demonstrate that the opposite guessing strategy for *d*-dimensional binary space produces the most diverse possible solutions. They also demonstrated the effectiveness of this approach on different benchmark datasets, including 10 common Traveling Salesman Problem (TSP) instances (Ventresca & Tizhoosh, 2008). In addition, the opposite of every feature subset is a unique subset as every point in multidimensional binary space has a unique opposite (Seif & Ahmadi, 2015).

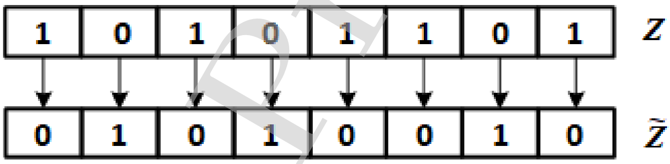


Figure 2: Representation of solution $\mathbf{z}$ and its Type-I opposite $\tilde{\mathbf{z}}$ (“1” indicates feature is selected and “0” indicates feature is not selected)

Several opposition-based metaheuristics have been proposed in the literature based on the different OBL variants, such as generating well-diversified initial solutions for DE (Rahnamayan et al., 2007; Basu, 2016) and PSO (Wang et al., 2009; Si at al., 2014). This is because the convergence speed of metaheuristics to the global optimum is crucially related to the distance between the initial solutions and the unknown optimal solution. Initial solutions are usually randomly generated (using uniform distribution) because of the lack of information regarding an optimal solution in most cases. If the optimal solution is far away from the initial solutions, the convergence speed to global optimum might be slow. Therefore, well-diversified initial solutions can improve the convergence of metaheuristics as demonstrated in the empirical study in (Tizhoosh, 2005b). In addition, OBL has been frequently used in the literature to update the population during the search (Rojas-Morales et al., 2017). OBL was applied to different metaheuristics, such as DE (Ahandani & Alavi-Rad, 2012; Basu, 2016; Ahandani, 2016), Harmony Search (HS) (Banerjee et al., 2014), PSO (Dong et al., 2017), Sine Cosine Algorithms (SCA) (Elaziz et al., 2017), and GWO (Heidari et al., 2019).

3. Background on self-adaptive cohort intelligence

SACI was proposed by Aladeemy et al. (2017) to address the limitations of CI. In CI and SACI, each candidate aims to improve its associated set of qualities (solutions) by controlling its behavior (fitness) and observing other candidates' behaviors. The exploitation (local search) is implemented by allowing each candidate to sample Q solutions from its sampling interval and the solution with the best behavior is selected. On the other hand, exploration (global search) is implemented by allowing each candidate to observe other candidates' behaviors to follow the candidate with the best behavior. This mechanism implies that candidates learn from each other and compete to attain the best behavior. CI suffers from several limitations, such as premature convergence because each candidate follows other candidates during exploration. Also, CI is originally a real-coded method that is not directly applicable to combinatorial problems, e.g., feature selection. In contrast, SACI is directly applicable to the binary domain. In addition, SACI employs a self-adaptive scheme that results in fewer control parameters compared with that of CI, which requires a sampling interval reduction parameter.

In SACI, S initial candidates, $\mathbf{Z} = [\mathbf{z}^1, \mathbf{z}^2, \dots, \mathbf{z}^S]$, are randomly generated and their behaviors are evaluated, i.e., $F(\mathbf{Z}) = [f(\mathbf{z}^1), f(\mathbf{z}^2), \dots, f(\mathbf{z}^S)]$. In each learning attempt (iteration), the candidate with the best behavior is kept, whereas $S - 1$ candidates are selected using tournament selection to undergo mutation with a rate of 1/*S*. Then, each candidate updates its search area by selecting a candidate's behavior to follow using tournament selection. The self-adaptive scheme is employed during the local search to update the mutation rate based on the candidate's behavior. Specifically, when a candidate gives a high-quality solution, the self-adaptive scheme reduces its mutation rate. In contrast, when the candidate gives a low-quality solution, the self-adaptive scheme increases its mutation rate. Thus, the local search is allows each candidate to sample a set of *Q* qualities within the updated search area with a mutation rate of m_l^s,

$$m_l^s = \max\left(\frac{1}{d}, \max\big(F(\mathbf{Z})\big) - f(\mathbf{z}^s)\right). \tag{5}$$

Then each candidate updates its behavior by selecting the quality with the best behavior among the sampled qualities. This procedure is repeated until the cohort saturation condition (for maximization problems) is reached for τ^{max} successive learning attempts,

$$\begin{gathered} |\max(F(\mathbf{Z})^l) - \max(F(\mathbf{Z})^{l-1})| \leq \varepsilon \text{ and,} \\ |\min(F(\mathbf{Z})^l) - \min(F(\mathbf{Z})^{l-1})| \leq \varepsilon \text{ and,} \\ |\max(F(\mathbf{Z})^l) - \min(F(\mathbf{Z})^l)| \leq \varepsilon, \end{gathered} \tag{6}$$

where ε is the convergence tolerance and *l* is the number of the current learning attempt. Cohort saturation implies there is no improvement in the best behavior of the cohort and the difference between candidates' behaviors is less than the convergence tolerance for several successive learning attempts. However, this termination condition is excessively conservative compared to most algorithms in the literature and may cause SACI to run longer without significant improvement in the best cohort behavior. The use of SACI in this research is motivated by its robustness and effectiveness in feature selection as it generally outperformed GA, PSO, and DE according to the numerical results reported in Aladeemy et al. (2017).

However, the efficiency of SACI can be improved by accelerating its convergence speed to make more suitable for large datasets. The following section presents the data used in this research and proposed methods.

4. Research methodology

The research methodology of this paper, illustrated in Figure 3, includes the following: (1) determination of features of interest, (2) data acquisition and preprocessing, (3) application of models used to predict patient no-shows in the literature using all original features and selected feature subset through statistical analysis, (4) development of wrapper methods using GA, PSO, DE, and SACI as search strategy, (5) development of new wrapper methods based on the proposed variants of OSACI; namely: OSACI-Init, OSACI-Update, and OSACI-Init_Update.

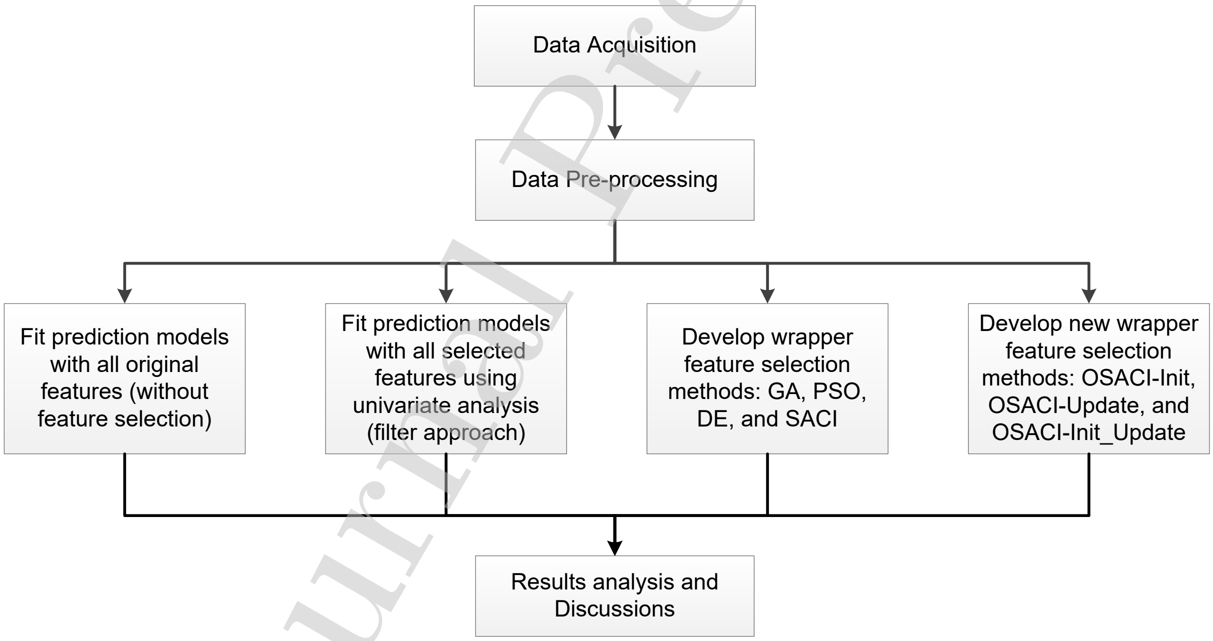


Figure 3: Framework of the research methodology

4.1 Data description and pre-processing

The data used in this research was extracted from the information system of a primary care clinic in upstate New York, USA. The retrieved data spans the period from January 1, 2016, to November 30, 2016, and includes information related to 6,599 appointments scheduled at this clinic. The features of interest were identified according to the prevalent and significant features documented in the patient no-show literature (Section 2.1) and a feature's availability in the information system of the clinic. The data showed that the clinic has a no-show rate of 18.58%. There are 21 features of interest were identified and categorized into four main groups as in (Gonfiantini et al., 2015): clinic and appointment-related, environment-related, patient-related, and appointment outcome-related features. Appendix B provides a description of the 21 original features used in this research.

The dataset was split into a training set (70%) and a testing set (30%); a split ratio that was used in numerous research works in the literature (e.g., Coelho & Neto, 2017; Ferrari et al., 2017; Antunes et al., 2017). This split ratio allows better representation of the data variance in the testing set than that when a smaller testing set is used, which will permit better assessment of the prediction performance of all used methods. Due to class imbalance issue, the training set was balanced by undersampling the majority class (i.e., shows) to achieve a balanced training set. Balancing the training set by undersampling the majority class outperforms other class balancing techniques, e.g., SMOTE (Wallace et al., 2011).

4.2 Opposition-based strategies

This section describes the OBL strategies implemented in the three variants of the proposed algorithm, OSACI. In this research, OBL initialization is implemented by generating $S/2$ random initial candidates (S denotes the number of candidates in this paper) and mapping these solutions to their opposites using Type-I opposition instead of selecting best S solutions from the union of S random initial solutions and their S opposites. This implementation of OBL initialization requires the same number of cost function evaluations as that required by a typical random initialization. This is because half of the initial solutions are composed of random initial solutions and the other half are their opposites as described in Algorithm 1. The diversity of the initial candidates generated in Algorithm 1 is greater than that when all solutions are randomly generated, which is needed to improve the exploration of the search space.

On the other hand, updating current solutions at each learning attempt using Type-I opposition can be implemented as described by Rahnamayan et al. (2006). Specifically, the opposites of all solutions in the current iteration are generated using Type-I opposition, i.e., the total of $2S$ solutions, and then the best S candidates are selected. A major limitation of this implementation is that it increases the computation time because of the evaluation of additional S cost functions at each iteration. In this research, OBL update is implemented at each learning attempt in which $S/2$ of the solutions with the lowest qualities are replaced with their opposite solutions, whereas the other $S/2$ solutions undergo mutation as illustrated in Algorithm 2. One advantage of this strategy is that it does not require additional S/2 cost function evaluations. In addition, this strategy does not introduce additional control parameters, such as the jumping rate as in (Esmailzadeh & Rahnamayan, 2011). The third OBL strategy examined in this research is OBL initialization and update, which combines both OBL initialization and OBL update strategies. Thus, three variants of OSACI are proposed in this paper; namely: OSACI with OBL initialization (OSACI-Init), OSACI with OBL update (OSACI-Update), and OSACI with OBL initialization and update (OSACI-Init_Update), which are described in Section 4.3.

Algorithm 1: OBL Initialization

Input:
S: No. of candidates;
d: No. of original features;
Procedure:
  Generate $S/2$ random candidates, $\mathbf{Z}_r \leftarrow \left[\mathbf{z}_r^1, \mathbf{z}_r^2, \dots, \mathbf{z}_r^{S/2}\right]$;
  Allocate memory for $S/2$ opposite candidates, $\tilde{\mathbf{Z}} \leftarrow \left[\tilde{\mathbf{z}}^1, \tilde{\mathbf{z}}^2, \dots, \tilde{\mathbf{z}}^{S/2}\right]$;
  for $s = 1$ to $S/2$ **do**
    for $j = 1$ to d **do** /* *Create Type-I opposite candidate* */
      $\tilde{z}_j^s \leftarrow 1 - z_{r,j}^s$;
    end for
  end for
  $\mathbf{Z} \leftarrow \mathbf{Z}_r \cup \tilde{\mathbf{Z}}$; /* *Initial candidates* */
Output: Initial candidates, $\mathbf{Z} \leftarrow [\mathbf{z}^1, \mathbf{z}^2, \dots, \mathbf{z}^S]$

Algorithm 2: OBL Update

Input:
$\mathbf{Z}$: Candidates in the current learning attempt;
d: No. of original features;
Procedure:
  Sort the candidates $\mathbf{Z} \leftarrow [\mathbf{z}^1, \mathbf{z}^2, \dots, \mathbf{z}^S]$ in ascending order with respect to behaviors;
  Allocate memory for $S/2$ opposite candidates, $\tilde{\mathbf{Z}} \leftarrow \left[\tilde{\mathbf{z}}^1, \tilde{\mathbf{z}}^2, \dots, \tilde{\mathbf{z}}^{S/2}\right]$;
  Allocate memory for $S/2$ mutant candidates, $\mathbf{Z}_m \leftarrow \left[\mathbf{z}_m^1, \mathbf{z}_m^2, \dots, \mathbf{z}_m^{S/2}\right]$;
  for $s = 1$ to $S/2$ **do**
    for $j = 1$ to d **do** /* *Create Type-I opposite candidate* */
      $\tilde{z}_j^s \leftarrow 1 - z_j^s$;
    end for
  end for
  for $s = (S/2)+1$ to S **do**
    Apply mutation on candidate $\mathbf{z}^s$ to create mutant candidate $\mathbf{z}_m^s$;
  end for
  $\mathbf{Z} \leftarrow \mathbf{Z}_m \cup \tilde{\mathbf{Z}}$**;** /* *New candidates* */
Output: New candidates in current learning attempt, $\mathbf{Z} \leftarrow [\mathbf{z}^1, \mathbf{z}^2, \dots, \mathbf{z}^S]$

4.3 The proposed algorithms

This section presents the proposed variants of OSACI: OSACI-Init, OSACI-Update, and OSACI-Init_Update, which are formed by integrating SACI (Section 3) with the Opposition-based strategies presented in Section 4.2. Similar to SACI, OSACI-Init, OSACI-Update, and OSACI-Init_Update are directly applicable to the binary domain, i.e., directly applicable to the feature selection problem. As illustrated in Algorithm 3 and Figure 4, OSACI-Init is implemented if the initial candidates are generated using Algorithm 1, whereas OSACI-Update is implemented if Algorithm 2 is used with random initial candidates. OSACI-Init_Update is implemented if both algorithm 1 and 2 are implemented. In addition to adopting OBL strategies, the proposed methods employ less conservative stopping criterion than that of SACI (Aladeemy et al., 2017). The stopping criterion of SACI is met only when there is no significant change in the minimum and maximum fitness values of the cohort in addition to the best fitness attained for a specific number of successive learning attempts, τ^{max}. This stopping criterion is less likely to be met before the maximum number of learning attempts, L, is reached. Thus, the stopping criterion used in the proposed algorithms is met if there is no significant improvement in the best cohort behavior for a specific number of successive learning attempts. Based on the classification of opposition-based metaheuristics presented in the recent review by Rojas-Morales et al. (2017), the proposed algorithms belong to the class of metaheuristics that directly apply OBL. The performance of proposed algorithms, OSACI-Init, OSACI-Update, and OSACI-Init_Update, is discussed in Section 5.

Algorithm 3: Opposition-based self-adaptive cohort intelligence

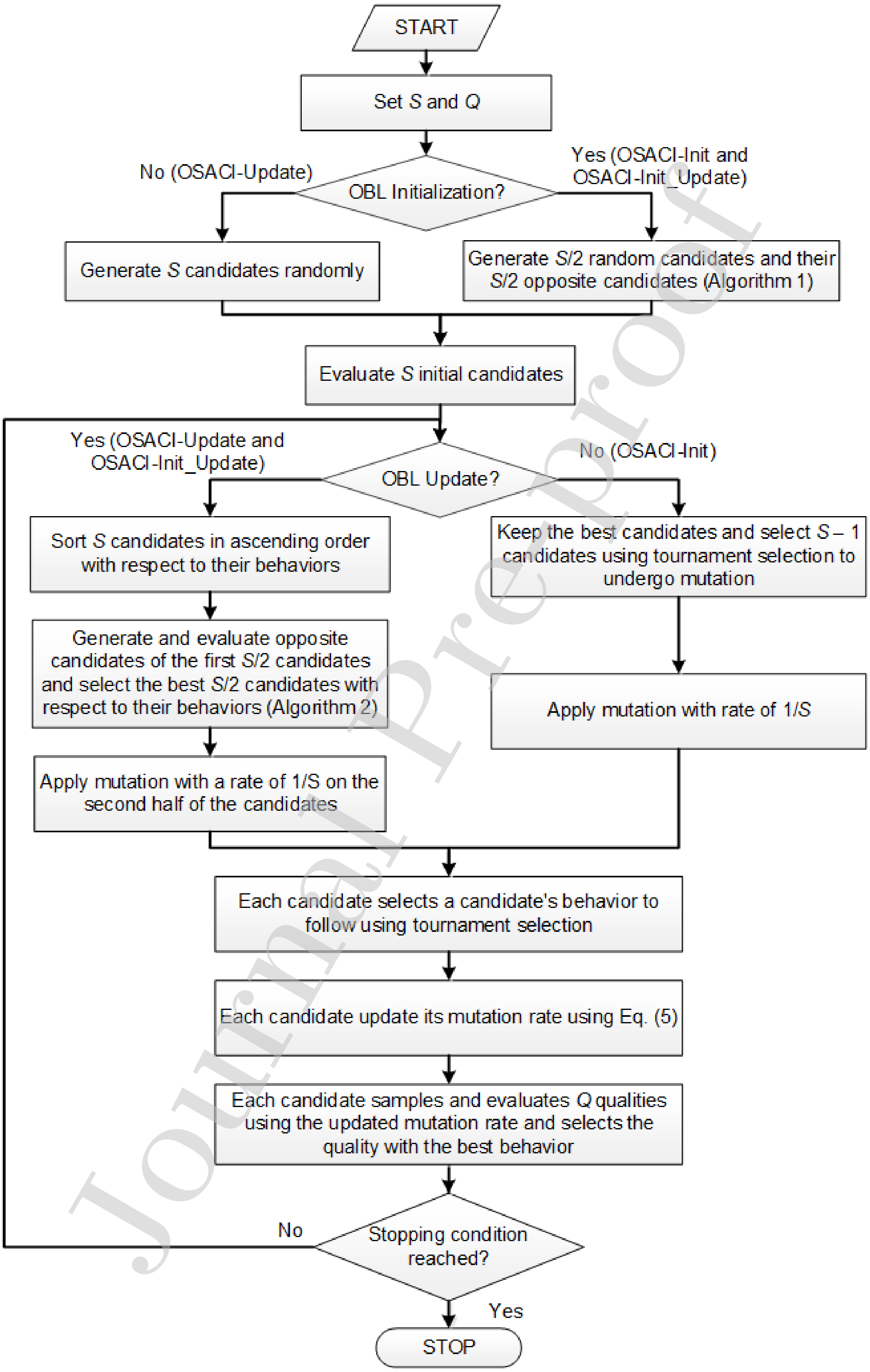
Input:

S: No. of candidates;
Q: No. of quality variations;
L: Max. no. of learning attempts;
ε: Convergence tolerance;
τ^{max}: Max. no. of successive learning attempts for saturation;

Procedure:

  Set $\tau \leftarrow 0$;
  Generate random initial candidates, $\mathbf{Z} \leftarrow [\mathbf{z}^1, \mathbf{z}^2, \ldots, \mathbf{z}^S]$; /* *apply **Algorithm 1** to generate **Z** to use OSACI-Init or OSACI-Init_Update**/
  Evaluate initial candidates, $F(\mathbf{Z}) \leftarrow [f(\mathbf{z}^1), f(\mathbf{z}^2), \ldots, f(\mathbf{z}^S)]$;
  for l = 1 to L **do**
    Keep the best candidate and select $S - 1$ candidates using tournament selection to undergo mutation with rate of $1/S$; /* *apply **Algorithm 2** to use OSACI-Update or OSACI-Init_Update**/
    Evaluate the new candidates;
    for s = 1 to S **do**
      Select a candidate's behavior to follow using tournament selection;
      Update the mutation rate using
        $m_l^s \leftarrow \max\left(\frac{1}{d}, \max\big(F(\mathbf{Z})\big) - f(\mathbf{z}^s)\right)$;
      Sample Q qualities (solutions) using the updated mutation rate, m_l^s;
      Evaluate the sampled qualities, $f(\mathbf{z}_1^s), f(\mathbf{z}_2^s), \ldots, f(\mathbf{z}_Q^s)$;
      Select the quality with the best behavior and update the behavior of candidate s;
    end for
    if $\left|\max\big(F(\mathbf{Z})^l\big) - \max\big(F(\mathbf{Z})^{l-1}\big)\right| \leq \varepsilon$ **then** /* Stopping criterion */
      $\tau \leftarrow \tau + 1$
    end if
    if $\tau = \tau^{max}$ **then**
      break
    end if
  end for

Output: Best feature subset attained by the cohort


START
Set S and Q
OBL Initialization?
No (OSACI-Update)
Yes (OSACI-Init and OSACI-Init_Update)
Generate S candidates randomly
Generate S/2 random candidates and their S/2 opposite candidates (Algorithm 1)
Evaluate S initial candidates
OBL Update?
Yes (OSACI-Update and OSACI-Init_Update)
No (OSACI-Init)
Sort S candidates in ascending order with respect to their behaviors
Generate and evaluate opposite candidates of the first S/2 candidates and select the best S/2 candidates with respect to their behaviors (Algorithm 2)
Apply mutation with a rate of 1/S on the second half of the candidates
Keep the best candidates and select S – 1 candidates using tournament selection to undergo mutation
Apply mutation with rate of 1/S
Each candidate selects a candidate's behavior to follow using tournament selection
Each candidate update its mutation rate using Eq. (5)
Each candidate samples and evaluates Q qualities using the updated mutation rate and selects the quality with the best behavior
Stopping condition reached?
No
Yes
STOP

456

Figure 4: Flowchart of the proposed variants of OSACI algorithm

4.4 Computational complexity analysis

This section discusses the computational complexity of the proposed algorithms, which depends on three input parameters: the number of candidates, *S*, the number of quality variations, *Q*, and the number of original features, *d*. The overall computational complexity of OSACI-Init is

$$\mathcal{O}(\mathrm{OSACI-Init}) = \mathcal{O}(dS) + L \times [\mathcal{O}(d(S-1)) + \mathcal{O}(dSQ)], \quad (7)$$

where L is maximum number of learning attempts (iterations). The computational complexity of SACI is the same as that of OSACI-Init, i.e., $\mathcal{O}(\mathrm{OSACI-Init}) = \mathcal{O}(\mathrm{OSACI})$. Thus, OBL initialization as described in Algorithm 1 does not introduce additional cost function evaluations compared to that of SACI. On other hand, the computational complexity of OSACI-Update and OSACI-Init_Update can be defined as follows:

$$\mathcal{O}(\mathrm{OSACI-Update}) = \mathcal{O}(\mathrm{OSACI-Init\_Update}) = \mathcal{O}(dS) + L \times [\mathcal{O}(dS) + \mathcal{O}(dSQ)]. \quad (8)$$

As demonstrated in Eq. (8), employing OBL in updating the cohort introduces one additional cost function evaluation in each learning attempt compared to that of SACI. However, updating the cohort during the search process using OBL results in higher cohort diversification that can improve the convergence rate. In addition, the computational complexity of the proposed algorithms, Eq. (7) and (8), defines the worst-case scenario or upper bound of running time. Given the fact that metaheuristics are stochastic optimization methods, the upper bound of running time does not necessarily reflect their empirical time complexity because the number of iterations they require to converge to optimal solution is not known in advance. Therefore, convergence curves were used in this research to assess empirical efficiency of all compared algorithms in Section 5.3 similar to previous research works (Ventresca & Tizhoosh, 2008; Nakisa et al., 2018; Mafarja et al., 2019).

5. Experimental results and discussions

This section presents and discusses the numerical experiments conducted in this research. All methods used in the numerical experiments were implemented in R on Windows operating system using a machine with Intel Xeon CPU E5-2699A v4 @ 2.40GHz processor and 256 GB RAM. Section 5.1 presents the fitness function and Section 5.2 presents the parameter settings of all algorithms used in the numerical experiments. Section 5.3 presents the comparative results, and Section 5.4 discusses the practical implications of the major findings in this research.

5.1 Fitness function

The wrapper methods used in this research evaluate every feature subset (solution) based on the AUC score of DT classifier trained using 3×5-fold Cross Validation (CV) and the complexity of the solution (dimensionality reduction). The AUC represents a trade-off between sensitivity (proportion of correctly identified positive instances) and specificity (proportion of correctly identified negative instances) under varying threshold values. In model training, the training set is split into five subsets where four subsets are used for model fitting and the remaining subset is used to evaluate the fitted model. This procedure is repeated five times with each of the five subsets used only once as a validation subset, then the AUC scores are averaged to obtain the training AUC score, $f(\mathbf{z}^s) \in [0,1]$. The following function calculates fitness of the solution,

$$Fitness = \lambda\, f(\mathbf{z}^s) + (1-\lambda)\left(\frac{|\mathcal{J}| - |\mathcal{S}|}{|\mathcal{J}|}\right), \tag{9}$$

where $\lambda \in [0,1]$ is a trade-off factor, whereas $|\mathcal{J}|$ and $|\mathcal{S}|$ are the cardinalities of the original features index set (the number of original features) and selected features index subset (the number of selected features), respectively. The solution structure shown in Figure 2 indicates that every feature subset is represented by a binary vector such that each feature is represented by “1” if selected and “0” otherwise. Thus, $|\mathcal{S}|$ of a given solution is equal to the number of 1s in that solution. The fitness function, Eq. (9), penalizes the solutions with higher complexity, i.e., a higher $|\mathcal{S}|$, to allow the selection of smaller feature subset. The trade-off factor, λ, can be set by the user to determine the importance of each objective, which can be determined experimentally. In this research, λ was set to 0.8 similar to previous research (Huang & Wang, 2006; Aladeemy et al., 2017).

5.2 Parameter settings

In this research, preliminary experiments were conducted to determine the best combination of parameter settings of the compared algorithms. The range of each initial parameter values of GA, PSO, and DE considered in the preliminary experiments (Table 1) is based on previous research (Huang & Wang, 2006; Chen et al., 2015). There are 36 combinations of parameter settings of GA, PSO, and DE that were simulated in the preliminary experiments in which the maximum number of iterations and the population size were set to 20 and 100, respectively. The parameter values that resulted in the highest AUC scores in the preliminary experiments were considered the best parameter values (Table 1), which were later used in the numerical experiments presented in Section 5.3. In GA, roulette selection and single-point crossover operators were implemented in the numerical experiments, whereas the binary versions of PSO (Khanesar et al., 2007) and DE (Chen et al., 2015) were adopted in this research. To compare the convergence speed of all algorithms based on similar number of cost function evaluations, the population/swarm size used in GA, PSO, and DE was set to 100, whereas the number of candidates and quality variations of SACI, OSACI-Init, OSACI-Update, and OSACI-Init_Update were set to 10 and 9, respectively. This parameter setting requires 99 cost function evaluations in each learning attempt in SACI, OSACI-Init, OSACI-Update, OSACI-Init_Update, which is the closest number to the 100 cost function evaluations per iteration used in GA, PSO, and DE. The maximum number of iterations and learning attempts of all compared algorithms was to 50.

Table 1: Initial and best parameter settings

Algorithm	**Parameter**	**Initial values**	**Best value**
GA	Mutation rate	0.01;0.05;0.10;0.15	0.01
	Crossover rate	0.75;0.80;0.85;0.90	0.80
PSO	c_1, c_1	1.60;1.80;2	1.80
	w_1, w_2	0.50;0.60;0.70;0.80	0.60
DE	Mutation rate	0.01;0.05;0.10;0.15	0.05

5.3 Numerical results

The comparative results include the performance metrics on training and testing sets for: (1) patient no-show prediction models using all original features (without feature selection, i.e., baseline), (2) patient no-show prediction models using significant features only determined through statistical analysis, and (3) wrapper methods that use GA, PSO, DE, SACI, OSACI-Init, OSACI-Update, and OSACI-Init_Update as search strategy. DT classifier was used to evaluate every selected feature subset. The prediction models used in the numerical experiments are: Random Forest (RF), AdaBoost.M1, Support Vector Machines (SVM), Naive Bayes (NB), Deep Learning Neural Network (DNN), K-Nearest Neighbors (KNN), Logistic Regression (LR), and Elastic-Net Regularized Generalized Linear Model, which have been widely used in the patient no-show literature. The performance of the compared algorithms were evaluated using average dimensionality reduction, AUC, sensitivity, specificity, and convergence speed over 30 independent runs. The nonparametric Wilcoxon rank sum test at 0.05 significance level was conducted to assess the statistical significance of the results achieved by the proposed algorithms versus the existing algorithms. The AUC, sensitivity, and specificity for training and testing sets when using all original features are shown in Tables 2 and 3, respectively. The differences between AUC scores on the training and testing sets for RF, AdaBoost.M1, NB, DNN, and KNN are relatively high (> 10%), which indicate their poor generalization capabilities when using all original features (without feature selection). In addition, the low sensitivity scores (< 50%) achieved by RF, AdaBoost.M1, SVM, NB, DNN, and KNN indicate their limited capacities in identifying the no-show instances correctly. On the other hand, DT, LR, and LR with regularization achieved good AUC scores: 0.72, 0.70, and 0.70, respectively, and good sensitivity scores: 82.30%, 80.40%, and 79.60%, respectively. Moreover, DT, LR, and LR with regularization achieved reasonable specificity scores (≥ 60%).

Statistical analysis was included in the numerical experiments to identify the significant variables for predicting patients' no-shows, as it is the commonly used feature selection method in the patient no-show literature. Specifically, t-test and Chi-square test were used to analyze the continuous and categorical features, respectively. The identified significant features were used to train all prediction models. The significant features identified using statistical analysis include two continuous features and six categorical features as shown in Table 4, which represents 61.90% dimensionality reduction (eight selected features out of 21 original features). Using the significant features shown in Table 4, the AUC, sensitivity, and specificity for both training and testing sets are demonstrated in Tables 5 and 6, respectively.

Table 2: AUC, sensitivity, and specificity on the training set using 21 original features (baseline)

Prediction Model	**AUC**	**Sensitivity (%)**	**Specificity (%)**
RF	0.80	78.81	66.16
AdaBoost.M1	0.78	74.54	66.20
SVM	0.67	46.43	72.45
NB	0.75	74.31	64.45
DNN	0.72	91.49	27.12
DT	0.77	83.33	64.37
KNN	0.70	74.50	57.04
LR	0.74	79.43	58.01
Elastic-Net Regularized Generalized Linear Model	0.75	81.18	56.81

Table 3: AUC, sensitivity, and specificity on the testing set using 21 original features (baseline)

Prediction Model	**AUC**	**Sensitivity (%)**	**Specificity (%)**
RF	0.66	22.30	46.30
AdaBoost.M1	0.67	20.70	46.30
SVM	0.61	47.70	73.70
NB	0.41	10.40	70.90
DNN	0.61	33	**88.50**
DT	**0.72**	**82.30**	62.60
KNN	0.58	44.70	39.90
LR	0.70	80.40	60.01
Elastic-Net Regularized Generalized Linear Model	0.70	79.60	60

Note: best performance is marked in boldface.

Table 4: Significant features based on statistical analysis

Feature	Type
Indirect waiting time	Continuous
Appointment duration	Continuous
Age group	Categorical
Marital status	Categorical
Adjusted risk score	Categorical
Provider type	Categorical
Visit type	Categorical
Appointment confirmation status	Categorical

The significant features identified using the statistical analysis did not improve the generalization capabilities of RF, AdaBoost.M1, NB, DNN, and KNN in terms of AUC and sensitivity scores (differences between AUC scores on training and testing sets are > 10%). As shown in Table 6, DT outperformed all prediction models using the significant features in terms of AUC although it is 1% less than its AUC score when all original features are used.

Table 5: AUC, sensitivity, and specificity on the training set based on the significant features using statistical analysis

Prediction Model	AUC	Sensitivity (%)	Specificity (%)
RF	0.79	81.60	62.90
AdaBoost.M1	0.75	64.99	69.77
SVM	0.72	69.49	61.77
NB	0.77	76.05	63.25
DNN	0.74	87.29	38.90
DT	0.78	78.46	66.74
KNN	0.73	75.51	60.61
LR	0.76	84.21	55.25
Elastic-Net Regularized Generalized Linear Model	0.75	83.51	54.91

Table 6: AUC, sensitivity, and specificity on the testing set based on the significant features using statistical analysis

Prediction Model	AUC	Sensitivity (%)	Specificity (%)	Dimensionality Reduction (%)
RF	0.61	38.40	39	62
AdaBoost.M1	0.65	23.40	46	62
SVM	0.56	21.80	**90.50**	62
NB	0.40	9.80	69.20	62
DNN	0.56	21.80	90.50	62
DT	**0.71**	76.80	65.70	62
KNN	0.62	33	43.30	62
LR	0.69	80.70	56.50	62
Elastic-Net Regularized Generalized Linear Model	0.69	**80.90**	56.60	62

Note: best performance is marked in boldface.

The AUC, sensitivity, and specificity on the testing set in addition to the dimensionality reduction based on 30 independent runs for all wrapper methods are summarized in Figure 5 and Table 7. The results showed that all wrapper methods led to higher dimensionality reduction compared with that achieved through statistical analysis. For instance, the lowest dimensionality reduction achieved by the wrapper methods is 12.86% higher than that achieved through statistical analysis. In addition, the performance of wrapper methods in terms of AUC and sensitivity scores is comparable to or better than that when using all original features or features selected by statistical analysis. This indicates the effectiveness of the wrapper methods for patient no-show prediction.

The average dimensionality reduction achieved by OSACI-Init, OSACI-Update, and OSACI-Init_Update is higher than those achieved by GA, PSO, and DE and this difference is statistically significant (p-value < 0.05) as shown in Figure 5. Moreover, OSACI-Init, OSACI-Update, and OSACI-Init_Update converged to the same solution (same two selected features) in all 30 independent runs. The features selected by OSACI-Init, OSACI-Update, and OSACI-Init_Update are: appointment indirect waiting time and appointment confirmation status, i.e., 90.48% dimensionality reduction. In addition, SACI converged to the same solution obtained by OSACI-Init, OSACI-Update, and OSACI-Init_Update in 29 out of 30 independent runs. DT induced using the two selected features by the proposed algorithms achieved comparable performance to that when using all 21 original features or the significant features identified by statistical analysis (eight selected features) in terms of the AUC and sensitivity scores. These results demonstrate that the proposed methods can provide promising performance for patient no-show prediction.

In terms of the average AUC and sensitivity scores, OSACI-Init, OSACI-Update, and OSACI-Init_Update achieved comparable performance to those of GA, PSO, DE, and SACI as shown in Table 7, and the p-values in Figure 5 indicate that the differences in these scores are not statistically significant (p-value ≥ 0.05). In addition, the performance metrics achieved by OSACI-Init, OSACI-Update, OSACI-Init_Update, and SACI over the 30 independent runs have the lowest variability as shown in Figure 5 and Table 7. For instance, PSO and OSACI-Init_Update achieved 81.27% and 81.16% of sensitivity, respectively. However, the standard deviation of the sensitivity achieved by OSACI-Init_Update is 0.47% compared to that achieved by PSO (1.90%). Additionally, DE achieved the highest specificity (proportion of the correctly identified shows) of 64.05%, and lowest sensitivity (proportion of the correctly identified no-shows) of 79.09%. However, the cost of misclassifying a no-show case is generally higher than that of misclassifying a show case. In terms of average convergence speed, OSACI-Init_Update has the highest convergence speed followed by OSACI-Init and OSACI-Update as shown in Figure 6. On the other hand, SACI required more learning attempts to converge compared with

OSACI-Init, OSACI-Update, and OSACI-Init_Update. This indicates that the proposed algorithms have better convergence speed than that of SACI. In general, the results demonstrated that OBL strategies, described in Section 4.2, can effectively diversify the cohort and improve the exploration capability to find promising solutions. Thus, the proposed feature selection methods can result in good prediction performance with fewer features and a higher convergence speed.

Table 7: Average (Avg.) and standard deviation (Std.) of AUC, sensitivity, and specificity on the testing set, and dimensionality reduction based on 30 independent runs

	AUC		**Sensitivity (%)**		**Specificity (%)**		**Dimensionality Reduction (%)**	
	Avg.	Std.	Avg.	Std.	Avg.	Std.	Avg.	Std.
DE	0.72	0.01	79.09	4.04	**64.05**	2.29	74.76	4.36
GA	0.72	0.01	80.87	1.31	62.14	1.09	87.94	3.69
PSO	0.72	0.01	**81.27**	1.90	61.68	1.36	87.14	3.78
SACI	**0.72**	**0**	81.03	0.53	61.79	0.15	90.31	0.87
OSACI-Init	0.71	0	80.97	0.28	61.79	0.06	**90.48**	**0**
OSACI-Update	0.71	0	80.97	0.28	61.79	**0.05**	**90.48**	**0**
OSACI-Init_Update	**0.72**	**0**	81.16	0.47	61.75	0.09	**90.48**	**0**

Note: best performance is marked in boldface.

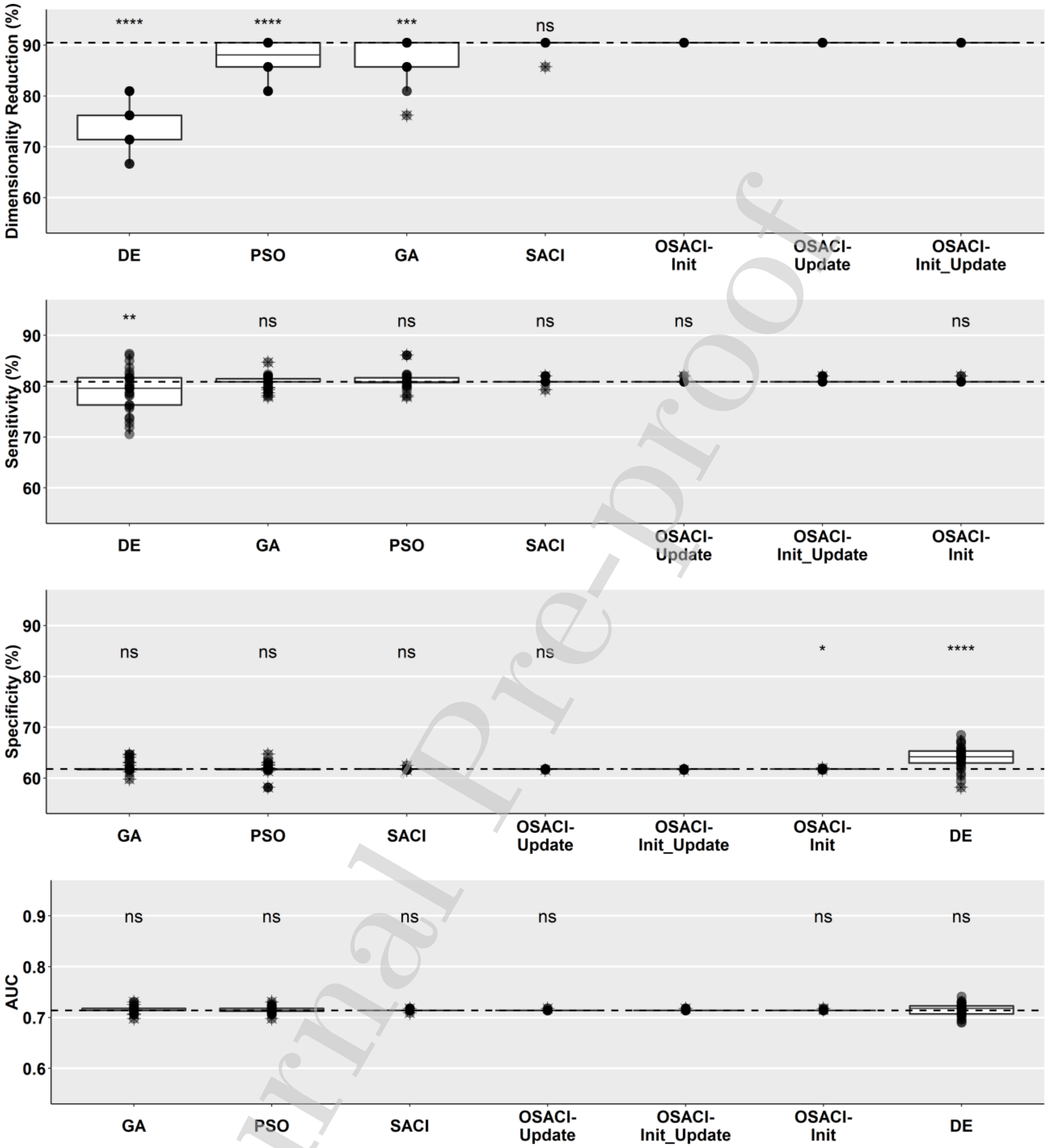


Figure 5: AUC, sensitivity, and specificity on the testing set and dimensionality reduction based on 30 independent runs with p-values of Wilcoxon test of OSACI-Init_Update versus other algorithms (**Notes**: (1) Horizontal dashed line represents the median of the performance metric of OSACI-Init_Update, (2) ns: p-value > 0.05, *: p-value <= 0.05, **: p-value <= 0.01, ***: p-value <= 0.001, ****: p-value <= 0.0001).

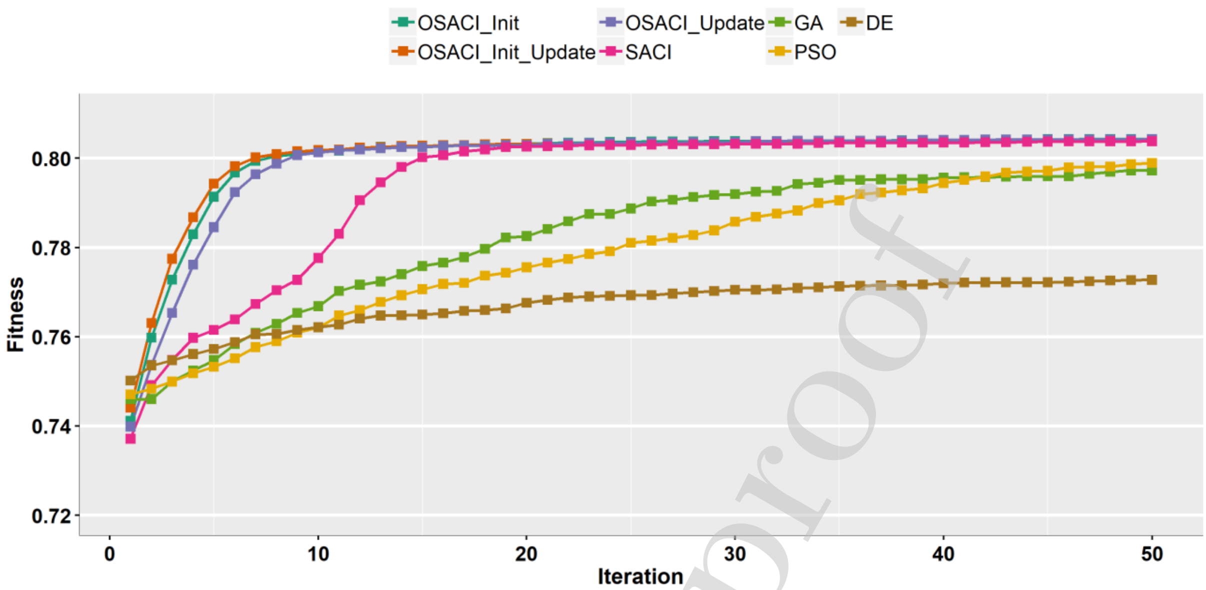


Figure 6: Average convergence curves of GA, PSO, DE, SACI, OSACI-Init, OSACI-Update, and OSACI-Init_Update

5.4 Practical implications and discussions

The simplicity of the solution achieved by the proposed methods, as illustrated in Figure 7, makes it easier to understand patient no-show behavior in the considered clinic in comparison with DT when using 21 original features or when using significant features identified by statistical analysis (eight selected features). A complex model is less reliable as it most likely overfits the data; therefore, it may exhibit poor generalization capability (Sung & Lu, 2013). OSACI with OBL initialization achieved good AUC score with only two features; namely: appointment indirect waiting time and appointment confirmation status. This finding can be of prime interest to managers of primary care clinics. This is because those two features can be easily retrieved from information systems without requiring calculations (e.g., converting patient's ZIP code to distance to the clinic) or using estimations from external sources (e.g., weather conditions). Another interesting finding is that the selected two features by the proposed methods can generally be controlled by primary care clinics. Therefore, the management of the considered clinic can benefit from this finding when planning preventive interventions to reduce its no-show rate and/or its adverse effect on operational performance.

The tree shown in Figure 7 can be interpreted as follows: the no-show probability of a patient who confirmed or did not confirm his or her appointment is 0.02 (very low no-show probability). The discussions with collaborators in the considered clinic helped to clarify and better understand these results. Having an appointment confirmation status as “Confirmed” or “Unconfirmed” means that the patient was contacted using an automatic appointment reminder phone call and the patient either confirmed the appointment or just hung up after hearing the reminder message. However, in both cases the patient was reminded about the appointment a couple of days ahead. This reveals the importance of relying on the automatic appointment reminder phone call. On the other hand, a patient whose confirmation status is “Cancelled Doctor” (cancelled by the doctor and rescheduled, but the patient was not contacted again for a reminder) or “Pre-Reg”, i.e., pre-registration by a real person for appointment reminder, has a show probability of 0.41 and no-show probability of 0.59. For these types of appointments, if the appointment indirect waiting time is < 1.5 days, then the no-show probability is low (0.28) and show probability becomes high (0.72). In contrast, if the indirect waiting time is ≥ 1.5 days then the no-show probability becomes higher (0.66); therefore, lowering the indirect waiting time can reduce the no-show rate.

The lead time feature has been reported to be significant for patient no-show prediction in some previous research works, including (Lee et al., 2005; Daggy et al., 2010; Norris et al., 2014; Gonfiantini et al., 2017). However, the significance of the confirmation type, especially the confirmation through pre-registration in predicting patient no-shows has not been reported in the literature. The results suggest that there is an issue in the pre-registration process in the considered clinic that needs further investigation. Finally, the proposed methods could be adapted to be effective in other healthcare settings, including specialty care clinics, as search strategies based on metaheuristics have been demonstrated to be effective in various applications.

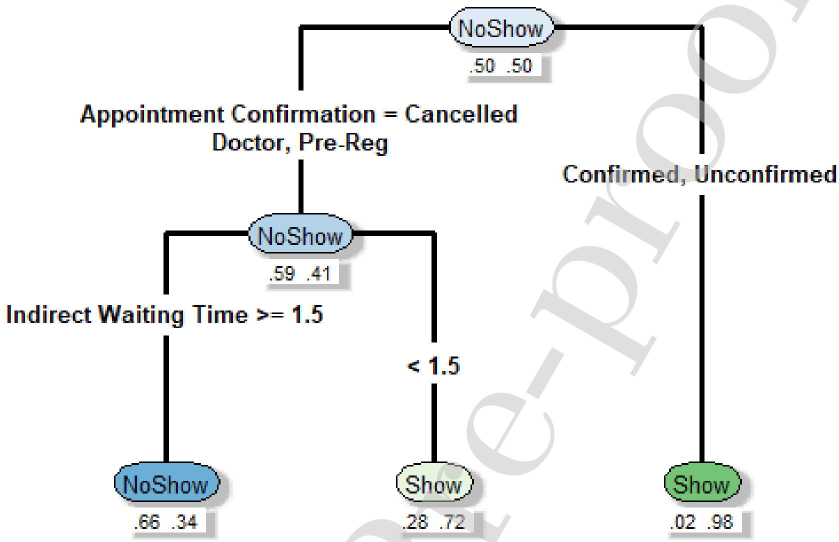


Figure 7: DT induced using two features selected by the proposed methods

6. Conclusions and future work

In this research, three proposed variants of OSACI algorithm were proposed for feature selection; namely: OSACI-Init, OSACI-Update, and OSACI-Init_Update. The proposed algorithms are based on integrating SACI with three OBL strategies, which are: OBL initialization, OBL update, and OBL initialization and update. The proposed algorithms were examined and compared with GA, PSO, DE, and SACI using patient no-show data in terms of AUC, sensitivity, specificity, dimensionality reduction, and convergence speed. The results showed that wrapper methods based on metaheuristics achieved higher dimensionality reduction than that achieved using statistical analysis; a widely used feature selection approach in the patient no-show literature. Additionally, the proposed algorithms outperformed GA, PSO, and

DE by achieving 90.48% dimensionality reduction over 30 independent runs. While achieving higher dimensionality reduction, the proposed algorithms achieved comparable performance to that of GA, PSO, and DE in terms of AUC, sensitivity, and specificity. In terms of convergence speed, OSACI-Init, OSACI-Update, and OSACI-Init_Update achieved the highest convergence speed compared to the other algorithms, including SACI. Additionally, the results revealed that OSACI-Init_Update has slightly better convergence speed compared to OSACI-Init and OSACI-Update. This indicates that employing OBL initialization and update could result in a better convergence than employing OBL initialization or OBL update, which is advantageous when applying wrapper methods to large patient no-show datasets (e.g., > 500,000 appointments). Based on these results, it can be concluded that the proposed feature selection methods can deliver promising performance in identifying feature subsets for patient no-show prediction that can be beneficial in developing preventive interventions to reduce the patient no-show rate or its adverse impact.

In future work, the impact of including other features can be investigated, such as direct waiting time (patients' waiting time at the clinic) and patient-to-provider ratio. In addition, the performance of the proposed variants of OSACI algorithm can be investigated on other datasets and problem domains. Furthermore, it is interesting to examine the impact of employing other OBL types, such as Generalized OBL, Center-based Sampling, and Quasi-Reflection OBL, for initialization and/or cohort update on the convergence speed, especially with highly-dimensional datasets.

Appendix A: Summary of the patient no-show related work

Author(s)/Year	Objective(s)	Sample Size	Methodology	Features	Major Results/Findings
Shonick and Klein (1977)	Reduce adverse impact on the no-show rate	1,178 appointments	Statistical methods such as Chi-square test	Age, gender, ethnicity, income, family size, total number of previous visits, referral source, appointment-breaking record	Significant factors: age and gender
Dove and Schneider (1981)	No-show prediction at day level	1,333 patients	Statistical analysis followed by decision tree	Race, age, residence place, service-connected disability, appointment interval, visit type, previous appointment-keeping record	Significant factors: appointment interval, age, travel distance and no-show history
Goldman et al. (1982)	Patients' no-show prediction	1,181 appointments / 376 patients	Multivariate LR analysis	Demographics, medical characteristic, satisfaction, waiting time at the center, time to arrive the center, no-show history, time since the patient had been using the center and the same doctor	Analysis showed that the no-show rate was correlated with race, age, percent of previously kept appointments and presence of psychosocial problems
Lagerlund et al. (2000)	Examine the impact of the beliefs, attitudes, awareness and knowledge about the treatment on the no-shows	949 patients	Univariate analysis followed by multivariate analysis	Sociodemographic factors, breast disease, health behavior, beliefs and knowledge about breast cancer, trust in the system, social influence, attitudes toward distribution of health-related information, self-efficacy, self-reported details for nonattendance	Higher no-shows for: women who perceived emotional barriers, who worried less about breast cancer, who have lower perception of mammogram benefits, who were not advised by a provider to participate, who have poor faith in healthcare and poor knowledge about breast cancer

Moore et al. (2001)	Investigated no-show impact by measuring two metrics: time and money lost	4,055 appointments	Statistical analysis	Appointment outcome, duration of the appointment or time allocated, visit charges, time of the visit, date, patients' demographics, provider type, ZIP codes	No-show rate was 24.4% per day. Unutilized appointment slots due to no-shows and cancellations were 31.1% of the scheduled appointments Over 1 FY revenue shortfall ranged from 3% to 14%
Lee et al. (2005)	Develop a patient no-show prediction model	22,864 patients	Univariate analysis and logistic regression	Appointment lead time, no-show history, distance from hospital, hospital admission, provision of phone number, department type, gender, hospital admission and phone number provision	Prediction model performance: Accuracy: 73% Specificity: 70% Sensitivity: 80% AUC: 0.84
Johnson et al. (2007)	Investigated what strategies are implemented at family residency practices to prevent or control patients' no-shows	141 practices	Survey questionnaire	Not applicable	To prevent no-shows, practices used: reminders, provide education and support to patients and used open-access. To control no-shows practices used: overbooking, encourage walk-ins and work-ins
Lehmann et al. (2007)	Evaluate no-show patients and their appointment features in order to be able to design intervention methods to reduce the no-show rates	1,296 patients	Compared both no-show and control patients' characteristics to find significant features using statistical tests	Patient related and appointment related information, residency status, language, coverage, family physician, expected medical consequences for a no-show, covert addiction	Significant factors: age, season of birth, citizenship, language, appointment type

Glowacka et al. (2009)	Patients' no-show prediction and use this information to improve patients' scheduling	Not mentioned	Associate rule mining (ARM) in addition to simulation-based optimization	Visit data, demographic information, medical history, family history	Using ARM they were able to define sets of rules that improved the clinic performance
Daggy et al. (2010)	Patients' no-show prediction and use prediction information to develop appointment schedules that optimize clinic overtime costs, utilization and patient waiting time	32,394 appointments / 5,446 patients	LR combined with optimal sequential scheduling	Patient related and appointment related information , previous hospital admissions, number of previous scheduled visits, clinical conditions, days since last scheduled visit	Model performance: AUC: 0.82. Simulation analysis demonstrated that incorporating no-show probabilities into the scheduling system increased profit, utilization with a slight increase in patient waiting time
Kheirkhah (2010)	Develop a decision support system for patient scheduling and overbooking by using predictive information about the show-up probability of a patient	691,100 appointments	LR (stepwise selection), support vector machines (SVM)	Patient related and appointment related information, home phone made, follow-up visit, current appointment status	All features are selected. SVM model performance: Accuracy: 81.25%, Sensitivity: 13.18% Specificity: 97.74% LR performance: Accuracy: 81.26%, Sensitivity: 12.36 % Specificity: 97.74% AUC: 0.72
Aladeddini et al. (2011)	Patients' no-show and cancellations prediction in real-time	1,543 appointments / 99 patients	Hybrid probabilistic model based on multinomial LR and Bayesian inference	Gender, date of birth, marital status, medical coverage, distance, clinic type, no-show history	Proposed method has mean squared error (MSE) = 0.01 and accuracy of 88%

Rinder (2012)	Patient no-show prediction at individual level. Integrate predictions into schedule overbooking policy	87,173 appointments	Artificial neural networks and decision tree (DT) mining followed by optimal sequential scheduling	Clinic ZIP code, appointment type, date and time, provider type and name, patient related information, parents' employment status, check-in time, vitals time, discharge time, encounter number, weather-related factors, distance, appointment time, average income, poverty level and education	. Neural networks accuracy: 90%
Norris et al. (2014)	Predict patients' arrival, cancellation, and no-show	858,579 appointments/ 88,345 patients	Multinomial LR combined with decision tree	Lead time, attendance history, payer, age, weather, day of week, time of day, weather, day of week, time of day	Significant factors: lead time, attendance history, payer and age. Prediction accuracy was not reported
Huang and Hanauer (2012)	Utilize no-show prediction information to improve overbooking strategy	104,799 appointments/ 7,988 patients	Likelihood ratio Chi-square tests for feature selection and LR for no-show prediction	Patient related and appointment related information, distance, primary insurance, main insurance holder and total insurance carriers	Excluded factors: religion and gender Proposed overbooking lowered average waiting time by: 6%-8%, overtime by: 24%-29%
Samorani and LaGanga (2015)	Maximize the volume of patients being seen and minimize staff overtime and patient waiting time	50,000 appointments/ 6700 patients	Dynamic programming, cost-sensitive Bayesian network and overbooking heuristic	Patients demographics, RMI and BRM scores, patient's ID code, lead time, attendance history, past appointments information, day of week, location, weather, age and diagnosis	Accuracy: 70% Sensitivity: 71% Specificity: 67% Authors also provided a heuristic scheduling policy

Gonfiantini et al. (2015)	Propose an overbooking strategy utilizing two predictive pieces of information: patients' no-shows and unpunctuality	8,921 appointments	C4.5 decision tree, LR, Support vector machine, Artificial neural network, Bayesian network	Patient related and appointment related information	LR was selected. Performance metrics: Accuracy: 81% Sensitivity: 65% Specificity: 87% AUC: 0.83 Proposed overbooking strategy outperformed all other overbooking strategies
Huang and Hanauer (2016)	Develop patient no-show prediction model	7,291 patients	LR	Patient related and appointment related information, total insurance carriers, primary insurance, total insurance carriers and main insurance holder	Prediction model performance: AUC: 0.69 on testing data set and 0.70 on training data set
Al-Mashraie (2016)	Patients' no-show prediction	55,699 appointments	Deep learning neural network	Patient related and appointment related information, portal registration, contact type, reminder patient, communication id, relationship id, number of previous visits, primary language id, average precipitation, average temperature	Most important factor: no-show history, followed by age and insurance type. All features were included for prediction. Prediction model performance on Accuracy: 74% AUC: 0.72
Elvira et al. (2017)	Patient no-show probability prediction	2,234,119 appointments	Gradient Boosting Algorithm	Patient related and appointment related information, medical specialty, monographic, consultation, center name	Prediction model performance: AUC: 0.74

Ding et al. (2018)	Derive a risk score for patient no-show	2,232,737 outpatient appointments, across 14 different specialties and 55 clinics	Regularized LR	Patient demographics, comorbidities, service utilization history, appointment information, financial information, patient engagement with the online portals and automated phone call prior to the appointment	No-show rates varied from 13% to 32% across the 14 specialties. Risk score for patient no-show was derived, average C-statistic: 0.83 Most important factor was whether the physician rescheduled the appointment
Mohammadi et al. (2018)	Develop patient no-show prediction model to understand appointment adherence in underserved populations	73,811 unique appointments	LR, artificial neural network (ANN), and Naive Bayes classifier	Patient characteristics, translator needed or not, insurance information, employment status, tobacco use, annual income and patient portal use. Clinic and appointment characteristics such as specialty, season, lead time, visit type, weekday	No-show rate: 16.7%. LR performance: AUC: 0.81, Sensitivity: 0.72, Specificity: 0.54, Accuracy: 73%. ANN performance: AUC: 0.66, Sensitivity: 0.63, Specificity: 0.35 Accuracy 71%. Naïve Bayes performance: AUC: 0.86, Sensitivity: 0.73, Specificity: 0.54 Accuracy: 82%
Topuz et al. (2018)	Propose a hybrid probabilistic prediction framework for patient no-show	16,345 unique patients	Elastic net (EN) variable-selection methodology integrated with probabilistic Bayesian Belief Network (BBN)	Patient socioeconomic status, demographic, appointment information, patient no-show history	Proposed hybrid framework was compared with BNN without feature selection, DT, logistic regression and ANN. Proposed EN-BNN framework outperforms other methods AUC: 0.69 Accuracy: 74.3%

Lenzi et al. (2019)	Assess factors associated with patient no-show. Develop a no-show prediction model	57,586 appointments	Stepwise naïve and mixed-effect logistic regression. In addition to the Akaike Information Criteria for model selection	Patient related and appointment related information, if it is a same-day appointment, provider type and specialty	No-show rate: 13%. AUC: 0.80
Dantas et al. (2019)	Examine factors associated with patient no-show	2,660 patients	Multiple LR analyses	Age, gender, payment form, clinic specialty, lead time, appointment time, no-show history, number of previous appointments, appointment type, month and weekday, distance	No-show rate: 21.9%. Eight factors were found to be associated. Accuracy: 71%.
Li et al. (2019)	Provide an estimated no-show probability at the patient level	42,903 patients	Bayesian nested logit	Appointment time, provider type, cancellation and no-show history, confirmation status, patients demographics, marital status, season, insurance type, age	Bayesian model resulted in 30% improvement in model fit. Simulation shows that utilizing patient no-show probability for scheduling will increase clinic profit. Patients were grouped into 3 categories: most likely, less likely and least likely to confirm
Chua and Chow (2019)	Develop patient no-show prediction model. Create a risk stratification model for patient no-show	75,677 appointments	Univariate analysis and multiple LR using SPSS software	Patient related and appointment related information, if appointment was after a public holiday or on a school holiday, referral source, previous visit status and type	No-show rate: 28.6%. AUC: 0.72 Five no-show risk groups were created, ranging from extremely low no-show risk to extremely high no-show

Appendix B: Description of all 21 original features used in this research

Factor	Description	Type	Note
Indirect waiting time	Days between appointment created date and date of the scheduled appointment	Continuous	-
Weekday	Day of the week that appointment was scheduled	Categorical	-
Part of day	Whether the appointment start time was before or after 12 pm	Categorical	-
Appointment date	Date of appointment	Continuous	-
Appointment start time	Time appointment is scheduled to start	Continuous	-
Appointment duration	Scheduled duration of the appointment	Continuous	-
Age group	Which age group does the patient belong to	Categorical	Five levels: greatest generation, baby boomers, generation X, millennials, and pediatric
Gender	Patient's gender documented in the information system	Categorical	Two levels: female and male
Marital status	Marital status of the patient	Categorical	Four levels: divorced, married, single, and widowed
Financial class	Insurance type mentioned in patient record	Categorical	-
Adjusted risk score	Modified total risk score of patient	Categorical	Three levels: low, medium, and high
No-show history	Number of previous no-shows	Categorical	Discrete numbers 0-24
Precipitation	Rainfall in appointment day (inches)	Continuous	-
Snowfall	Amount of snow in appointment day (inches)	Continuous	-
Average temperature	Average of hourly temperature in appointment day (Fahrenheit unit)	Continuous	-
Days from last ED visit	Number of days between scheduled appointment and last ED discharge	Continuous	-
Provider type	Healthcare provider type	Categorical	Nine levels including: dieticians, doctor, lab

			personnel, nurse practitioner, nurse, physician assistant, resident, social worker
Appointment created same day of ED discharge	Whether appointment was created on the same day of the ED discharge or not	Categorical	Two levels: yes or no
Visit type	Type of scheduled visit	Categorical	Seven levels: new visit, nurse visit, office visit (follow-up visit with a doctor), procedure, resident new visit, resident visit (follow-up visit with resident), and transitional care management
Appointment confirmation status	Represents the outcome of the appointment confirmation process	Categorical	Six levels: cancelled by doctor, cancelled by patient, confirmed, pre-registered, rescheduled, and unconfirmed
Distance	Distance in miles between patient and clinic ZIP codes	Continuous	-
Class	Response variable	Categorical	Two levels: show and no-show

References

Ahandani, Morteza Alinia, and Hosein Alavi-Rad. "Opposition-based learning in the shuffled differential evolution algorithm." *Soft Computing* 16, no. 8 (2012): 1303-1337.

Ahandani, Morteza Alinia. "Opposition-based learning in the shuffled bidirectional differential evolution algorithm." *Swarm and Evolutionary Computation* 26 (2016): 64-85.

Aladeemy, Mohammed, Salih Tutun, and Mohammad T. Khasawneh. "A new hybrid approach for feature selection and Support Vector Machine model selection based on Self-Adaptive Cohort Intelligence." *Expert Systems with Applications* 88 (2017): 118-131.

Alaeddini, Adel, Kai Yang, Chandan Reddy, and Susan Yu. "A probabilistic model for predicting the probability of no-show in hospital appointments." *Health Care Management Science* 14, no. 2 (2011): 146-157.

Aljarah, Ibrahim, Majdi Mafarja, Ali Asghar Heidari, Hossam Faris, Yong Zhang, and Seyedali Mirjalili. "Asynchronous accelerating multi-leader salp chains for feature selection." *Applied Soft Computing* 71 (2018): 964-979.

Al-Mashraie, M. "Simulation-based evaluation of no-show prediction using deep learning neural networks." Master Thesis, State University of New York at Binghamton, 2016.

Alpert, Joel J. "Broken appointments." *Pediatrics* 34, no. 1 (1964): 127-132.

Amoshahy, Mohammad Javad, Mousa Shamsi, and Mohammad Hossein Sedaaghi. "A novel flexible inertia weight particle swarm optimization algorithm." *PloS one* 11, no. 8 (2016): e0161558.

Antunes, Francisco, Bernardete Ribeiro, and Francisco Pereira. "Probabilistic modeling and visualization for bankruptcy prediction." *Applied Soft Computing* 60 (2017): 831-843.

Banerjee, Abhik, V. Mukherjee, and S. P. Ghoshal. "An opposition-based harmony search algorithm for engineering optimization problems." *Ain Shams Engineering Journal* 5, no. 1 (2014): 85-101.

Basu, M. "Quasi-oppositional differential evolution for optimal reactive power dispatch." *International Journal of Electrical Power & Energy Systems* 78 (2016): 29-40.

Centers for Medicare and Medicaid Services, "National Health Expenditure Accounts: Methodology Paper, 2014". CMS.gov. https://www.cms.gov/research-statistics-data-and-systems/statistics-trends-and-reports/nationalhealthexpenddata/downloads/dsm-14.pdf. (accessed July 4, 2017).

Centers for Medicare and Medicaid Services, "National Health Expenditure Accounts: Methodology Paper, 2015". CMS.gov. https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/NationalHealthExpendData/downloads/dsm-15.pdf. (accessed July 4, 2017).

Centers for Medicare and Medicaid Services. "National Health Expenditure Projections 2016-2025". CMS.gov. https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/NationalHealthExpendData/Downloads/proj2016.pdf (accessed July 4, 2017).

Chen, Yu, Weicheng Xie, and Xiufen Zou. "A binary differential evolution algorithm learning from explored solutions." *Neurocomputing* 149 (2015): 1038-1047.

Chua, Siang Li, and Wai Leng Chow. "Development of predictive scoring model for risk stratification of no-show at a public hospital specialist outpatient clinic." *Proceedings of Singapore Healthcare* 28, no. 2 (2019): 96-104.

Coelho, Francisco, and João Pedro Neto. "A method for regularization of evolutionary polynomial regression." *Applied Soft Computing* 59 (2017): 223-228.

Cortés, Pablo, Rodrigo A. Gómez-Montoya, Jesús Muñuzuri, and Alexander Correa-Espinal. "A tabu search approach to solving the picking routing problem for large-and medium-size distribution centres considering the availability of inventory and K heterogeneous material handling equipment." *Applied Soft Computing* 53 (2017): 61-73.

Daggy, Joanne, Mark Lawley, Deanna Willis, Debra Thayer, Christopher Suelzer, Po-Ching DeLaurentis, Ayten Turkcan, Santanu Chakraborty, and Laura Sands. "Using no-show modeling to improve clinic performance." *Health Informatics Journal* 16, no. 4 (2010): 246-259.

Dantas, Leila F., Silvio Hamacher, Fernando L. Cyrino Oliveira, Simone DJ Barbosa, and Fábio Viegas. "Predicting patient no-show behavior: a study in a bariatric clinic." *Obesity surgery* 29, no. 1 (2019): 40-47.

Deyo, Richard A., and Thomas S. Inui. "Dropouts and broken appointments: a literature review and agenda for future research." *Medical Care* 18, no. 11 (1980): 1146-1157.

Ding, Xiruo, Ziad F. Gellad, Chad Mather III, Pamela Barth, Eric G. Poon, Mark Newman, and Benjamin A. Goldstein. "Designing risk prediction models for ambulatory no-shows across different specialties and clinics." *Journal of the American Medical Informatics Association* 25, no. 8 (2018): 924-930.

Dong, Wenyong, Lanlan Kang, and Wensheng Zhang. "Opposition-based particle swarm optimization with adaptive mutation strategy." *Soft Computing* 21, no. 17 (2017): 5081-5090.

Dove, Henry G., and Karen C. Schneider. "The usefulness of patients' individual characteristics in predicting no-shows in outpatient clinics." *Medical Care* 19, no. 7 (1981): 734-740.

Eberhart, Russell, and James Kennedy. "A new optimizer using particle swarm theory." In *MHS'95. Proceedings of the Sixth International Symposium on Micro Machine and Human Science*, pp. 39-43. Ieee, 1995.

Elaziz, Mohamed Abd, Diego Oliva, and Shengwu Xiong. "An improved opposition-based sine cosine algorithm for global optimization." *Expert Systems with Applications* 90 (2017): 484-500.

Elvira, Carlos, Alberto Ochoa, Juan Carlos Gonzalvez, and Francisco Mochón. "Machine-Learning-Based No Show Prediction in Outpatient Visits." *International Journal of Interactive Multimedia and Artificial Intelligence* 4, no. Special Issue on Big Data and e-Health (2018).

Esmailzadeh, Ali, and Shahryar Rahnamayan. "Opposition-based differential evolution with protective generation jumping." In *Differential Evolution (SDE), 2011 IEEE Symposium on*, pp. 1-8. IEEE, 2011.

Faris, Hossam, Majdi M. Mafarja, Ali Asghar Heidari, Ibrahim Aljarah, Al-Zoubi Ala'M, Seyedali Mirjalili, and Hamido Fujita. "An efficient binary salp swarm algorithm with crossover scheme for feature selection problems." *Knowledge-Based Systems* 154 (2018): 43-67.

Ferrari, Alessandro, Stefano Lombardi, and Alberto Signoroni. "Bacterial colony counting with convolutional neural networks in digital microbiology imaging." *Pattern Recognition* 61 (2017): 629-640.

Gandomi, Amir H. "Interior search algorithm (ISA): a novel approach for global optimization." *ISA transactions* 53, no. 4 (2014): 1168-1183.

Glover, Fred. "Tabu search—part I." *ORSA Journal on computing* 1, no. 3 (1989): 190-206.

Glowacka, Karolina J., Raymond M. Henry, and Jerrold H. May. "A hybrid data mining/simulation approach for modelling outpatient no-shows in clinic scheduling." *Journal of the Operational Research Society* 60, no. 8 (2009): 1056-1068.

Goldberg, David E. "Genetic algorithms in search, optimization and machine learning." *Reading: Addison-Wesley, 1989*(1989).

Goldman, Lee, Ralph Freidin, E. Francis Cook, John Eigner, and Pamela Grich. "A multivariate approach to the prediction of no-show behavior in a primary care center." *Archives of Internal Medicine* 142, no. 3 (1982): 563-567.

Gonfiantini, Michael J. "Utilizing individual patient unpunctuality and no-show predictions to develop optimal appointment overbooking strategies in an outpatient setting." PhD dissertation. State University of New York at Binghamton, 2015.

Guedria, Najeh Ben. "Improved accelerated PSO algorithm for mechanical engineering optimization problems." *Applied Soft Computing* 40 (2016): 455-467.

Heidari, Ali Asghar, Rahim Ali Abbaspour, and Huiling Chen. "Efficient boosted grey wolf optimizers for global search and kernel extreme learning machine training." *Applied Soft Computing* 81 (2019): 105521.

Huang, Cheng-Lung, and Chieh-Jen Wang. "A GA-based feature selection and parameters optimization for support vector machines." *Expert Systems with applications* 31, no. 2 (2006): 231-240.

Huang, Y., and David A. Hanauer. "Patient no-show predictive model development using multiple data sources for an effective overbooking approach." *Applied Clinical Informatics* 5, no. 3 (2014): 836-860.

Huang, Y., and P. Zuniga. "Dynamic overbooking scheduling system to improve patient access." *Journal of the Operational Research Society* 63, no. 6 (2012): 810-820.

Huang, Yu-Li, and David A. Hanauer. "Time dependent patient no-show predictive modelling development." *International journal of health care quality assurance* 29, no. 4 (2016): 475-488.

Johnson, Bradley J., James W. Mold, and J. Michael Pontious. "Reduction and management of no-shows by family medicine residency practice exemplars." *The Annals of Family Medicine* 5, no. 6 (2007): 534-539.

Katrutsa, Alexandr, and Vadim Strijov. "Comprehensive study of feature selection methods to solve multicollinearity problem according to evaluation criteria." *Expert Systems with Applications* 76 (2017): 1-11.

Khanesar, Mojtaba Ahmadieh, Mohammad Teshnehlab, and Mahdi Aliyari Shoorehdeli. "A novel binary particle swarm optimization." *In Control & Automation*, 2007. MED'07. Mediterranean Conference on, pp. 1-6. IEEE, 2007.

Kheirkhah, Parviz. "Development of the patient scheduling system to handle patient no-show problem using data mining and simulation-based optimization techniques." PhD dissertation. University of Houston, 2010.

Kirkpatrick, Scott, C. Daniel Gelatt, and Mario P. Vecchi. "Optimization by simulated annealing." *science* 220, no. 4598 (1983): 671-680.

Kulkarni, Anand J., Ishan P. Durugkar, and Mrinal Kumar. "Cohort intelligence: a self supervised learning behavior." In *2013 IEEE international conference on systems, man, and cybernetics*, pp. 1396-1400. IEEE, 2013.

LaGanga, Linda R., and Stephen R. Lawrence. "Appointment overbooking in health care clinics to improve patient service and clinic performance." *Production and Operations Management* 21, no. 5 (2012): 874-888.

Lagerlund, Magdalena, Anna Hedin, Pär Sparén, Erik Thurfjell, and Mats Lambe. "Attitudes, beliefs, and knowledge as predictors of nonattendance in a Swedish population-based mammography screening program." *Preventive Medicine* 31, no. 4 (2000): 417-428.

Lee, Vernon J., Arul Earnest, Mark I. Chen, and Bala Krishnan. "Predictors of failed attendances in a multi-specialty outpatient centre using electronic databases." *BMC Health Services Research* 5, no. 1 (2005): 51.

Lehmann, T. N. O., A. Aebi, D. Lehmann, M. Balandraux Olivet, and H. Stalder. "Missed appointments at a Swiss university outpatient clinic." *Public Health* 121, no. 10 (2007): 790-799.

Lenzi, Henry, Ângela Jornada Ben, and Airton Tetelbom Stein. "Development and validation of a patient no-show predictive model at a primary care setting in Southern Brazil." *PloS one*14, no. 4 (2019): e0214869.

Li, Yutian, Sammi Yu Tang, Joseph Johnson, and David A. Lubarsky. "Individualized No-show Predictions: Effect on Clinic Overbooking and Appointment Reminders." *Production and Operations Management* (2019).

Lin, Shih-Wei, Kuo-Ching Ying, Shih-Chieh Chen, and Zne-Jung Lee. "Particle swarm optimization for parameter determination and feature selection of support vector machines." *Expert systems with applications* 35, no. 4 (2008): 1817-1824.

Mafarja, Majdi, and Seyedali Mirjalili. "Whale optimization approaches for wrapper feature selection." *Applied Soft Computing* 62 (2018): 441-453.

Mafarja, Majdi, Ibrahim Aljarah, Ali Asghar Heidari, Abdelaziz I. Hammouri, Hossam Faris, Al-Zoubi Ala'M, and Seyedali Mirjalili. "Evolutionary population dynamics and grasshopper optimization approaches for feature selection problems." *Knowledge-Based Systems* 145 (2018): 25-45.

Mafarja, Majdi, Ibrahim Aljarah, Ali Asghar Heidari, Hossam Faris, Philippe Fournier-Viger, Xiaodong Li, and Seyedali Mirjalili. "Binary dragonfly optimization for feature selection using time-varying transfer functions." *Knowledge-Based Systems* 161 (2018): 185-204.

Mafarja, Majdi, Ibrahim Aljarah, Hossam Faris, Abdelaziz I. Hammouri, Al-Zoubi Ala'M, and Seyedali Mirjalili. "Binary grasshopper optimisation algorithm approaches for feature selection problems." *Expert Systems with Applications* 117 (2019): 267-286.

Mahdavi, Sedigheh, Shahryar Rahnamayan, and Kalyanmoy Deb. "Opposition based learning: A literature review." *Swarm and evolutionary computation* 39 (2018): 1-23.

Masood, M. K., Yeng Chai Soh, and Chaoyang Jiang. "Occupancy estimation from environmental parameters using wrapper and hybrid feature selection." *Applied Soft Computing* 60 (2017): 482-494.

Medicare Payment Advisory Commission, June 2016 "A Data Book: Health Care Spending and the Medicare Program." Medpac.gov. http://www.medpac.gov/docs/default-source/data-book/june-2016-data-book-health-care-spending-and-the-medicare-program.pdf (accessed July 19, 2016).

Mirjalili, Seyedali, and Andrew Lewis. "The whale optimization algorithm." *Advances in engineering software* 95 (2016): 51-67.

Mirjalili, Seyedali, Seyed Mohammad Mirjalili, and Andrew Lewis. "Grey wolf optimizer." *Advances in engineering software* 69 (2014): 46-61.

Mirjalili, Seyedali. "Dragonfly algorithm: a new meta-heuristic optimization technique for solving single-objective, discrete, and multi-objective problems." *Neural Computing and Applications* 27, no. 4 (2016): 1053-1073.

Mohammadi, Iman, Huanmei Wu, Ayten Turkcan, Tammy Toscos, and Bradley N. Doebbeling. "Data Analytics and Modeling for Appointment No-show in Community Health Centers." *Journal of primary care & community health* 9 (2018): 2150132718811692.

Moore, Charity G., Patricia Wilson-Witherspoon, and Janice C. Probst. "Time and money: effects of no-shows at a family practice residency clinic." *Family Medicine-Kansas City*- 33, no. 7 (2001): 522-527.

Nakisa, Bahareh, Mohammad Naim Rastgoo, Dian Tjondronegoro, and Vinod Chandran. "Evolutionary computation algorithms for feature selection of EEG-based emotion recognition using mobile sensors." *Expert Systems with Applications* 93 (2018): 143-155.

Norris, John B., Chetan Kumar, Suresh Chand, Herbert Moskowitz, Steve A. Shade, and Deanna R. Willis. "An empirical investigation into factors affecting patient cancellations and no-shows at outpatient clinics." *Decision Support Systems* 57 (2014): 428-443.

Nuti, Lynn A., Mark Lawley, Ayten Turkcan, Zhiyi Tian, Lingsong Zhang, Karen Chang, Deanna R. Willis, and Laura P. Sands. "No-shows to primary care appointments: subsequent acute care utilization among diabetic patients." *BMC Health Services Research* 12, no. 1 (2012): 304.

Perron, Noelle Junod, Melissa Dominicé Dao, Michel P. Kossovsky, Valerie Miserez, Carmen Chuard, Alexandra Calmy, and Jean-Michel Gaspoz. "Reduction of missed appointments at an urban primary care clinic: a randomised controlled study." *BMC Family Practice* 11, no. 1 (2010): 79.

Rahnamayan, Shahryar, Hamid R. Tizhoosh, and Magdy MA Salama. "Opposition-based differential evolution (ODE) with variable jumping rate." *In Foundations of Computational Intelligence*, 2007. FOCI 2007. IEEE Symposium on, pp. 81-88. IEEE, 2007.

Rahnamayan, Shahryar, Hamid R. Tizhoosh, and Magdy MA Salama. "Opposition versus randomness in soft computing techniques." *Applied Soft Computing* 8, no. 2 (2008): 906-918.

Rahnamayan, Shahryar, Hamid R. Tizhoosh, and Magdy MA Salama. "Opposition-based differential evolution algorithms." In *Evolutionary Computation, 2006. CEC 2006. IEEE Congress on*, pp. 2010-2017. IEEE, 2006.

Rashedi, Esmat, Hossein Nezamabadi-Pour, and Saeid Saryazdi. "GSA: a gravitational search algorithm." *Information sciences* 179, no. 13 (2009): 2232-2248.

Rinder, Maria M. "An Integrated Decision-Support Tool to Forecast and Schedule No-Show Appointments in Healthcare." PhD dissertation. Ohio University, 2012.

Rojas-Morales, Nicolas, María-Cristina Riff Rojas, and Elizabeth Montero Ureta. "A survey and classification of opposition-based metaheuristics." *Computers & Industrial Engineering* 110 (2017): 424-435.

Samorani, Michele, and Linda R. LaGanga. "Outpatient appointment scheduling given individual day-dependent no-show predictions." *European Journal of Operational Research* 240, no. 1 (2015): 245-257.

Schwartz, Mark D. "Health care reform and the primary care workforce bottleneck." *Journal of General Internal Medicine* 27, no. 4 (2012): 469-472.

Seif, Z., and Mohammad Bagher Ahmadi. "Opposition versus randomness in binary spaces." *Applied Soft Computing* 27 (2015): 28-37.

Shonick, William, and Bonnie W. Klein. "An approach to reducing the adverse effects of broken appointments in primary care systems: development of a decision rule based on estimated conditional probabilities." *Medical Care* (1977): 419-429.

Si, Tapas, Arunava De, and Anup Kumar Bhattacharjee. "Particle swarm optimization with generalized opposition based learning in particle's pbest position." In *Circuit, Power and Computing Technologies (ICCPCT), 2014 International Conference on*, pp. 1662-1667. IEEE, 2014.

Storn, Rainer, and Kenneth Price. "Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces." *Journal of global optimization* 11, no. 4 (1997): 341-359.

Tizhoosh, Hamid R. "Opposition-based learning: a new scheme for machine intelligence." In *Computational Intelligence for Modelling, Control And Automation, 2005 And International Conference On Intelligent Agents, Web Technologies And Internet Commerce, International Conference On*, vol. 1, pp. 695-701. IEEE, 2005.

Tizhoosh, Hamid R. "Reinforcement learning based on actions and opposite actions." In *International conference on artificial intelligence and machine learning*, vol. 414. 2005.

Tizhoosh, Hamid R., and Mario Ventresca, eds. *Oppositional concepts in computational intelligence*. Vol. 155. Springer, 2008.

Topuz, Kazim, Hasmet Uner, Asil Oztekin, and Mehmet Bayram Yildirim. "Predicting pediatric clinic no-shows: a decision analytic framework using elastic net and Bayesian belief network." *Annals of Operations Research* 263, no. 1-2 (2018): 479-499.

Ventresca, Mario, and Hamid R. Tizhoosh. "A diversity maintaining population-based incremental learning algorithm." *Information Sciences* 178, no. 21 (2008): 4038-4056.

Ventresca, Mario, Shahryar Rahnamayan, and Hamid R. Tizhoosh. "A note on “Opposition versus randomness in soft computing techniques”[Appl. Soft Comput. 8 (2)(2008) 906–918]." *Applied Soft Computing* 10, no. 3 (2010): 956-957.

Wallace, Byron C., Kevin Small, Carla E. Brodley, and Thomas A. Trikalinos. "Class imbalance, redux." In *Data Mining (ICDM), 2011 IEEE 11th International Conference on*, pp. 754-763. IEEE, 2011.

Wang, Hui, Zhijian Wu, Shahryar Rahnamayan, and Lishan Kang. "A scalability test for accelerated DE using generalized opposition-based learning." In *Intelligent Systems Design and Applications, 2009. ISDA'09. Ninth International Conference on*, pp. 1090-1095. IEEE, 2009.

Xue, Bing, Mengjie Zhang, and Will N. Browne. "Particle swarm optimisation for feature selection in classification: Novel initialisation and updating mechanisms." *Applied Soft Computing* 18 (2014): 261-276.

Xue, Bing, Mengjie Zhang, Will N. Browne, and Xin Yao. "A survey on evolutionary computation approaches to feature selection." *IEEE Transactions on Evolutionary Computation* 20, no. 4 (2016): 606-626.

Highlights

- Three variants of Opposition-based Self-Adaptive Cohort Intelligence (OSACI) algorithm are proposed.
- New wrapper feature selection methods based on OSACI are presented.
- The proposed wrapper methods were tested for patient no-show prediction.
- The results revealed the effectiveness of the proposed algorithms for feature selection.

The authors declare that there is no conflict of interest.