Exploring Social and Environmental Factors Contributing to Smoking Initiation Among Thai Adolescents Using Advanced Feature Selection Techniques

The proposed algorithm for feature selection

In recent decades, meta-heuristic algorithms have demonstrated their effectiveness across various research domains, prompting researchers to dedicate significant efforts to enhance their performance. Moreover, developing algorithms that offer superior performance is driven by the goal of further enhancing the capabilities of meta-heuristic algorithms. Ultimately, the objective is to discover the most optimal solution for NP-hard problems. However, it is important to note that there is no definitive "best result" by which one can always improve the findings with new or modified algorithms.

Additionally, the No Free Lunch (NFL) theorem, proposed by Wolpert and Macready in 1997 [1], states that any two algorithms would yield equivalent results when evaluated across all possible optimization problems. Consequently, while an algorithm may demonstrate superior performance on certain problems, it does not guarantee similar outcomes on all other problems. Therefore, it can be concluded that there is no universally applicable algorithm capable of effectively addressing all optimization problems, and which can consistently generate the best possible results.

These observations contribute to the resilience of research in this field, particularly in the context of feature selection (FS) being recognized as an optimization problem [2]. As a result, researchers are actively developing novel and efficient FS methods by leveraging meta-heuristic algorithms.

Chaotic map

Chaos is a phenomenon characterized by its sensitivity to initial conditions wherein even minor changes in the initial conditions can result in nonlinear variations in future behavior. Chaos optimization represents a recent class of search algorithms that leverage this concept. The fundamental concept behind chaos optimization involves mapping parameters or variables from a chaotic system into the solution space. The search for global optima in chaos optimization depends on the inherent properties of chaotic motion such as ergodicity, regularity, and stochasticity. These properties play a crucial role in enabling chaos optimization algorithms to explore and identify global optimum solutions effectively. COA (Chaos Optimization Algorithm) offers several key advantages, including a fast convergence rate and the ability to avoid local minima. These privileges contribute to a significant improvement in the overall performance of evolutionary algorithms [3]. Chaotic maps exhibit deterministic behavior without the use of random factors.

Feature selection theory

In the context of a classification task with N features, the problem of feature selection is commonly formulated as a single-objective optimization task, which can be expressed as follows: [1].

$$\begin{aligned} &\min\ E(x) \\ &s.t.\ x = (x_1, x_2, \ldots, x_N) \\ &\qquad x_i \in \{0,1\}, i = 1, \ldots, N. \end{aligned} \tag{1}$$

In Eq. (1), the feature selection problem for a classification task with N features involves optimizing the classification error, denoted as $E(x)$, where x represents a feature subset. Each feature in the subset can take on a binary value of either 0 or 1, with "1" indicating the selection of the corresponding feature and "0" indicating its exclusion. The problem of feature selection poses significant challenges due to its extensive search space, rendering it an NP-hard combinatorial optimization problem. To tackle such complexities, meta-heuristic algorithms have emerged as a viable solution [4], demonstrating their effectiveness in solving various combinatorial optimization problems and finding practical applications in numerous fields. Most of the research on feature selection primarily emphasizes the identification and elimination of irrelevant or redundant features to enhance classification accuracy, often overlooking other crucial factors. Nevertheless, for high-dimensional datasets, it is equally vital to minimize the number of selected features to optimize computational resources and expedite computational speed.

Furthermore, it has been established that feature selection constitutes a multi-objective optimization problem [5]. Consequently, several meta-heuristic algorithms designed for multi-objective optimization have been employed to address feature selection challenges effectively. Hence, the feature selection problem is regarded as a multi-objective optimization problem due to the inherent conflict between its primary objectives: enhancing classification performance and minimizing the number of selected features. These objectives often exhibit a trade-off relationship, further emphasizing the multi-objective nature of the problem. Dealing with the huge search space makes the algorithm slow down. Therefore, to solve this problem, this research focuses on the combination of feature selection and the metaheuristic optimization algorithm. In addition, there are also modern techniques that are integrated into many parts.

Grey Wolf Optimization

The Grey Wolf Optimization (GWO) algorithm is a metaheuristic optimization algorithm inspired by the social behavior of grey wolves in nature. It was proposed by [6]. The GWO algorithm is a population-based optimization technique that mimics the hunting behavior of grey wolves in a hierarchical structure. The algorithm iteratively searches for the optimal solution by imitating the social interactions and hunting mechanisms of wolf packs. Here's a brief overview of the steps involved in the GWO algorithm [7]:

Algorithm 1: Pseudo Code of GWO algorithm

input: $\boldsymbol{NP}$, the population size; $\boldsymbol{T}$, the maximum number of iterations.
output: The optimal individual $\vec{X}_\alpha$, the best fitness value $f(\vec{X}_\alpha)$.

Initialize a population $X_i (i = 1,2,3, \ldots, n)$ of $\boldsymbol{NP}$ individuals randomly.
Initialize parameters t = 1, $\vec{\alpha}$.
Calculate the fitness value of each individual and record the best individual, second-best individual, and the third-best individual as $\vec{X}_\alpha$, $\vec{X}_\beta$ and $\vec{X}_\delta$.
while $t \leq T$ **do**
    for everyone **do**
        Update the positions of the remaining wolves based on the positions of $\vec{X}_\alpha$, $\vec{X}_\beta$ and $\vec{X}_\delta$, simulating the hunting behavior.
    end
    Calculate the fitness value of everyone in the population and update $\vec{X}_\alpha$, $\vec{X}_\beta$ and $\vec{X}_\delta$.
    $t \leftarrow t + 1$.
    Update $\vec{\alpha}$
end
Output $\vec{X}_\alpha$ and $f(\vec{X}_\alpha)$.

The GWO algorithm leverages cooperation and communication among wolves to explore the search space effectively and converge toward the optimal solution. The alpha, beta, and delta wolves play key roles in guiding the search process and driving the exploration and exploitation balance. The GWO algorithm has been applied to various optimization problems and has shown promising results in terms of convergence speed and accuracy. However, like other metaheuristic algorithms, its performance can vary depending on the problem at hand and the appropriate tuning of its parameters. The designing of an ACBGWO algorithm to address feature selection issues is described in sub-Section "Feature selection by ACBGWO approach".

Feature selection by ACBGWO approach

Feature selection can be conceptualized as a search problem, wherein each state within the search space represents a potential subset of features for the task at hand. However, conducting an exhaustive evaluation of all possible feature subsets is often impractical due to the substantial computational effort involved. Numerous search techniques have been developed to address the feature selection problem, employing intelligent search methods within the solution space. One such approach is the GWO algorithm, an evolutionary search algorithm that has been effectively employed as the search engine in the feature selection process [2]. However, the GWO algorithm has not been able to accurately and efficiently analyze the patterns of new smoker initiation from large-scale data on risky smoking behavior. Therefore, the research team has improved the original GWO algorithm by incorporating additional techniques called Adaptive Chaotic and Binary. These enhancements have made the GWO algorithm more effective in selecting accurate and fast feature selection.

In this section, the researchers propose a novel feature selection method, namely Adaptive Chaotic Binary Grey Wolf Optimization (ACBGWO), building upon the principles of the optimization algorithm. First, one establishes the connection between the ACBGWO search algorithm and the problem of feature selection, elucidating how the ACBGWO algorithm can effectively address feature selection challenges. In the second subsection, the study provides a comprehensive explanation of the evaluation function utilized in the feature selection process. In the third subsection, there are presented several key considerations regarding the final evaluation process. These considerations encompass crucial aspects of the feature selection process. Additionally, the fourth subsection outlines the stopping criterion employed in the ACBGWO algorithm to determine the termination conditions for the search. Several considerations about the final evaluation process are given in the third subsection, and the stopping criterion of ACBGWO is presented in the fourth subsection.

The GWO algorithm was originally introduced by [6]. Since then, several researchers and academics have endeavored to enhance its applicability and expand its capabilities in various domains. For example, there are Binary Grey Wolf Optimization approaches for feature selection [8]; a Chaotic Grey Wolf Optimization algorithm for constrained optimization problems [9]; a feature selection using Binary Grey Wolf Optimizer with elite-based crossover for Arabic text classification [10]; and finally, an Improved Binary Grey Wolf Optimizer and its application for feature selection [11].

Nevertheless, the challenges of premature convergence and susceptibility to local optima remain prevalent issues associated with the GWO algorithm. Therefore, it is proposed that the Adaptive Chaotic Binary Grey Wolf Optimization (ACBGWO) is implemented in order to improve upon the optimization performance of the original GWO. The two major highlights of the ACBGWO algorithm are described below.

Firstly, the primary objective of the proposed solution is to enhance global convergence and mitigate the risk of getting trapped in local minima within the GWO algorithm. To achieve this, the proposed solution incorporates a replacement of the r_1 (random parameter) in GWO [6] with chaotic mapping. This integration introduces ergodic irregularity and stochastic properties, thereby enhancing the exploration capability of the algorithm. In order to enhance the search capability of the GWO algorithm by utilizing chaotic (r_1) parameters, the experiments have shown that the logistic map [3] exhibits a smooth and effective integration with GWO. The iterative form of the logistic map, as employed in the algorithm, is defined by Eq. (2) as follows:

$$z^{t+1} = az^t(1 - z^t), \quad (2)$$

where $a = 4$, and t denotes the iteration number. This logistic chaotic map is utilized to generate a chaotic signal within the intervals $z^t \in (0, 1)$, with z^t representing the initial value of the t^{th} chaotic number. Within the method of embedding chaotic maps into the GWO, it is defined through Eq. (3).

$$r_1^{t+1} = 4c^t(1 - c^t), t = 1,2,3, \dots, m, \tag{3}$$

To introduce chaotic behavior, the researcher has dynamically modified the value of r_1 in a chaotic manner while decreasing it in each iteration [12]. This approach enabled ongoing exploration even in the later stages of each iteration. By implementing this method, the process reduced the range of normalization proportionally with each iteration, as depicted in Eq. (4) of reference.

$$V(t) = MAX - \frac{t}{T}(MAX - MIN) \tag{4}$$

In Eq. (4), the variable t represents the current iteration, while T represents the maximum number of iterations. Additionally, MAX and MIN correspond to the adaptive intervals used in the computation. The study mapped to $r_1 \in (0, V(t))$ in each iteration, which is also decreased. The steps of this $V(t)$ process is visualized above. Note that these experiments utilize $MAX = 20$, and $MIN = 1E\text{-}10$ [12].

Secondly, as FS is a binary optimization problem [13], the transfer function of FS produces binary outputs $\in \{0, 1\}$. A value of zero signifies the rejection of a feature as redundant, while a value of one indicates the selection of a feature as useful. However, it is essential to consider the possibility of the obtained results falling outside the desired range. To maintain the output within the expected range, a binarization function is applied to each agent. In this study, the sigmoid (S-shaped) transfer function, as proposed by [14], is employed for this purpose. The S-shaped transfer function is defined as follows:

$$T(x) = \frac{1}{1 + e^{-x}} \tag{5}$$

$$X^d(t) = \begin{cases} 1 & \text{if } rnd < T(X^d(t)) \\ 0 & \text{if } rnd \geq T(X^d(t)) \end{cases} \tag{6}$$

In the Eq. (5)-(6), range of this function is $\in [0, 1]$. If the output of the transfer function exceeds rnd, where rnd is a random number uniformly distributed in the range (0, 1), the value is set to 1, indicating that the attribute is considered useful. Conversely, if the output is less than or equal to rnd, the value is set to 0, indicating that the attribute is deemed redundant and will not be considered [14].

In conclusion, considering the stochastic nature of metaheuristics, the robustness of a new approach holds significant importance. In this proposed chaotic approach, the adaptive behavior of the attraction power of the objective function has been altered, which contributes to its robustness. The introduction of chaotic changes occurs alongside the adaptive decrease of random parameters. This necessitates the search agents to explore the search space up until the final iteration, albeit with a reduced scale and magnitude of exploration due to the adaptive method. Essentially, there is proposed a novel adaptive chaotic approach where the r1 parameters of the GWO are automatically determined through randomization. This eliminates the need for manual parameter tuning by researchers. It is believed that this approach represents the first systematic attempt to automatically tune the attraction power of the GWO's goal bower, enabling effective optimization in both the exploration and exploitation phases. The process of the ACBGWO algorithm is represented in Algorithm 2.

Algorithm 2: Pseudo Code of ACBGWO algorithm for feature selection

input: $\boldsymbol{NP}$, the population size; $\boldsymbol{T}$, the maximum number of iterations; Dataset.
output: The featured individual $\vec{X}_{\alpha}$, the best fitness value $f(\vec{X}_{\alpha})$.

Initialize a population $X_i(i = 1,2,3, \ldots, n)$ of $\boldsymbol{NP}$ individuals randomly.
Initialize parameters $t = 1$, $\vec{\alpha}$.
Calculate the fitness value of each individual and record the best individual, second-best individual, and the third-best individual as $\vec{X}_{\alpha}$, $\vec{X}_{\beta}$ and $\vec{X}_{\delta}$.
while *Stopping criteria not met* **do**
    Update the adaptively chaotic number, according to Eq. (4)
    Update the chaotic number using the chaotic map equation, according to Eq. (2)-(3)
    for $i = 1: N$ **do**
        Update the position ($\vec{X}_{\alpha}$, $\vec{X}_{\beta}$ and $\vec{X}_{\delta}$) to a binary position, according to Eq. (1)
    end
    Calculate the fitness value of everyone in the population and update $\vec{X}_{\alpha}$, $\vec{X}_{\beta}$ and $\vec{X}_{\delta}$.
    Update the current number of iterations, $t \leftarrow t + 1$.
    Update $\vec{\alpha}$
end
Output feature individual $\vec{X}_{\alpha}$ and $f(\vec{X}_{\alpha})$.
The search population NP's classification accuracy is calculated by dividing the total number of training vectors by the constant t. The KNN classifier's error rate, while employing the features chosen by the search population NP, is calculated as 1 minus the KNN classifier's classification accuracy.

Experimental

The research team has studied and developed a basic algorithm called Grey Wolf Optimizer (GWO), which has been enhanced to be more intelligent for feature selection and is applied to analyze the problem of new smoker initiation among Thai adolescents, ages 15-18. The research team collected a large-scale dataset from 1991 to 2021. For the experimental process, the team used MATLAB R2016a software running on a desktop computer with an Intel® Core™ i7-6770HQ processor, 8.00GB of RAM, 500GB of HD, and a Microsoft Windows 10 Professional 64-bit operating system. This computer was used for data analysis throughout this research study. The values of the parameters utilized by the ACBGWO algorithms are presented in Table 1.1.

Table 1.1

Parameter setting for the ACBGWO algorithm.

Parameter	Value
Number of independent runs	30
Number of iterations	1000
Number of search agents (n)	100
Dimension (d)	59

Table 1.2. Displaying the results of feature selection for the data.

#	Details of the feature selection process																														Chosen
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	
Number of features	27	26	38	30	29	24	32	31	31	21	30	30	21	21	31	32	18	35	30	23	24	20	14	35	28	30	36	32	21	30	**27.67**
Accuracy	99.56	99.63	99.58	99.56	99.56	99.58	99.58	99.58	99.71	99.63	99.67	99.63	99.69	99.63	99.71	99.61	99.71	99.63	99.64	99.64	99.70	99.67	99.66	99.60	99.65	99.62	99.57	99.63	99.63	99.56	**99.63**

Table 1.2 illustrates the results of the feature selection process performed using the ACBGWO algorithm. Each iteration in the table provides details on the number of features selected, the achieved accuracy, and the stability of the model across repeated trials. The final selection of features, presented in the last column, reflects the optimal balance between predictive performance and model simplicity, ensuring both high accuracy and practical interpretability. The selection of the final feature set was justified based on several analytical criteria. First, the accuracy metric was prioritized, with iterations achieving a consistent accuracy of 99.60% or higher being considered optimal. Additionally, the standard deviation of the accuracy values was closely examined to assess model stability, with a low standard deviation is 0.0479 indicating robust performance across iterations. The trade-off between the number of features and model performance was also carefully evaluated to avoid overfitting while retaining interpretability. Consequently, the final selection comprised 21 features, representing a parsimonious yet highly predictive set of variables.

The final feature set encapsulates the most significant social, environmental, and behavioral factors influencing smoking initiation among Thai adolescents. These features were identified through a rigorous optimization process, ensuring that they not only capture the complexity of the phenomenon but also provide actionable insights for targeted interventions. The inclusion of these features in subsequent multivariable models underscores their importance in understanding the predictors of youth smoking behaviors. This comprehensive approach ensures that the selected features are both statistically robust and relevant for practical applications in public health and tobacco control research.

Figure 1. The feature selection process for general data on smoking prevalence among the Thai population.

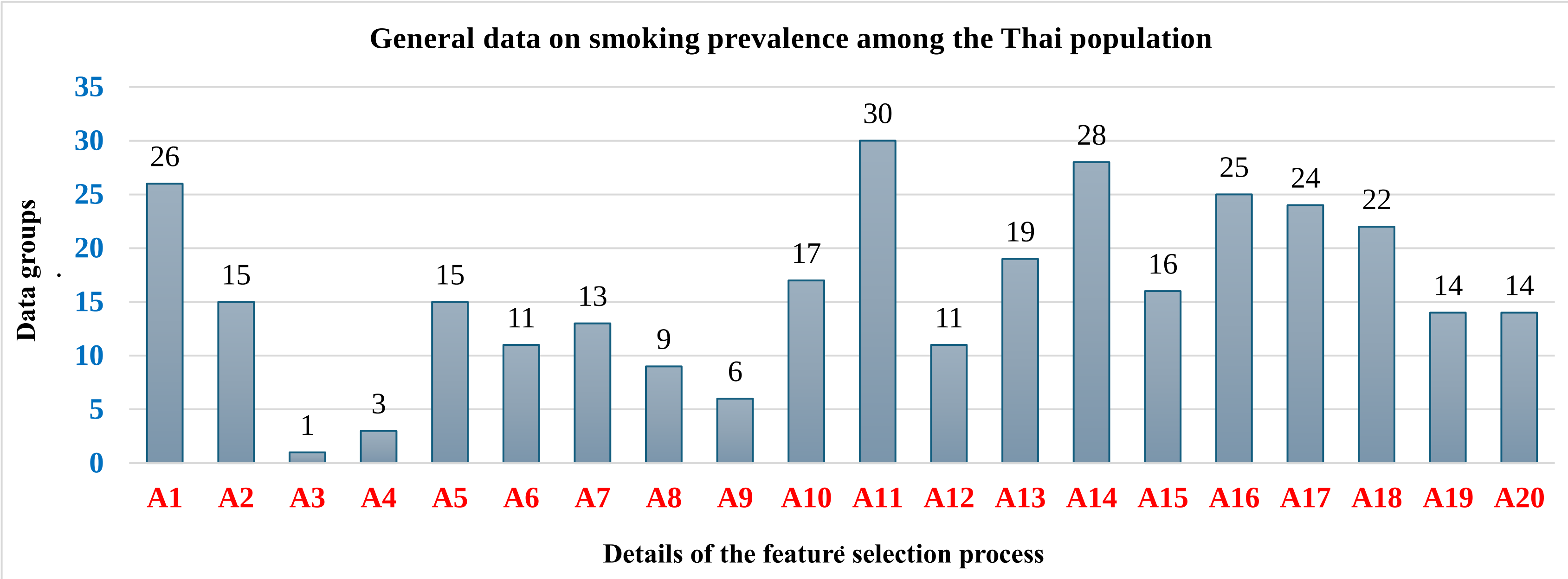

Figure 1 illustrates the feature selection process for general smoking prevalence among the Thai population. The ACBGWO algorithm identified key variables such as regional smoking patterns, which demonstrated significant predictive value with a classification accuracy of 99.63% and a low standard deviation of 0.0479. Features such as gender and age group contributed substantially to distinguishing smoking from non-smoking behaviors, as evidenced by their high selection frequencies during the optimization process.

Figure 2. The feature selection process for have you noticed anyone smoking near you or any cigarette butts in any public place.

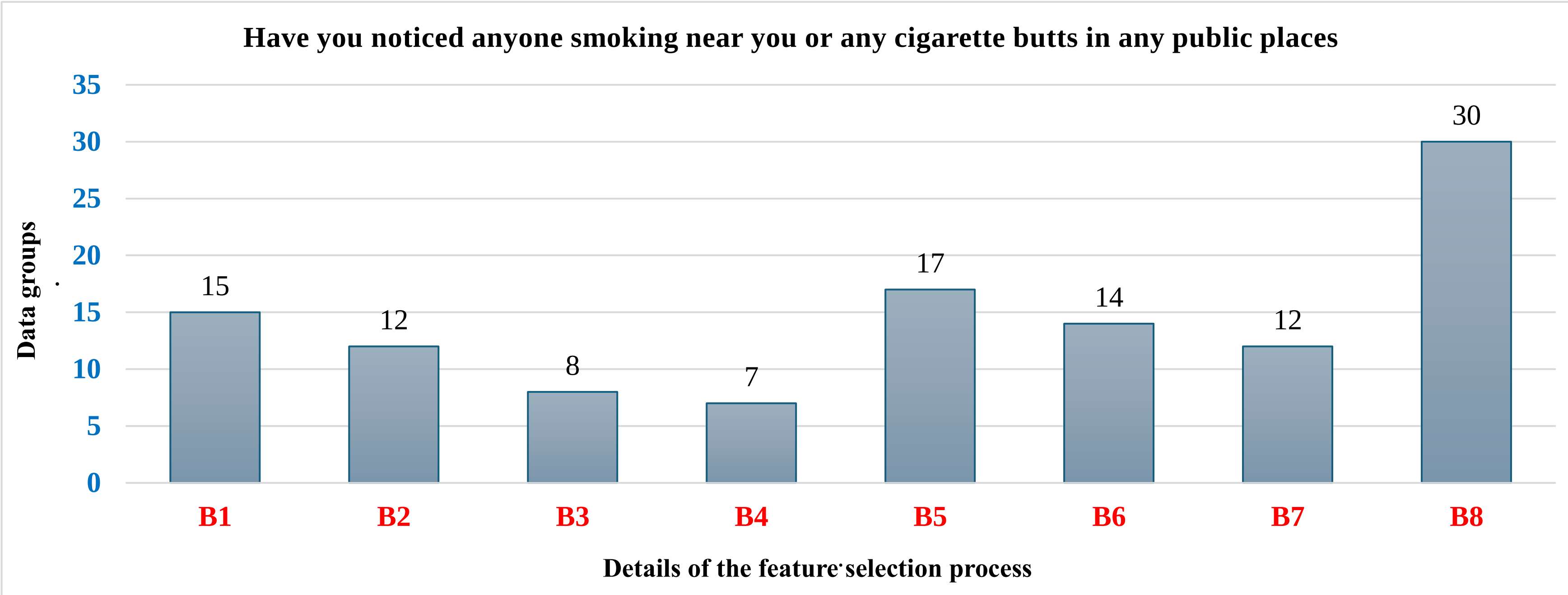

Figure 2 focuses on environmental exposure to smoking cues, such as noticing someone smoking or the presence of cigarette butts in public places. Among the analyzed features, the presence of cigarette butts in fresh markets was reported by 5,621 individuals, highlighting its strong association with smoking initiation. These features consistently showed high relevance in multiple iterations of the algorithm, contributing to the algorithm's overall accuracy.

Figure 3. The feature selection process for have you noticed information about the danger of smoking cigarettes or that encourages quitting smoking in any sources of media.

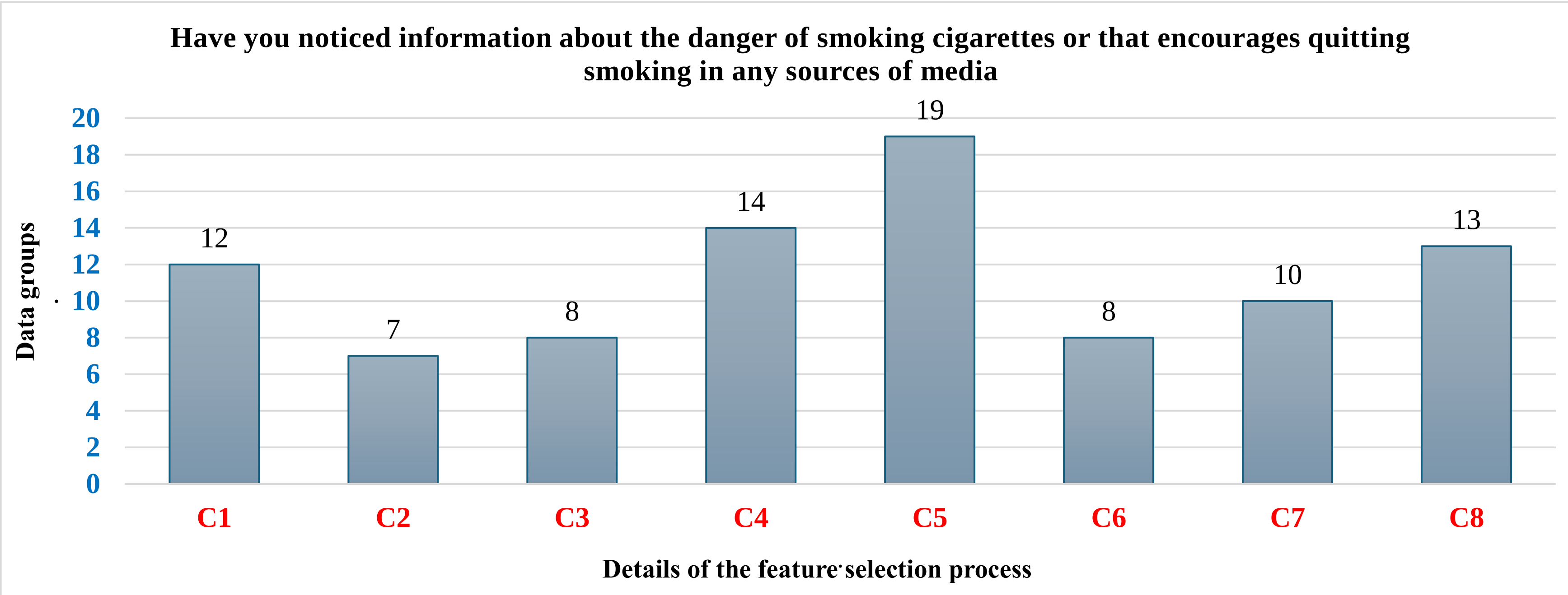

Figure 3 represents the analysis of media influence, particularly anti-smoking messages and warnings. The data revealed that 15.86% (5,719 participants) reported noticing health warnings on cigarette packs, with these warnings significantly associated with lower smoking prevalence. This figure highlights the selection of media-related variables as critical predictors of smoking behavior.

Figure 4. The feature selection process for have you noticed any of following types of cigarettes promotions.

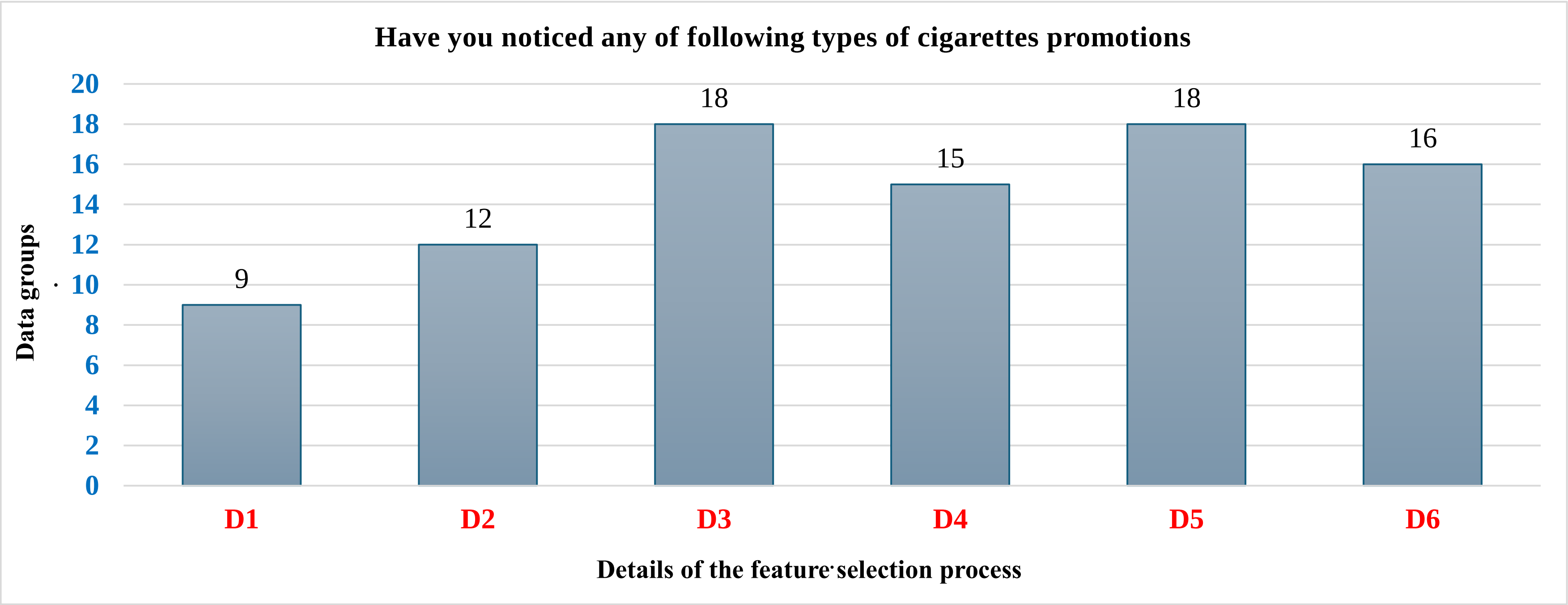

Figure 4 highlights the feature selection process for data on exposure to cigarette promotions. Among adolescents exposed to promotional campaigns, those who reported noticing cigarette promotions showed a higher likelihood of smoking initiation. The feature selection process prioritized these variables, achieving accuracy levels above 99.60% across iterations. This underscores the importance of regulating tobacco advertisements.

Figure 5. The feature selection process for have you noticed any advertisements or signs that encourage smoking in any place.

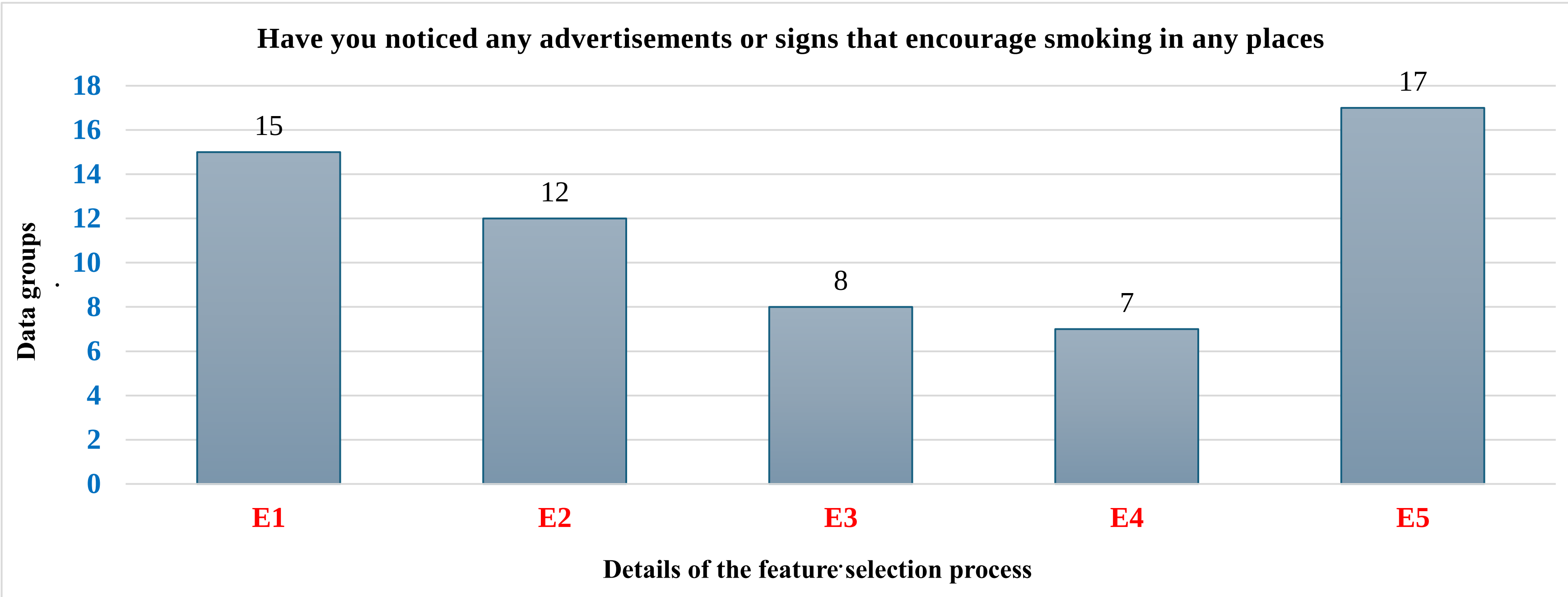

Figure 5 examines the influence of advertisements and signage promoting smoking in various locations. Exposure to such advertisements was significantly correlated with smoking initiation. Statistical analysis revealed that 80.34% of smokers (28,977 participants) reported limited awareness of health warnings due to purchasing cigarettes through informal channels or consuming hand-rolled cigarettes.

Figure 6. The feature selection process for what are any of following diseases caused by smoking tobacco.

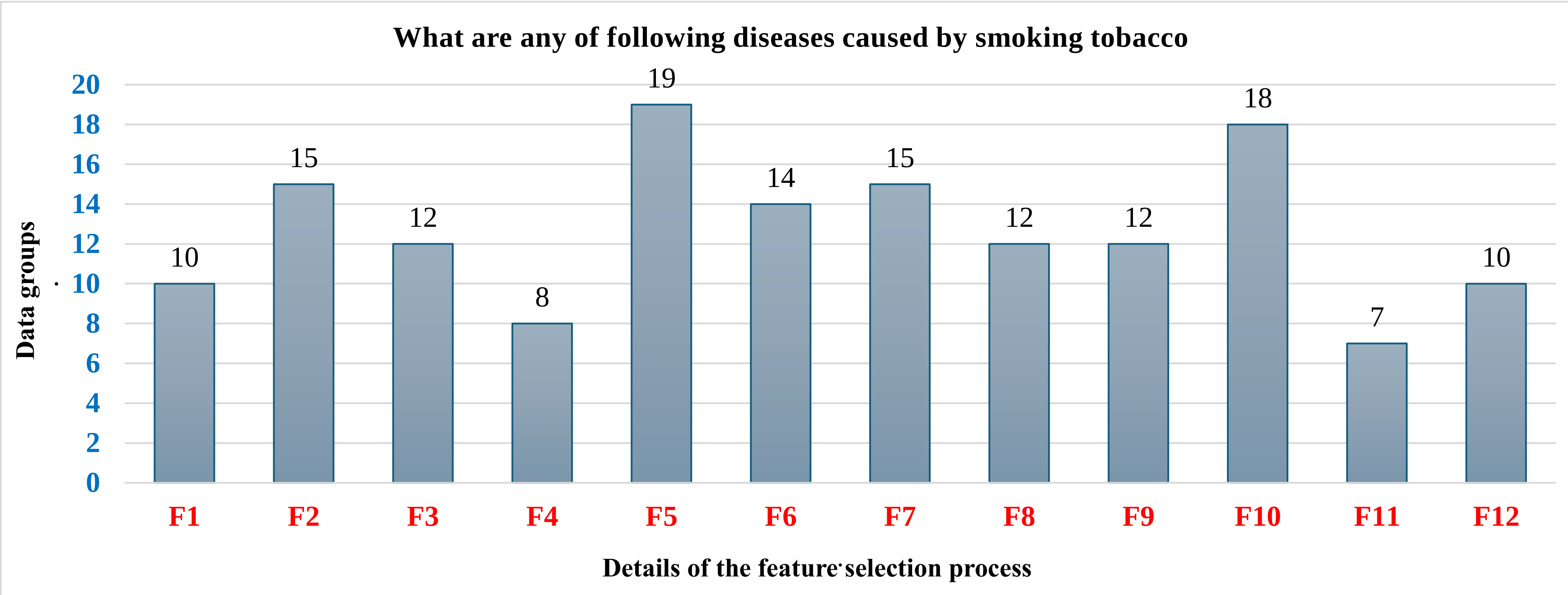

Figure 6 delves into health awareness data, focusing on diseases caused by smoking tobacco. The three most recognized smoking-related diseases COPD (22.27%), laryngeal cancer (21.58%), and oral cancer (21.48%) were identified as strong predictors of health-conscious behavior among adolescents. However, despite high awareness, many participants continued to engage in smoking, highlighting the gap between knowledge and behavioral change.

The statistical results reflected in Figures 1-6 showcase the robustness of the ACBGWO algorithm in identifying and prioritizing critical factors influencing smoking initiation. These insights, combined with high accuracy and stability metrics, provide a valuable framework for designing targeted interventions and public health strategies aimed at curbing adolescent smoking behaviors in Thailand.

1. Bai L, Li H, Gao W, Xie J, Wang H. A joint multiobjective optimization of feature selection and classifier design for high-dimensional data classification. *Inf Sci (N Y)*. 2023;626:457-473. doi:10.1016/j.ins.2023.01.069

2. Han X, Chang X, Quan L, et al. Feature subset selection by gravitational search algorithm optimization. *Inf Sci (N Y)*. 2014;281:128-146. doi:10.1016/j.ins.2014.05.030

3. Wangchamhan T, Chiewchanwattana S, Sunat K. Efficient algorithms based on the k-means and Chaotic League Championship Algorithm for numeric, categorical, and mixed-type data clustering. *Expert Syst Appl*. 2017;90:146-167. doi:10.1016/j.eswa.2017.08.004

4. Majmundar A, Chou CP, Cruz TB, Unger JB. Relationship between social media engagement and e-cigarette policy support. *Addictive Behaviors Reports*. 2019;9. doi:10.1016/j.abrep.2018.100155

5. Lai CM. Multi-objective simplified swarm optimization with weighting scheme for gene selection. *Applied Soft Computing Journal*. 2018;65:58-68. doi:10.1016/j.asoc.2017.12.049

6. Mirjalili S, Mirjalili SM, Lewis A. Grey Wolf Optimizer. *Advances in Engineering Software*. 2014;69:46-61. doi:10.1016/j.advengsoft.2013.12.007

7. Wang J, Lin D, Zhang Y, Huang S. An adaptively balanced grey wolf optimization algorithm for feature selection on high-dimensional classification. *Eng Appl Artif Intell*. 2022;114. doi:10.1016/j.engappai.2022.105088

8. Emary E, Zawbaa HM, Hassanien AE. Binary grey wolf optimization approaches for feature selection. *Neurocomputing*. 2016;172:371-381. doi:10.1016/j.neucom.2015.06.083

9. Kohli M, Arora S. Chaotic grey wolf optimization algorithm for constrained optimization problems. *J Comput Des Eng*. 2018;5(4):458-472. doi:10.1016/j.jcde.2017.02.005

10. Chantar H, Mafarja M, Alsawalqah H, Heidari AA, Aljarah I, Faris H. Feature selection using binary grey wolf optimizer with elite-based crossover for Arabic text classification. *Neural Comput Appl*. 2020;32(16):12201-12220. doi:10.1007/s00521-019-04368-6

11. Hu P, Pan JS, Chu SC. Improved Binary Grey Wolf Optimizer and Its application for feature selection ☆. 2020;195:105746. doi:10.1016/j.knosys

12. Wangkhamhan T. Adaptive Chaotic Satin Bowerbird Optimization Algorithm for Numerical Function Optimization. *Journal of Experimental and Theoretical Artificial Intelligence*. 2020;33(5):719-746.

13. KUSHAL KANTI GHOSH SAPKSZWGARS. Improved Binary Sailfish Optimizer Based on Adaptive β-Hill Climbing for Feature Selection. Published online 2020.

14. Ahmed S, Sheikh KH, Mirjalili S, Sarkar R. Binary Simulated Normal Distribution Optimizer for feature selection: Theory and application in COVID-19 datasets. *Expert Syst Appl*. 2022;200. doi:10.1016/j.eswa.2022.116834

© 2025 Muangkote N. et al.