



Use of Artificial Bee Swarm Optimization (ABSO) for Feature Selection in System Diagnosis for Coronary Heart Disease

Wiharto^{1*}, Yaumi A. Z. A. Fajri², Esti Suryani³, and Sigit Setyawan⁴

^{1,2,3}Department of Informatics, Sebelas Maret University, 57126, Indonesia

⁴Department of Medicine, Sebelas Maret University, 57126, Indonesia

Abstract

The selection of the correct examination variables for diagnosing heart disease provides many benefits, including faster diagnosis and lower cost of examination. The selection of inspection variables can be performed by referring to the data of previous examination results so that future investigations can be carried out by referring to these selected variables. This paper proposes a model for selecting examination variables using an Artificial Bee Swarm Optimization method by considering the variables of accuracy and cost of inspection. The proposed feature selection model was evaluated using the performance parameters of accuracy, area under curve (AUC), number of variables, and inspection cost. The test results show that the proposed model can produce 24 examination variables and provide 95.16% accuracy and 97.61% AUC. These results indicate a significant decrease in the number of inspection variables and inspection costs while maintaining performance in the excellent category.

Index Terms: Bee Swarm Optimization, feature selection, examination fees, coronary heart disease

I. INTRODUCTION

Heart disease is a non-communicable disease that is the leading cause of death worldwide, including in Indonesia. Based on the Basic Health Research (RISKESDAS) data from 2018, the incidence of heart disease has shown an increasing trend, with the prevalence of heart disease in Indonesia at 1.5%. This means that 15 of 1,000 Indonesians suffer from heart disease. Heart disease is still the number one cause of death since Covid-19 and is referred to as a “silent killer”. Most people consider medical checkups after facing significant heart issues. Therefore, it is important for everyone to play a role in preventing the high number of deaths from heart disease. Prevention can be achieved through regular checks. Routine checkups certainly time-consuming and expensive but bring many health benefits.

The development of artificial intelligence has affected the development of diagnostic models for coronary heart disease. Many studies have developed artificial intelligence-based diagnostic models, namely, models that focus on the use of machine learning algorithms for classification. The performance of diagnostic system models with machine learning is mainly determined by the accuracy of the classification algorithm; however, determining the appropriate examination variables is also very important. Determining the correct examination variable requires a suitable feature selection method. The selection of inappropriate features will affect the performance of the diagnostic system model. Feature selection methods have been developed using several approaches, including Wrapper [1]. Feature selection using the Wrapper approach is largely determined by the method used to determine the selected feature subset. The determina-

Received 15 October 2022, Revised 26 March 2023, Accepted 26 March 2023

*Corresponding Author Wiharto (E-mail: wiharto@staff.uns.ac.id)

Department of Informatics, Sebelas Maret University, 57126, Indonesia

Open Access <https://doi.org/10.56977/jicce.2023.21.2.130>

print ISSN: 2234-8255 online ISSN: 2234-8883

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Copyright © The Korea Institute of Information and Communication Engineering

tion of feature subsets in the Wrapper approach was developed using a metaheuristic algorithm [2]. Several metaheuristic algorithms can be used, including genetic algorithms (GAs), Particle Swarm Optimization (PSO), Artificial Bee Swarm Optimization (ABSO), and Artificial Bee Colony (ABC). However, they have both advantages and disadvantages. The accuracy of the chosen algorithm has an impact on the performance of the proposed system.

This paper proposes a coronary heart disease diagnosis model using the ABSO-based feature selection method. The ABSO-based feature selection model uses an objective function that considers system performance and inspection costs. The system performance was measured using the area under curve (AUC) performance parameters, accuracy, number of features, and total inspection costs.

II. LITERATURE REVIEW

Metaheuristic algorithms are inspired by the behaviors of ants, insects, bees, and butterflies. Metaheuristic algorithms that consider bee behavior have been developed and applied in various engineering fields [3-5], mostly numerical optimization. Karaboga et al. [6] proposed an artificial bee colony (ABC) algorithm. In the ABC algorithm, bees attempt to find food sources and advertise them. Onlookers follow their attractive employed bees and scout bees fly spontaneously to find better food sources. Regarding bee behavior, Yang [7] proposed a virtual bee algorithm (VBA). The aim of VBA is to optimize two-dimensional numerical functions using a collection of virtual bees that move randomly in the phase space and interact by searching for food sources that match the coded function values. The intensity of the interaction between these bees yields a solution to the optimization problem. Sundareswaran et al. [8] proposed a different approach based on the natural behavior of honeybees during nectar collection, where randomly generated employed bees are forced to move towards elite bees. This represents the optimal solution [8]. Bees move based on a probabilistic approach. The flight step distance of the bees was used as a variable parameter in the algorithm. Experiments show that the algorithm developed based on the intelligent behavior of honeybees successfully solves numerical optimization problems and provides better performance than a number of population-based algorithms, such as PSO, GA, and ACO [8,9].

The ABC algorithm has several technical weaknesses, including slow convergence and becoming stuck at a local optimum. The improved ABC algorithm is also known as the Bee Swarm Optimization (BSO) algorithm [10]. The BSO algorithm works similarly to the ABC algorithm and is based on the behavior of honeybees foraging for food. The BSO algorithm uses different types of bees to optimize the numer-

ical functions. Each type of bee exhibits a different movement pattern. The scout bees fly randomly over their nearest area. A watcher bee selects an experienced hunter bee because it attracts the elite and moves towards it. Experienced wandering bees remember the best food sources found so far. Bees select the most experienced foragers as elite bees and adjust their positions based on cognitive and social knowledge. The BSO algorithm uses a set of approaches to reduce the stagnation and premature convergence problems [11]. In the feature selection process, BSO can outperform several other metaheuristic algorithms, such as GA, PSO, ACO, and ABC [12-14].

A good way to control coronary heart disease is to perform regular checks. However, routine inspections require time and money. The duration of the examination depends on the number of variables examined, while variables that require low prices but are able to produce optimal diagnostic results are preferable. Many models of coronary heart disease diagnosis systems have been developed using metaheuristic algorithms that significantly optimize the feature selection process. In the feature selection process, a metaheuristic algorithm is used to determine the correct type of inspection attributes. Wiharto et al. [15] proposed a feature selection model using a genetic algorithm with an objective function considering the cost of examination. The proposed model produced an AUC of 95.1% using 20 examination variables; unfortunately, this study tested only one dataset. In this dataset, the feature selection process appears to eliminate the high-cost inspection variables immediately. In the research of Wiharto et al. [15], performance was not significantly different from that of Wiharto et al. [16], who used a stepwise greedy combination with Best First Search (BFS). This model can provide an AUC of 95.4% with a few features but at a much higher cost. A similar study, which resulted in expensive inspection fees and good performance, was conducted by Wiharto et al. [17]. This study used a GA for the feature selection process.

Tama et al. [18] proposed a feature selection model based on PSO. This investigation identified 27 variables for the diagnosis of coronary heart disease. Examining these 27 variables resulted in a high total cost. The resulting AUC performance parameter was 98.7%. The artificial bee colony (ABC) has also been used in the feature selection process [19,20]. Kilic et al. [19] was able to produce 16 examination variables, and the best performance was achieved with an accuracy of 89.44%. The number of features and performance are relatively good; however, if viewed from the cost of inspection, using the selected features requires a relatively high price. This is because the selected features incur high inspection costs. In addition, reference to a number of existing studies confirmed that BSO has better optimization capabilities compared to GA, PSO, ACO, and ABC.

III. SYSTEM MODEL AND METHODS

We developed an ABSO-based feature selection model for a coronary heart disease diagnosis system using the Z-Alizadeh Sani, Cleveland, and Statlog datasets. The datasets can be accessed online at <https://archive.ics.uci.edu/ml/datasets.php>. The examination variables and amount of data for each dataset are listed in Table 1. The examination variables in the dataset confirmed the cost of the examination at the Prodia Surakarta Indonesia Laboratory and Sebelas Maret University Hospital, Surakarta, Indonesia. The examination fee is in the form of Indonesian Rupiah (IDR). In the Z-Alizadeh Sani dataset, one attribute was added, namely, the examination fee; thus, the total number of attributes used was 56. There were 14 attributes in the Cleveland and Statlog datasets. The inspection cost attributes before the feature selection process were normalized using the min-max method. The proposed system model develops a feature selection model using the ABSO algorithm. The ABSO algorithm follows the structure and flight patterns of bees, as shown in Fig. 1, which shows the scout bees walking randomly around their current position. An onlooker bee probabilistically selects an experienced forager bee as the elite bee that attracts and follows it. Experienced forager bees remember their previous information, like the global best bees as elite bees, and update their positions according to social and cognitive knowledge.

Table 1. Datasets

No	Dataset	#Feature	#Instance data
1	Z-Alizadeh Sani	55	303
2	Cleveland	14	303
3	Statlog	14	303

The structure of the bee swarm and its flight path feature selection using ABSO by considering costs is divided into five stages: (1) initialization of the population of bees, (2) initialization of parameters, (3) calculating the objective function, (4) updating bees, and (5) information selection [2,11,21,22].

1. Initial population of bees. At this stage, the bee population is determined, which is a representation of a number of selected alternative features. The bee population comprises experienced foragers, onlookers, and scouts:

$$b = e \cup o \cup s \tag{1}$$

where, e , o , and s represent the collections of experienced forager bees, onlookers, and scouts, respectively. The selected feature set is represented by Equation (2), where each bee, m , represents each feature.

$$\bar{x}(b, m) = (x(b, m_1), x(b, m_2), \dots, x(b, m_D)) \tag{2}$$

The variable $\bar{x}(b, m)$ represents alternative solutions or feasible features of the problem in D -dimensional space S , where $S \subset R^D$.

2. The second step is initializing the parameters, as shown in Equation (3). Determination of the number of bees expressed as $n(b)$, maximum number of iterations as $Iter_{max}$, and initialization of the function:

$$\bar{x}_0(b, m) = Init(m, S) \quad \forall m \in b \tag{3}$$

The variable $Init(m, s)$ refers to the initialization function in the search space S , which is associated with a random bee position.

3. Determination of objective function $f_0(\bar{x}(b, m))$. In the case to be solved, the objective function is a function of the accuracy and cost of examination for each variable examined, $ci(b)$, where i refers to each feature.

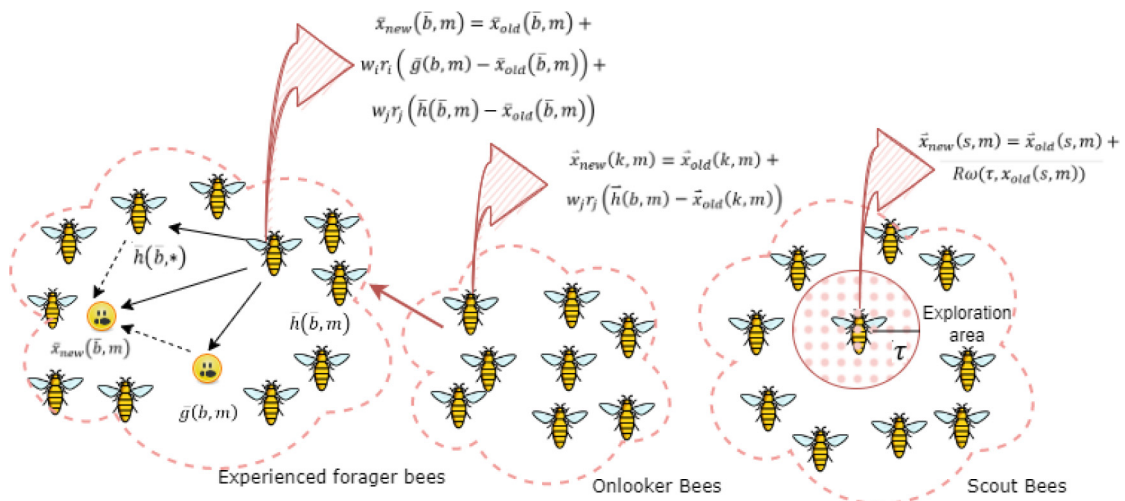


Fig. 1. Structure of the bee swarm and its flight pattern.

The examination fee is calculated as the average cost of each selected examination variable in each feature subset $\bar{x}(b, m)$. The total cost of the variables examined can be written as

$$C(\bar{x}(b, m)) = \frac{\sum_{i=1}^k c_i(b)}{k} \quad (4)$$

Furthermore, by referring to the selected features $\bar{x}(b, m)$, classification was performed using machine learning algorithms. The algorithms tested were SVM, kNN, Random Forest (RF), lightGBM, and XGBoost. The algorithm was used to calculate the accuracy (ACC) performance parameter, which was used as one of the objective function variables. The calculation of its accuracy is given by Eq. (5).

$$f(\bar{x}(b, m)) = \text{ACC} = \frac{\text{TN} + \text{TP}}{\text{TN} + \text{TP} + \text{FN} + \text{FP}} \quad (5)$$

True Positive (TP): When the actual patient is positive, predicted by the system model as positive results. True Negative (TN): When the actual patient is negative, predicted by the system model as negative results. False Positive (FP): When the actual patient is positive, predicted by the system model as negative results. False Negative (FN): When the actual patient is negative, predicted by the system model as positive results. Referring to Eqs. (4) and (5), the ABSO objective function can be written as

$$f_0(\bar{x}(b, m)) = f(\bar{x}(b, m)) - \theta * C(\bar{x}(b, m)) \quad (6)$$

where θ is a weight parameter of the cost effect on evaluation, with values in the range [0,1]. In this study, the value of $\theta = 0.25$ was used, so Eq. (6) becomes Eq. (7).

$$f_0(\bar{x}(b, m)) = f(\bar{x}(b, m)) - 0.25 * C(\bar{x}(b, m)) \quad (7)$$

When the objective function in the ABSO algorithm does not consider costs, it can be expressed as

$$f_0(\bar{x}(b, m)) = f(\bar{x}(b, m)) \quad (8)$$

4. Perform the bee update process. At this stage, the positions of bees change, namely, those of experienced forager bees, onlookers, and scouts.

- a. The position of the experienced forager bee is determined by

$$\begin{aligned} \bar{x}_{\text{new}}(\bar{b}, m) &= \bar{x}_{\text{old}}(\bar{b}, m) + w_i r_i (\bar{g}(\bar{b}, m) - \\ &\bar{x}_{\text{old}}(\bar{b}, m)) + w_j r_j (\bar{h}(\bar{b}, m) - \bar{x}_{\text{old}}(\bar{b}, m)) \end{aligned} \quad (9)$$

where $\bar{x}_{\text{new}}(\bar{b}, m)$ represents the position of the new food source for the bee. Parameters r_i, r_j are random variables with a uniform distribution in the range [0,1], whereas w_i, w_j represent parameters that con-

trol the best food sources found by the m -th and elite bees. Equation (8) for the ABSO algorithm can be explained by dividing the right-hand side into three parts. The first part, $\bar{x}_{\text{old}}(\bar{b}, m)$, shows the position of the old food found by the experienced forager bee. The second part represents the complete knowledge of $w_i r_i (\bar{g}(\bar{b}, m) - \bar{x}_{\text{old}}(\bar{b}, m))$, which pulls the experienced forager bee to the best food position. The third part represents the social knowledge that pulls experienced forager bees to the best position $\bar{h}(\bar{b}, m)$ found by the elite bees.

- b. Experienced forager bees share social knowledge with onlooker bees (k) and update their positions using Equation (10):

$$\bar{x}_{\text{new}}(k, m) = \bar{x}_{\text{old}}(k, m) + w_k r_k (\bar{h}(b, m) - \bar{x}_{\text{old}}(k, m)) \quad (10)$$

where $\bar{x}_{\text{new}}(k, m)$ is the position of the new food source selected by the onlooker bee (m), $w_k r_k$ is a parameter used to control the attractiveness of the bees to the food source, and $\bar{h}(b, m)$ is the elite bee position vector. An onlooker bee uses social knowledge provided by the experienced forager bee to adjust its movement trajectory next time. In each algorithm cycle, the information of the food source and its position provided by the experienced forager bee is shared in the dance area. Subsequently, the onlooker bee evaluates the information provided, uses a probabilistic approach to choose one of these food sources, and follows the experienced forager bee in relation to the selected food source. The probability is calculated as

$$P_n = \frac{f_0(\bar{x}(b, n))}{\sum_{c=1}^{N_b} f_0(\bar{x}(b, c))} \quad (11)$$

where $f_0(\bar{x}(b, c))$ is the value of the objective function of the food source found by experienced forager bees (n), and N_b is the number of experienced forager bees.

- c. The position of the scout bee, s , is fixed using Eq. (12).

$$\bar{x}_{\text{new}}(s, m) = \bar{x}_{\text{old}}(s, m) + R\omega(\tau, \bar{x}_{\text{old}}(s, m)) \quad (12)$$

where $\bar{x}_{\text{old}}(s, m)$ represents the position of the abandoned food source, and $R\omega$ is the random walk function that controls the current position of the scout bee within the search radius τ .

5. Information selection using Eq. (13):

$$\begin{cases} \text{IF } f_0(\bar{x}(b, m)) > f_0(\bar{g}(b, m)) \text{ THEN } \bar{g}(b, m) = \bar{x}(b, m) \\ \text{IF } f_0(\bar{g}(b, m)) > f_0(\bar{h}(b, *)) \text{ THEN } \bar{h}(b, *) = \bar{g}(b, m) \end{cases} \quad (13)$$

where $\bar{g}(b, m)$ is the best food source that the experienced forager bee remembers (m), and $\bar{h}(b, *)$ indicates the position of the best food source found by elite bees.

Several features were obtained from the feature selection process using the ABSO method, and then the classification process was performed. The classification process was performed using the same classification algorithm used to calculate the objective function in ABSO. The classification algorithms are SVM, kNN, RF, lightGBM, and XGBoost. The parameters used to measure the performance of the proposed model were the number of features, total inspection cost, accuracy, and AUC.

IV. RESULTS

The feature selection model testing using ABSO in cases of coronary heart disease diagnosis is divided into two parts. The first is ABSO feature selection with an objective function that does not consider inspection costs. Both objective functions consider the cost of examination. The test results for the ABSO objective function, which do not consider costs, are presented in Tables 2, 4, and 6. The results of the objective function that considers audit fees are presented in Tables 3, 5, and 7. Costs of examinations were determined based on exchange rates for Indonesian Rupiah (IDR). The proposed model was implemented in Python programming using Jupyter Notebook. The model ran on a computer system with an Intel(R) Core (TM) i5-8250U CPU @ 1.60 GHz, 1800 Mhz, 4 Core(s), 8 Logical Processor(s), and 8.0 GHz memory.

Table 2 shows the feature selection without considering cost for the Z-Alizadeh Sani dataset. The best performance was produced on the number of features 22, with a total

Table 2. System performance **without** considering inspection costs (Z-Alizadeh Sani)

Algorithm	ACC	AUC	#Feature	Cost (IDR)
SVM	0.9613	0.9742	22	468,644
kNN	0.9581	0.9594	21	561,108
LightGBM	0.9032	0.9516	24	667,744
RF	0.8548	0.9390	17	643,944
XGBoost	0.8839	0.9003	20	709,408

Table 3. System performance **with** considering inspection costs (Z-Alizadeh Sani)

Algorithm	ACC	AUC	#Feature	Cost (IDR)
SVM	0.9516	0.9761	24	239,294
LightGBM	0.9226	0.9626	29	485,572
RF	0.7516	0.9536	22	363,058
KNN	0.9452	0.9434	17	146,508
XGBoost	0.8742	0.8782	28	135,800

Table 4. System performance **without** considering inspection costs (Cleveland)

Algorithm	#Feature	Accuracy	AUC	Cost (IDR)
LightGBM	7	0.861	0.910	11,800,000
RF	7	0.828	0.906	11,210,000
SVC	6	0.844	0.901	11,085,000
kNN	9	0.818	0.889	10,535,000
XGBoost	4	0.845	0.884	10,095,000

Table 5. System performance **with** considering inspection costs (Cleveland)

Algorithms	#Feature	Accuracy	AUC	Cost (IDR)
RF	9	0.828	0.897	7,135,000
LightGBM	8	0.809	0.896	6,210,000
kNN	9	0.818	0.889	10,535,000
SVC	10	0.821	0.880	6,355,000
XGBoost	10	0.802	0.874	6,480,000

inspection fee of IDR 468,644, AUC performance parameters reaching 97.42%, and an accuracy of 96.13%. This was achieved using the SVM algorithm. If feature selection considers inspection cost, the best performance is obtained with 24 features, and the total inspection fee is IDR 239,294. Diagnosis using these 24 features provided an AUC of 97.61% with an accuracy of 95.16%, as shown in Table 3. This indicates a significant reduction in inspection costs. However, the resulting performance was not significantly different.

Table 4 shows the results of testing using the Cleveland dataset, where the feature selection process did not consider inspection costs. The best performance was obtained with 9 features, with an inspection fee of IDR 11,800,00. The resulting AUC performance was 91% and the accuracy was 86.1%. If the feature selection considers costs, the best performance is achieved when the number of features is 9, with a price of IDR 7,135,000. The use of these nine features provided an AUC performance of 89.7% and accuracy of 82.8%, as shown in Table 5.

The next test used the Statlog dataset. Table 6 shows the test results not considering costs, whereas those that consider price are listed in Table 7. Referring to the two tables, the resulting performances were not significantly different when using the Cleveland dataset. The features in the Cleveland dataset were the same as those in the Statlog dataset; therefore, the only difference was the cost of the inspection results. For feature selection without considering the cost, the number of features required was 6, with AUC performance reaching 90.94%, accuracy 84.44%, and an inspection fee of IDR 11,675,000. If feature selection considers the cost of inspection, it requires a total of 8 features, with a resulting performance of 89% AUC, 82.59% accuracy, and an inspection fee of IDR 6,105,000.

Table 6. System performance **without** considering inspection costs (Statlog)

Algorithm	#Feature	ACC	AUC	Cost (IDR)
LightGBM	6	0.8444	0.9094	11,675,000
kNN	9	0.8440	0.9020	11,820,000
SVC	7	0.8481	0.8811	11,020,000
RF	5	0.8333	0.8678	10,325,000
XGBoost	6	0.8296	0.8667	11,010,000

Table 7. System performance **with** considering inspection costs (Statlog)

Algorithm	#Feature	ACC	AUC	Cost (IDR)
SVC	8	0.8259	0.8900	6,105,000
XGBoost	10	0.8259	0.8875	6,355,000
kNN	9	0.8370	0.8830	7,010,000
LightGBM	7	0.7519	0.8550	1,230,000
RF	6	0.7704	0.8422	1,210,000

The results of testing the feature selection model based on ABSO, where the objective function is a function of accuracy and cost of inspection, show good performance. Referring to the performance parameters, especially AUC, the proposed model can provide relatively the same performance as the feature selection model, which does not consider inspection cost. In addition, the results shown in Tables 2-7 indicate that the proposed model requires a much cheaper total inspection cost with relatively similar performance.

V. DISCUSSION AND CONCLUSIONS

A. DISCUSSION

The ABSO-based feature selection model has relatively good capabilities, both when the feature selection process does and does not consider inspection costs. An ABSO-based feature selection model, when it does not consider cost, tends to choose expensive features; thus, it will require a high inspection cost. This is because it focuses only on one variable of high accuracy, regardless of the costs involved. The cost of an inspection will increase because the examined attributes are high in price; however, in the Z-Alizadeh Sani

dataset, the difference in examination costs is not too high between one feature and another. There is a stark contrast in the Cleveland and Statlog datasets, in which there are two expensive examinations in both datasets: fluoroscopy and Thallium-201 stress scintigraphy. The two examinations are always selected during the feature selection process without considering the cost of the examination. This is because these two attributes are significant in determining the success of heart disease diagnosis. The use of these two examinations will be able to provide a high accuracy, as shown in Table 4, where seven features were selected, including two examinations. Table 6 also shows the same, which requires six features that include both investigations. These results are supported by several previous studies [16,17,23].

Feature selection in the coronary heart disease diagnosis system can be used to select examination attributes that can improve the performance of the diagnosis system [24]. In addition to improving performance, it can also reduce complex computational processes during the classification process. Considering the cost of inspection, the results of system testing using feature selection based on ABSO are summarized in Table 8. Table 8 shows that feature selection using ABSO considering cost results in a larger number of features. This is because in the selection process, when a high-cost feature is obtained, the chance of being selected is lower than that of a low-cost feature. To maintain system performance, other features that are cheaper but have a significant effect on replacing high-cost features will be added. Using this pattern, the performance of the diagnostic system can be maintained. However, the consequence is an increase in the number of features. The addition of a number of features to the proposed feature selection model does not automatically increase the total cost required for inspection. This is because the combined cost of several features is sometimes lower than that of examining a single feature. This results in a higher number of features but lower total inspection cost while maintaining performance. This can be seen in Tables 5 and 7, where the results of feature selection take into account the cost, and Thallium-201 stress scintigraphy examination was not selected but was replaced with another examination at a lower cost.

If we look at the objective function of ABSO shown in Eq. (7), the system performance will be reduced by the magni-

Table 8. System performance comparison summary

Dataset	FS based cost	Method	#Feature	Cost (IDR)	ACC	AUC
z-Alizadeh Sani	No	SVM	22	468,644	96.13%	97.42%
	Yes	SVM	24	239,294	95.16%	97.61%
Clevelands	No	LightGBM	7	11,800,000	86.10%	91.00%
	Yes	RF	9	7,135,000	82.80%	89.70%
Statlog	No	LightGBM	6	11,675,000	84.44%	90.20%
	Yes	kNN	9	7,010,000	83.70%	88.30%

Table 9. Comparison of system performance with previous research

References	Method Feature Selection	Feature	AUC
[16]	CBFS + Greedy Stepwise Algorithm	Typical chest pain, Age, regional wall motion abnormality (Region RWMA), Q-wave, Nonanginal, Blood Pressure (BP), Poor R Progression, Valvular Heart Disease (VHD)	95.40%
[17]	Genetic algorithms + FCBFS	Typical Chest Pain, Diabetes Mellitus (DM), Nonanginal, HTN, Chronic Renal Failure (CRF), Airway disease, Age, Dyspnea, Lung rales, Function Class, Edema, Diastolic Murmur, Low Threshold Angina (Low Th Ang), Family History (FH), Congestive Heart Failure (CHF), Pulse Rate (PR), Weight, Obesity, Sex, Current Smoker.	97.50%
[25]	Random Forest	Typical chest pain, Triglyceride (TG), Body Mass Index (BMI), Age, Weight, BP, Potassium (K), Fasting Blood Sugar (FBS), Length, Blood Urea Nitrogen (BUN), PR, Hemoglobin (HB), Function class, Neutrophil (Neut), Ejection Fraction (EF-TTE), White Blood Cell (WBC), DM, Platelet (PLT), Atypical, FH, High Density Lipoprotein (HDL), Erythrocyte Sedimentation Rate (ESR), Creatine (CR), Low Density Lipoprotein (LDL), T inversion, Dyslipidemia (DLP), Region RWMA, HTN, Obesity, Systolic murmur, Sex, Dyspnea, Current smoker, Bundle Branch Block (BBB), left ventricular hypertrophy (LVH), Edema, Ex-smoker, valvular heart disease (VHD), ST depression, Lymph.	96.70%
[26]	Genetic algorithms and ANN	Typical chest pain, Atypical, Age, Nonanginal, DM, Tinversion, FH, Region RWMA, HTN, TG, PR, Diastolic murmur, Current smoker, Dyspnea, ESR, BP, Function class, Sex, FBS, ST depression, ST elevation, Q-wave	94.50%
[27]	Hybrid feature selection (chi-square, gain ratio, information gain, and relief)	Typical Chest Pain, Atypical, Nonanginal, Region RWMA, EF-TTE, Age, Tinversion, Q wave, VHD, ST Elevation, BP	90.90%
[28]	Ensemble method with PSO	The feature is not shown	92.20%
Proposed	Cost-based ABSO & SVM	Age, Length, BMI, DM, HTN, Current Smoker, Obesity, CRF, Airway disease, CHF, DLP, BP, Weak Peripheral Pulse, Lung rales, Typical Chest Pain, Dyspnea, Function Class, Nonanginal, Exertional Chest Pain, Q Wave, ST Elevation, Tinversion, BBB, TG	97.61%

tude of the normalized total cost of inspection. Based on the calculations from the data in Table 8, the inspection fee can be reduced by an average of 42.81% using the three datasets. The cost reduction was significant, with only an average increase in two features compared to feature selection without considering inspection costs. A feature selection model using ABSO can significantly reduce inspection costs; however, a decrease in inspection costs is accompanied by a reduction in performance. The decline in the average performance from the test results using the three datasets was 1.91%, whereas that for the AUC parameter was 1.11%. This decrease is relatively small; even for the Z-Alizadeh Sani dataset, there was an increase in AUC from 97.42 to 97.61%.

Many studies have been conducted on the use of feature selection in the diagnosis system for coronary heart disease. The feature selection methods used were genetic algorithms, particle swarm optimization, fast correlation-based filter (FCBF) [29], and greedy algorithms [16]. The proposed feature selection model can provide a relatively better performance than those in a number of previous studies. The feature selection model proposed by Kilic & Keles [19], which uses an artificial bee colony combined with Sequential Minimal Optimization (SMO) can only provide an accuracy of 89.4389%, which is much lower than that of the

proposed method. The proposed method was also better than that used by Tama et al. [18]. In this study, a two-tier ensemble PSO method was used for the feature selection. The resulting accuracy was 91.18%. The same was also done by Zomorodi-Moghadam et al. [30] using a hybrid PSO with an accuracy of 84.25%; the value of the performance parameter was still lower than that of the proposed method. In addition, the proposed method was better than that of Babic et al. [31], who used SVM. A complete comparison of the AUC performance parameters with those of previous studies is presented in Table 9. Table 9 shows that the proposed feature selection method has a relatively better performance in terms of AUC. Another advantage of the proposed model is that inspection costs are lower.

B. CONCLUSIONS

The feature selection model using ABSO for the diagnosis of coronary heart disease is able to provide relatively good performance. This performance was indicated by the accuracy of the performance parameters, which reached 95.10%, and the AUC reached 97.61%. When referring to the AUC parameter, the performance of the diagnostic system model shows that the performance is included in the excellent cate-

gory because it is above 90%. This method can reduce the number of features from 55 to 24 for the Z-Alizadeh Sani dataset at a relatively low cost. The same is true for the Cleveland and Statlog datasets, which can eliminate expensive checks by replacing them with cheaper ones while maintaining system performance. For future research, a feature selection model can be developed that is influenced not only by the cost factor but also by other factors, such as the availability of existing health services.

ACKNOWLEDGEMENTS

We thank the Prodia Laboratory and UNS Hospital for providing information on examination costs. In addition, we thank the National Research and Innovation Agency of the Republic of Indonesia, which provided research funding under the Basic Research Grant scheme under Contract No. 469.1/UN27.22/PT.01.03/2022.

REFERENCES

- [1] A. Bommert, X. Sun, B. Bischl, J. Rahnenführer, and M. Lang, "Benchmark for filter methods for feature selection in high-dimensional classification data," *Computational Statistics & Data Analysis*, vol. 143, no. 2020, p. 106839, Mar. 2020, DOI: 10.1016/j.csda.2019.106839.
- [2] R. Akbari, A. Mohammadi, and K. Ziarati, "A novel bee swarm optimization algorithm for numerical function optimization," *Communications in Nonlinear Science and Numerical Simulation*, vol. 15, no. 10, pp. 3142-3155, Oct. 2010, DOI: 10.1016/j.cnsns.2009.11.003.
- [3] N. Karaboga, "A new design method based on artificial bee colony algorithm for digital IIR filters," *Journal of the Franklin Institute*, vol. 346, no. 4, pp. 328-348, May 2009, DOI: 10.1016/j.jfranklin.2008.11.003.
- [4] D. Teodorović, "Bee Colony Optimization (BCO)," *Innovations in Swarm Intelligence*, Springer Berlin Heidelberg, pp. 39-60, 2009, DOI: 10.1007/978-3-642-04225-6_3.
- [5] D. T. Pham, M. Castellani, and A. A. Fahmy, "Learning the inverse kinematics of a robot manipulator using the Bees Algorithm," in *2008 6th IEEE International Conference on Industrial Informatics*, Daejeon, South Korea, pp. 493-498, 2008, DOI: 10.1109/INDIN.2008.4618151.
- [6] D. Karaboga and B. Basturk, "A powerful and efficient algorithm for numerical function optimization: artificial bee colony (ABC) algorithm," *Journal of Global Optimization*, vol. 39, no. 3, pp. 459-471, Oct. 2007, DOI: 10.1007/s10898-007-9149-x.
- [7] X-S. Yang, "Engineering Optimizations via Nature-Inspired Virtual Bee Algorithms," in *Artificial Intelligence and Knowledge Engineering Applications: A Bioinspired Approach, Las Palmas, Canary Islands*, Spain, pp. 317-323, 2005, DOI: 10.1007/11499305_33.
- [8] K. Sundareswaran and V. T. Sreedevi, "Development of novel optimization procedure based on honey bee foraging behavior," in *2008 IEEE International Conference on Systems, Man and Cybernetics*, Singapore, pp. 1220-1225, 2008, DOI: 10.1109/ICSMC.2008.4811449.
- [9] D. Karaboga and B. Akay, "A comparative study of Artificial Bee Colony algorithm," *Applied Mathematics and Computation*, vol. 214, no. 1, pp. 108-132, Aug. 2009, DOI: 10.1016/j.amc.2009.03.090.
- [10] D. Chaudhary, B. Kumar, S. Sakshi, and R. Khanna, "Improved Bee Swarm Optimization Algorithm for Load Scheduling in Cloud Computing Environment," in *Data Science and Analytics*, Springer Singapore, pp. 400-413, 2018, DOI: 10.1007/978-981-10-8527-7_33.
- [11] R. Akbari, A. Mohammadi, and K. Ziarati, "A powerful bee swarm optimization algorithm," in *2009 IEEE 13th International Multi-topic Conference*, Islamabad, Pakistan, pp. 1-6, 2009, DOI: 10.1109/INMIC.2009.5383155.
- [12] S. Sadeg, L. Hamdad, K. Benatchba, and Z. Habbas, "BSO-FS: Bee Swarm Optimization for Feature Selection in Classification," in *Advances in Computational Intelligence*, Springer International Publishing, pp. 387-399, 2015, DOI: 10.1007/978-3-319-19258-1_33.
- [13] H. Djellali, A. Djebbar, N. G. Zine, and N. Azizi, "Hybrid Artificial Bees Colony and Particle Swarm on Feature Selection," in *Computational Intelligence and Its Applications*, Springer International Publishing, pp. 93-105, 2018, DOI: 10.1007/978-3-319-89743-1_9.
- [14] S. Sadeg, L. Hamdad, A. R. Remache, M. N. Karech, K. Benatchba, and Z. Habbas, "QBSO-FS: A Reinforcement Learning Based Bee Swarm Optimization Metaheuristic for Feature Selection," in *Advances in Computational Intelligence*, Springer International Publishing, pp. 785-796, 2019, DOI: 10.1007/978-3-030-20518-8_65.
- [15] Wiharto, E. Suryani, S. Setyawan, and B. P. Putra, "The Cost-Based Feature Selection Model for Coronary Heart Disease Diagnosis System Using Deep Neural Network," *IEEE Access*, vol. 10, pp. 29687-29697, 2022, DOI: 10.1109/ACCESS.2022.3158752.
- [16] Wiharto, E. Suryani, and S. Setyawan, "Framework Two-Tier Feature Selection on the Intelligence System Model for Detecting Coronary Heart Disease," *Ingénierie des systèmes d'information*, vol. 26, no. 6, pp. 541-547, Dec. 2021, DOI: 10.18280/isi.260604.
- [17] W. Wiharto, E. Suryani, S. Setyawan, and B. P. Putra, "Hybrid Feature Selection Method Based on Genetic Algorithm for the Diagnosis of Coronary Heart Disease," *Journal of Information and Communication Convergence Engineering*, vol. 20, no. 1, pp. 31-40, Mar. 2022, DOI: 10.6109/JICCE.2022.20.1.31.
- [18] B. A. Tama, S. Im, and S. Lee, "Improving an Intelligent Detection System for Coronary Heart Disease Using a Two-Tier Classifier Ensemble," *BioMed Research International*, vol. 2020, pp. 1-10, Apr. 2020, DOI: 10.1155/2020/9816142.
- [19] U. Kilic and M. Kaya Keles, "Feature Selection with Artificial Bee Colony Algorithm on Z-Alizadeh Sani Dataset," in *2018 Innovations in Intelligent Systems and Applications Conference (ASYU)*, Adana, Turkey, pp. 1-3, 2018, DOI: 10.1109/ASYU.2018.8554004.
- [20] B. Subanya and R. R. Rajalaxmi, "Feature selection using artificial bee colony for cardiovascular disease classification," in *2014 International Conference on Electronics and Communication Systems (ICECS)*, Coimbatore, India, pp. 1-6, 2014, DOI: 10.1109/ECS.2014.6892729.
- [21] N-C. Yang and D. Mehmood, "Multi-Objective Bee Swarm Optimization Algorithm with Minimum Manhattan Distance for Passive Power Filter Optimization Problems," *Mathematics*, vol. 10, no. 1, p. 133, Jan. 2022, DOI: 10.3390/math10010133.
- [22] A. Askarzadeh and A. Rezazadeh, "Artificial bee swarm optimization algorithm for parameters identification of solar cell models," *Applied Energy*, vol. 102, pp. 943-949, Feb. 2013, DOI: 10.1016/j.apenergy.2012.09.052.
- [23] R. Spencer, F. Thabtah, N. Abdelhamid, and M. Thompson, "Exploring feature selection and classification methods for predicting heart disease," *DIGITAL HEALTH*, vol. 6, p. 205520762091477, Jan. 2020, DOI: 10.1177/2055207620914777.

- [24] M. S. Pathan, A. Nag, M. M. Pathan, and S. Dev, "Analyzing the impact of feature selection on the accuracy of heart disease prediction," *Healthcare Analytics*, vol. 2, p. 100060, Nov. 2022. DOI: 10.1016/j.health.2022.100060.
- [25] J. H. Joloudari, E. H. Joloudari, H. Saadatfar, M. Ghasemigol, S. M. Razavi, A. Mosavi, N. Nabipour, S. Shamshirband, and L. Nadai, "Coronary Artery Disease Diagnosis; Ranking the Significant Features Using a Random Trees Model," *International Journal of Environmental Research and Public Health*, vol. 17, no. 3, p. 731, Jan. 2020. DOI: 10.3390/ijerph17030731.
- [26] Z. Arabasadi, R. Alizadehsani, M. Roshanzamir, H. Moosaei, and A. A. Yarifard, "Computer aided decision making for heart disease detection using hybrid neural network-Genetic algorithm," *Computer Methods and Programs in Biomedicine*, vol. 141, no. 2017, pp. 19-26, ?. 2017. DOI: 10.1016/j.cmpb.2017.01.004.
- [27] B. Kolukisa, H. Hacilar, G. Goy, M. Kus, B. Bakir-Gungor, A. Aral, and V. C. Gungor, "Evaluation of Classification Algorithms, Linear Discriminant Analysis and a New Hybrid Feature Selection Methodology for the Diagnosis of Coronary Artery Disease," in *2018 IEEE International Conference on Big Data (Big Data)*, Seattle, USA, pp. 2232-2238, 2018. DOI: 10.1109/BigData.2018.8622609.
- [28] B. Kolukisa et al., "Coronary Artery Disease Diagnosis Using Optimized Adaptive Ensemble Machine Learning Algorithm," *International Journal of Bioscience, Biochemistry and Bioinformatics*, vol. 10, no. 1, pp. 58-65, 2020. DOI: 10.17706/ijbbb.2020.10.1.58-65.
- [29] B. Senliol, G. Gulgezen, L. Yu, and Z. Cataltepe, "Fast Correlation Based Filter (FCBF) with a different search strategy," in *2008 23rd International Symposium on Computer and Information Sciences*, Istanbul, Turkey, pp. 1-4, 2008. DOI: 10.1109/ISCIS.2008.4717949.
- [30] M. Hassoon, M. S. Kouhi, M. Zomorodi-Moghadam, and M. Abdar, "Using PSO Algorithm for Producing Best Rules in Diagnosis of Heart Disease," in *2017 International Conference on Computer and Applications (ICCA)*, Doha, Qatar, pp. 306-311, Sep. 2017. DOI: 10.1109/COMAPP.2017.8079784.
- [31] F. Babič, J. Olejár, Z. Vantová, and J. Paralič, "Predictive and Descriptive Analysis for Heart Disease Diagnosis," in *2017 Federated Conference on Computer Science and Information Systems (FedCSIS)*, Prague, Czech Republic, pp. 155-163, 2017. DOI: 10.15439/2017F219.



Wiharto

Wiharto is an Associate professor of Computer Science at the Department of Informatics, Sebelas Maret University, Surakarta, Indonesia. He received his Ph.D. degree from Gadjah Mada University, Indonesia in 2017. He is conducting research activities in the areas of artificial intelligence, computational intelligence, expert systems, machine learning, and data mining.



Yaumi A.Z.A. Fajri

Yaumi AZA Fajri is a 2017 undergraduate student of Informatics in the Faculty of Information Technology and Data Science, Universitas Sebelas Maret, Surakarta, Indonesia. Her research interests are swarm intelligence optimization algorithms and data mining.



Esti Suryani

Esti Suryani received her Bachelor of Science (B.S.) degree from Gadjah Mada University, Yogyakarta, Indonesia, in 2002 and Master's degree in computer science from Gadjah Mada University, Yogyakarta, Indonesia, in 2006. She is currently working as an Assistant professor in the Department of Informatics, Faculty of Mathematics and Natural Sciences, Sebelas Maret University, Surakarta, Indonesia. Her experience and areas of interest include image processing and fuzzy logic.



Sigit Setyawan

Sigit Setyawan received his Bachelor of Medicine degree from Sebelas Maret University, Surakarta, Indonesia, in 2005 and Master's degree in medicine from Gadjah Mada University, Yogyakarta, Indonesia, in 2015. He is currently working as an Assistant professor in the Department of Medicine, Faculty of Medicine, Sebelas Maret University, Surakarta, Indonesia. His experience and areas of interest include molecular biology, genomes, and health informatics.