JITK (JURNAL ILMU PENGETAHUAN DAN TEKNOLOGI KOMPUTER)

VOL. 11. NO. 4 MAY 2026
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v11i4.7826

COMPETITIVE COLLABORATION BETWEEN SVM AND NUTCRACKER OPTIMIZATION ALGORITHM FOR CARDIOVASCULAR DISEASE CLASSIFICATION

Florentina Yuni Arini[1*]; Itsna Sabila Hidayati[1]; Januar Pancaran Nur Fajri[1]; Muhammad Alvin Adinata[1]; Bhanu Rizqi Marzaki[1]; Rangga Wicaksana[1]; Poomin Duankhan[2]

Department of Computer Science[1]
Universitas Negeri Semarang, Indonesia[1]
https://unnes.ac.id[1]
floyuna@mail.unnes.ac.id*, itsnasabil@students.unnes.ac.id, januar352@students.unnes.ac.id, alvinadinata1325@students.unnes.ac.id, bhanuqiq01@students.unnes.ac.id, ranggawicaksana@students.unnes.ac.id

College of Computing[2]
Khon Kaen University, Khon Kaen, Thailand[2]
https://eng.kku.ac.th[2]
mini@kkumail.com

(*) Corresponding Author
(Responsible for the Quality of Paper Content)

Abstract—*Cardiovascular disease (CVD) remains the leading cause of death worldwide, emphasizing the need for accurate and reliable diagnostic models. Support Vector Machine (SVM) has demonstrated strong performance in medical data classification due to its ability to handle complex and high-dimensional data; however, its effectiveness depends heavily on appropriate hyperparameter selection. To address this limitation, this study integrates the Nutcracker Optimization Algorithm (NOA), a population-based metaheuristic inspired by squirrel foraging behavior, to optimize SVM hyperparameters for cardiovascular disease prediction. Using a standardized heart disease dataset, three models were evaluated: a baseline SVM, an NOA-optimized model, and an integrated SVM–NOA model. The proposed SVM–NOA approach achieved the best performance, improving accuracy from 83.61% to 88.52%, precision from 78.12% to 86.21%, and F1-score from 83.33% to 87.72%, while maintaining a recall of 89.29%. Although NOA incurs additional one-time optimization cost, the final optimized SVM required only 0.0095 s for training, compared to 0.0205 s for the baseline SVM. These results demonstrate that SVM–NOA provides a robust and computationally practical approach for enhancing cardiovascular disease prediction.*

Keywords: *Cardiovascular Disease, Hyperparameter Optimization, Machine Learning, Nutcracker Optimization Algorithm, Support Vector Machine*

Intisari—*Penyakit kardiovaskular (CVD) masih menjadi penyebab utama kematian di seluruh dunia, sehingga menekankan pentingnya pengembangan metode diagnosis yang akurat dan andal. Support Vector Machine (SVM) telah menunjukkan kinerja yang baik dalam klasifikasi data medis karena kemampuannya menangani data yang kompleks dan berdimensi tinggi; namun, performa SVM sangat bergantung pada pemilihan hyperparameter yang optimal. Untuk mengatasi keterbatasan tersebut, penelitian ini mengintegrasikan Nutcracker Optimization Algorithm (NOA), sebuah metaheuristik berbasis populasi yang terinspirasi oleh perilaku mencari makan tupai, untuk mengoptimalkan hyperparameter SVM dalam prediksi penyakit kardiovaskular. Menggunakan dataset penyakit jantung yang telah distandarisasi, tiga model dievaluasi, yaitu SVM dasar, SVM yang dioptimalkan dengan NOA, dan model terintegrasi SVM–NOA. Model SVM–NOA menghasilkan kinerja terbaik dengan peningkatan akurasi dari 83,61% menjadi 88,52%, precision*

dari 78,12% menjadi 86,21%, serta F1-score dari 83,33% menjadi 87,72%, dengan nilai recall yang tetap sebesar 89,29%. Meskipun proses optimasi NOA memerlukan biaya komputasi tambahan satu kali, model SVM hasil optimasi hanya membutuhkan waktu pelatihan sebesar 0,0095 detik, lebih cepat dibandingkan SVM dasar dengan waktu 0,0205 detik. Hasil ini menunjukkan bahwa integrasi SVM–NOA merupakan pendekatan yang andal dan praktis secara komputasi untuk meningkatkan akurasi prediksi penyakit kardiovaskular.

***Kata Kunci**: Penyakit Kardiovaskular, Optimasi Hyperparameter, Machine Learning, Algoritma Optimasi Nutcracker, Support Vector Machine.*

INTRODUCTION

Cardiovascular disease remains the leading cause of death globally, with an estimated 17.9 million deaths each year, or about 31% of all deaths worldwide [1]. Early detection and accurate diagnosis are crucial in reducing mortality rates, but the complexity of high-dimensional medical data and non-linear relationships between clinical variables often make it difficult for medical professionals to make decisions [2]. Machine learning has shown significant potential in supporting heart disease diagnosis by extracting hidden patterns from patient data, but model performance is highly dependent on feature selection and proper hyperparameter optimization [3]. Conventional approaches that rely on trial and error or grid search require extensive computation time and often fail to produce optimal configurations, necessitating more efficient optimization methods to improve prediction accuracy while maintaining model interpretability.

Classification based on SVM has been applied in various areas such as face recognition, disease diagnosis, text recognition, sentiment analysis, plant disease detection, and intrusion detection systems for network security applications [4]. The kernel trick applied in SVM enables the transformation of non-linear data into a higher-dimensional feature space, so that patterns that cannot be separated linearly in the original space can be classified accurately. However, SVM performance is highly sensitive to parameter (C), kernel coefficient (gamma), and the type of kernel used, where suboptimal configurations can cause overfitting or underfitting [5]. Several studies have proven the effectiveness of SVM in diagnosing heart disease with an accuracy of 80-85%, but these results can still be improved through more sophisticated optimization strategies to explore a wide parameter space [6].

The Nutcracker Optimization Algorithm (NOA) is a relatively new population-based metaheuristic inspired by the behavior of squirrels in cracking nuts and storing food for winter [7]. The nutcracker's behavior can be summarized in a single process where it first gathers and stores pine seeds, then later searches for and retrieves the seeds it previously stored [8]. The advantage of NOA lies in its ability to avoid local optima through adaptive diversification operators and rapid convergence towards global solutions, with lower computational complexity compared to conventional evolutionary algorithms such as Genetic Algorithm or Particle Swarm Optimization [9]. Recent research shows that NOA has superior performance in various optimization benchmark functions and has begun to be applied to hyperparameter tuning for machine learning. Although Support Vector Machines (SVMs) have been extensively employed for cardiovascular disease (CVD) classification and have frequently been integrated with various optimization and meta-heuristic techniques, such as Particle Swarm Optimization [10] and Grey Wolf Optimization [11], to improve performance through enhanced feature selection and hyperparameter tuning, the Nutcracker Optimization Algorithm (NOA) is a recently proposed approach that has primarily been applied to generic optimization problems and has not yet been investigated for enhancing SVM-based CVD classification.

The integration of SVM with the Nutcracker Optimization Algorithm opens up significant opportunities in overcoming the limitations of conventional hyperparameter optimization, which often results in suboptimal configurations for medical classification [12]. The NOA's superior balance between exploration and exploitation makes it an ideal candidate for optimizing sensitive SVM parameters, where small changes in the values of C and gamma can result in substantial differences in accuracy [13]. Recent research shows that nature-inspired metaheuristics can improve SVM performance by 8-12% compared to traditional grid search methods on medical datasets, with a 60-75% reduction in computation time through intelligent search strategies that avoid exhaustive parameter exploration [14]. A comprehensive study on noncommunicable disease datasets demonstrated that a hybrid PSO-based optimization combined with SVM can improve classification accuracy by up to 23.8% compared to approaches without feature selection, with the best performance achieved through the integration of rough set theory and

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

VOL. 11. NO. 4 MAY 2026
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v11i4.7826

particle swarm optimization for simultaneous feature reduction and parameter tuning [14]. NOA's advantage in avoiding premature convergence and its adaptive ability to adjust search behavior based on the landscape fitness function makes it particularly suitable for complex and multimodal SVM optimization landscapes, where multiple local optima can mislead conventional optimization algorithms.

The success of machine learning models in diagnosing cardiovascular disease is highly dependent on the quality and representation of the dataset used for training and validation [15]. A good medical dataset must have characteristics such as adequate sample size. Comprehensive clinical feature variation, and diagnosis labels that have been verified by medical professionals to ensure accurate ground truth [16]. Common challenges in medical datasets include class imbalance, missing values, noise, and variability in measurement scales, which can affect model performance if not handled with proper preprocessing. The selection of datasets that have been standardized and widely accepted in the research community is also important to enable fair comparison of results with previous studies, so that scientific contributions can be measured objectively [17].

The application of metaheuristic algorithms to cardiovascular disease prediction has received considerable attention in recent years, yet several critical gaps remain in current research. First, although systematic reviews have documented the effectiveness of various machine learning models in CVD prediction [18], exploration of newly developed nature-inspired algorithms such as NOA for SVM hyperparameter optimization in this domain is still lacking. Most published work continues to rely on traditional optimizers such as Particle Swarm Optimization, Genetic Algorithm, or conventional Grid Search approaches, potentially missing opportunities for performance improvements offered by newer algorithmic innovations [19]. Second, although hybrid approaches combining metaheuristic optimization with machine learning classifiers have shown promising results in cardiovascular disease detection [20], comparative analyses examining whether direct algorithm integration outperforms sequential optimization-then-classification workflows are still lacking in the literature. Third, computational efficiency considerations including convergence speed, required iterations, and resource utilization have received limited attention in the context of medical datasets, despite their importance for practical clinical applications [21]. These gaps collectively highlight the need for comprehensive investigations that not only evaluate novel optimization algorithms but also examine integration strategies and the feasibility of their practical implementation in real-world healthcare settings.

To address the identified research gaps, this investigation aims to advance the field of cardiovascular disease prediction through three interrelated objectives. The primary objective involves the development and systematic evaluation of three different modeling configurations: a baseline SVM implementation, a NOA-optimized SVM where hyperparameters are tuned separately prior to classification, and a fully integrated SVM-NOA framework where optimization occurs concurrently with model training [22]. This comparative framework allows for a rigorous assessment of whether algorithm integration provides meaningful advantages compared to conventional sequential approaches, a question that has not been adequately addressed by previous hyperparameter tuning studies [23]. The second objective focuses on comprehensive performance benchmarking across multiple evaluation dimensions accuracy, precision, recall, F1-score, and computational efficiency using the standard Heart Disease dataset from the UCI Machine Learning Repository to ensure reproducibility and allow for direct comparison with the existing literature [24]. The third objective involves conducting a detailed feature correlation analysis to identify which clinical variables exhibit the strongest predictive associations with cardiovascular disease outcomes, thus providing insights beyond algorithm performance to support clinical interpretation and potential feature engineering in future work [25]. Together, these objectives form a cohesive research strategy that evaluates not only "whether NOA improves SVM performance" but also "how NOA should be integrated" and "what clinical insights emerge from this approach."

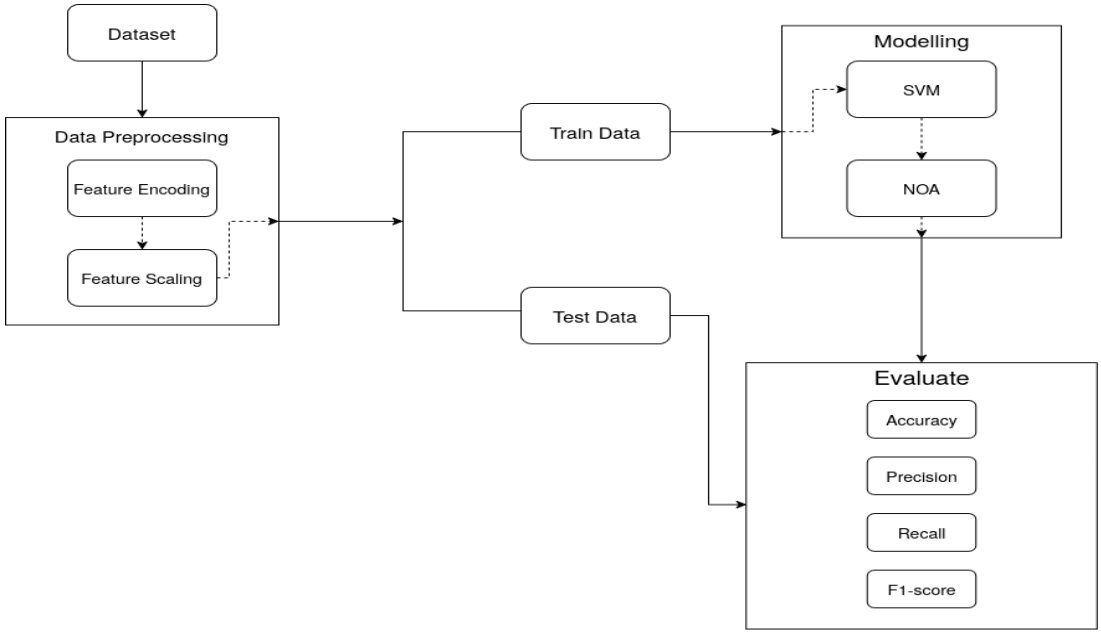
The main aim of this research is to design a cardiovascular disease (CVD) classification framework by collaborating a Support Vector Machine (SVM) with the Nutcracker Optimization Algorithm (NOA). The proposed approach focuses on the simultaneous optimization of SVM hyperparameters and feature subset selection to improve predictive accuracy, model stability, and generalization capability in the presence of complex, high-dimensional cardiovascular data. Moreover, the main contributions of this research are: (a) A competitive–collaborative Support Vector Machine (SVM) and the Nutcracker Optimization Algorithm (NOA) is proposed for cardiovascular

disease classification; (b) Nutcracker Optimization Algorithm tunes critical SVM hyperparameters, reducing manual intervention and improving model performance.

MATERIALS AND METHODS

This research focuses on enhancing the performance of the Support Vector Machine (SVM) by applying the Nutcracker Optimization Algorithm as a reliable method for tuning hyperparameters. By integrating the Nutcracker Optimization Algorithm, this study aims to improve the SVM's learning process, allowing the model to achieve better accuracy and reliability when handling complex and non-linear data sets.

Figure 1 shows the structure of this research which follows a series of steps. It starts with data preprocessing to ensure the data is accurate and consistent, then splits the dataset into training and testing parts. After that, an SVM model is built with optimized parameters using the Nutcracker algorithm to increase classification accuracy. Finally, the model's performance is assessed using various quantitative measures to evaluate its effectiveness and dependability.



Source: (Research Results, 2026)

Figure 1. Proposed Research Flowchart

Dataset Description

This research aims to classify cardiovascular disease using the Heart Disease dataset which was obtained from UCI Machine Learning Repository at https://archive.ics.uci.edu/dataset/45/heart+disease [26]. The dataset contains 303 rows of data and 13 columns of features, which contributed to the outcome of cardiovascular disease classification as the target output. Table 1 shows the Heart Disease dataset features.

Table 1. Heart Disease Dataset Features

Feature	Type	Range of Feature	Description
Age	number	28 - 77	Age in years
Sex	binary	0 - 1	Biological sex, where 1 denotes male and 0 denotes female
Cp	categorical	1 - 4	The category of chest pain experienced
trestbps	number	0 - 200	Resting blood pressure measured in millimeters of mercury (mmHg)
chol	number	0 - 603	Cholesterol level measured in milligrams per deciliter (mg/dl)
fbs	binary	0 - 1	Whether the individual's fasting blood glucose exceeds 120 mg/dl
restcg	categorical	0 - 2	The resting electrocardiographic (ECG) findings
thalach	number	60 - 202	The maximum heart rate achieved

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

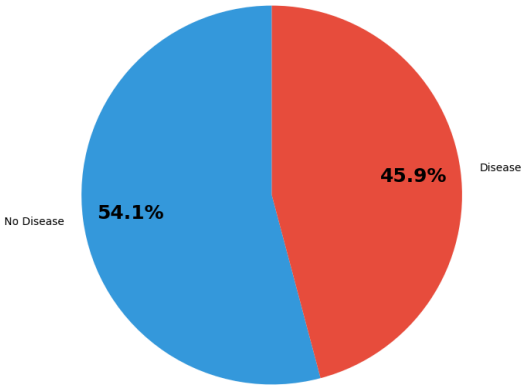
VOL. 11. NO. 4 MAY 2026
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v11i4.7826

Feature	Type	Range of Feature	Description
exang	binary	0 - 1	Whether the individual experienced exercise-induced angina
oldpeak	number	-2.6 - 6.2	The degree of ST-segment depression induced by exercise relative to rest
slope	categorical	1 - 3	The slope of the ST-segment during peak exercise
ca	number	0 - 3	The number of major coronary vessels visualized by fluoroscopy
thal	categorical	3 - 7	The thalassemia status of the individual

Source: (Research Results, 2026)

The features in the Heart Disease dataset can be broadly categorized into demographic attributes, clinical measurements, exercise-induced indicators, and diagnostic test results, as shown in Table 1. Demographic information includes age and biological sex, while clinical measurements such as resting blood pressure, cholesterol level, and fasting blood glucose reflect baseline cardiovascular risk factors. Exercise-related features, including maximum heart rate achieved, exercise-induced angina, ST-segment depression, and ST-segment slope, provide insight into cardiac response under physical stress. Diagnostic attributes derived from electrocardiographic and imaging examinations, such as resting ECG results, the number of major coronary vessels observed by fluoroscopy, and thalassemia status, further contribute to identifying underlying cardiac abnormalities. This structured grouping clarifies the clinical relevance of each feature and supports their suitability for cardiovascular disease classification.

Following the feature-level description, an exploratory analysis was conducted to examine the distribution of the target classes in the dataset. The target classes focus on the distribution of samples based on the heart disease diagnosis which is defined into two classes: Disease and No Disease (as shown in Figure 2). The pie chart in Figure 2 illustrates the proportion of dataset instances belonging to each class. From this visualization, it becomes evident that the dataset does not suffer from severe class imbalance, which is advantageous for developing a stable classification model. This observation also provides an initial indication that both classes are represented sufficiently to support reliable learning during the training phase.



Source: (Research Results, 2026)
Figure 2. Class Distribution of the Target Variable in the Heart Disease Dataset

Additionally, Figure 2 demonstrates a fairly balanced class distribution of heart disease dataset. The Disease class accounts for 54.1% of the data, while the No Disease class represents 45.9%. This suggests that the dataset includes a similar number of individuals without and with cardiovascular disease. While the distribution is not entirely balanced, the difference between the two classes is manageable and does not significantly affect the model training process.

This kind of class distribution is typical in medical datasets, where the occurrence of disease-positive and disease-negative cases can vary slightly. However, both classes are adequately represented, which supports the development and assessment of a reliable model. Additionally, descriptive statistics for each feature are included, offering a summary of the dataset's characteristics. These statistics provide insight into the central tendency and spread of each attribute, aiding in the interpretation of data patterns before proceeding with further preprocessing and modeling steps.

Data Preprocessing

Before developing the model, several preprocessing steps were carried out to prepare the Heart Disease dataset for classification. Initially, the target variable was transformed into a binary format, with 0 indicating individuals without heart disease and 1 indicating those diagnosed with the disease. Missing values in the dataset were then addressed using the mean imputation method [27]. Mean imputation was used as a simple and efficient approach to handle missing values while preserving the dataset size. However, this method may reduce data variability and introduce bias, particularly if the missing values are not completely at random. Therefore, results obtained using mean imputation should be interpreted with caution, and more

robust imputation methods may be considered in future analyses [28].

Following imputation, all numerical features were standardized using the StandardScaler technique to ensure that each feature contributes equally to the model and to prevent attributes with larger numerical ranges from disproportionately influencing model training. The processed dataset was then randomly split into training and testing sets, with 80% of the data allocated for training and 20% for testing [29]. Stratified sampling was applied during this process to preserve the original class distribution in both subsets [30]. Overall, this preprocessing strategy ensures that the dataset is appropriately prepared for classification, supporting stable and reliable baseline model performance in cardiovascular disease prediction.

Support Vector Machine (SVM)

Support Vector Machine (SVM) is a powerful machine learning method that operates based on the principle of structural risk minimization [31]. The Support Vector Machine (SVM) is a supervised learning algorithm based on statistical learning theory that selects key training samples to represent the entire dataset. It offers advantages such as high classification accuracy and short prediction time but has limitations due to its rigid model structure [32]. The Equation (1) present the computational formulation of SVM classification [33].

$$y_i(w^T + b) \geq 1, \forall_i \quad (1)$$

Consider training data consisting of pairs (x_i, y_i) where $x_i \in \mathbb{R}$, $y_i \in \{-1, +1\}$ and w^T denote as transpose of vector w then formula become Equation (2) and Equation (3).

$$\text{if } y_i = +1 \quad (w^T + b) \geq 1 \quad (2)$$
$$\text{if } y_i = -1 \quad (w^T + b) \leq -1 \quad (3)$$

The constraint of Equation (1) can be expressed as two separate conditions since the class label y_i takes only the values +1 and −1. For instances belonging to the positive class, the condition simplifies to Equation (2) which ensures that the samples lie on the positive side of the margin. Conversely, in the negative class, it becomes Equation (3), which places samples on the opposite side of the margin. As a result, this single inequality enforces correct classification for all samples while maintaining a minimum margin from the decision boundary. Moreover, this representation offers an intuitive geometric understanding of data separation in SVM using a maximal margin hyperplane.

Equation (4) presents the dual objective formulation and the kernel function, which together define the optimization structure of SVM and the mechanism used to measure similarity between data points in the transformed feature space.

$$\alpha_i^0 = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j K(x_i, x_j) \quad (4)$$

assigns the Lagrange multipliers α_i as the primary variables responsible for determining each sample's contribution to the decision boundary. The labels y_i and y_j encode how pairs of samples interact based on class membership, while the kernel function $K(x_i, x_j)$ quantifies their similarity in the transformed feature space. The kernel function measures the similarity between two data points x_i and x_j after they are implicitly mapped into a higher-dimensional feature space. This function allows SVM to construct nonlinear decision boundaries without explicitly computing the transformation of the input data.

Nutcracker Optimization Algorithm (NOA)

Nutcrackers exhibit two distinct behavioral phases that occur seasonally, collecting and storing seeds during warmer months, and later retrieving these caches using spatial cues or random search strategies during colder periods and inspired by these adaptive behaviors, the NOA's foraging and storage mechanisms can be mathematically formalized through a series of steps defining the foraging stage [34].

Foraging stage: Exploration phase 1

In this phase, nutcrackers disperse randomly throughout the search space. Each individual begins by evaluating the initial location of a cone containing seeds. If the seeds are deemed suitable, the nutcracker moves to that spot and stores them in a chosen cache area. If no appropriate seeds are found, it continues exploring other cones located on different trees, such as pines. This behavior is mathematically described by Equation (5).

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

VOL. 11. NO. 4 MAY 2026
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v11i4.7826

$$X_i^{new} = \begin{cases} X_{i,j}^{old} \; if \; \tau_1 < \tau_2 \\ \begin{cases} M_t + \gamma \times (A_j - B_j) + \mu \times (r^2 \times U - L), if \; t \le T_{max}/2.0 \\ C_t + \mu \times (A_j - B_j) + \mu \times (r_1 \le \delta) \times (r^2 \times U - L) \end{cases} \end{cases} \quad (5)$$

The term X_i^{new} denotes the updated position of the i^{th} dimension at iteration $t+1$, while $X_{i,j}^{old}$ represents its current position. The parameters U and L define the upper and lower boundaries of the j^{th} dimension, respectively. The variable γ is a random factor determined by the Lévy-flight mechanism. $X_{best,j}^{old}$ indicates the best solution achieved so far in the j^{th} dimension, and A, B, C corresponding to three nutcrackers chosen at random from the entire population. The coefficients τ_1, τ_2, r, r_1 are random values generated within the range [0,1]. μ denotes a parameter generated randomly, as defined in Equation (6). where r_2, r_3 and τ_4 are randomly generated values within the interval 0 to 1. τ_4 follows a normal distribution, while τ_5 is a random value derived using the Lévy-flight mechanism.

$$\mu = \begin{cases} \tau_3 \; if \; r_1 < r_2 \\ \tau_4 \; if \; r_2 < r_3 \\ \tau_5 \; if \; r_1 < r_3 \end{cases} \quad (6)$$

Storage stage: Exploitation phase 1

The nutcrackers initiate the exploitation phase by transferring the food collected during the prior exploration stage to temporary storage sites. In this early stage, they focus on gathering and storing pine seeds. This behavior is mathematically represented by Equation (7).

$$X_i^{new} = \begin{cases} X_i^{old} + \mu \times (X_{best}^{old} - X_i^{old}) \times |\lambda| + r_1 \times (A_t - B_t), if \; \tau_1 < \tau_2 \\ X_i^{old} + \mu \times (A_t - B_t), if \; \tau_1 < \tau_3 \\ best_t \times l \end{cases} \quad (7)$$

where λ is a randomly generated value derived from the Lévy-flight distribution, and l serves as a control parameter that gradually decreases from 1 to 0. During the optimization process, Equation (8) illustrates how the algorithm maintains a balance between the foraging and caching phases to achieve an effective trade-off between exploration and exploitation.

$$X_i^{new} = \begin{cases} Equation \; (5), if \; \varphi < P_{a_i} \\ Equation \; (7) \end{cases} \quad (8)$$

where φ is a randomly generated value within the range 0 to 1 and P_{a_i} represents a probability that decreases linearly from 1 to 0.

Cache-search stage: Exploration phase 2

In the NOA framework, each nutcracker utilizes two reference points $(\boldsymbol{RP_s})$ to mark every cache location, as formulated in Equation (9). These reference points guide the search process by defining directional cues that influence the movement of each nutcracker within the solution space. This mechanism ensures that the algorithm balances exploration and exploitation effectively, thereby improving the likelihood of identifying optimal hyperparameters.

$$RP_s = \begin{cases} \overrightarrow{RP_{1,1}^t} \; \overrightarrow{RP_{1,2}^t} \\ \vdots \quad \vdots \\ \overrightarrow{RP_{i,1}^t} \; \overrightarrow{RP_{i,2}^t} \\ \vdots \quad \vdots \\ \overrightarrow{RP_{N,1}^t} \; \overrightarrow{RP_{N,2}^t} \end{cases} \quad (9)$$

Here, $RP_{i,1}^t$ represents the first reference point (RP) for the i^{th} nutcracker at iteration t. Two distinct formulations are used to construct the first and second reference points (RP_s), each contributing to improved exploration efficiency as the nutcracker searches for hidden caches. The first RP is computed using Equation (10), whereas the second RP is derived from Equation (11).

$$RP_{i,1}^t = \begin{cases} X_i^t + (\alpha \times \cos(\theta) \times (X_A^t - X_A^t) + \alpha \times RP), if \; \theta = \pi/2 \\ X_i^t + \alpha \times \cos(\theta) \times (X_A^t - X_A^t) \end{cases} \quad (10)$$

$$RP_{i,1}^{t} = \begin{cases} X_i^t + \left(\alpha \times \cos(\theta) \times \left((U - L) \times \tau_3 + L\right) + \alpha \times RP\right) \times U_2, if\ \theta = \pi/2 \\ X_i^t + \alpha \times \cos(\theta) \times (U - L \times \tau_3 + L) \times U_2 \end{cases} \quad (11)$$

Recovery stage: Exploitation phase 2

In the first case, the nutcracker is able to remember the cache location through the use of the first reference point (RP), and this behavior is mathematically described by Equation (12).

$$X_{i,j}^{new} = \begin{cases} X_{i,j}^{old}, if \tau_3 < \tau_4 \\ X_{i,j}^{old} + r_1 \times \left(X_{best,j}^{old} - X_{i,j}^{old}\right) + r_2 \times \left(RP_{i,1}^{t} - C_j\right) \end{cases} \quad (12)$$

where r_1, r_2, τ_3 and τ_4 are randomly generated values within the range of 0 to 1, and C represents the index of a randomly chosen solution from the current population. In the alternative case, the nutcracker fails to recall the cache location indicated by the first reference point (RP) and instead depends on the second RP to find it. During the initial storage phase, nutcrackers store multiple RPs linked to their caches. Although they often succeed in retrieving the cache on their first attempt using the primary RP, the proposed model incorporates the possibility of retrieval errors. As stated in Equation (13), when a nutcracker cannot find its cache using the first RP, it triggers the spatial memory associated with the second RP.

$$X_i^{new} = \begin{cases} X_i^{old}, if\ f\left(X_i^{old}\right) < f\left(RP_{i,2}^{t}\right) \\ RP_{i,2}^{t} \end{cases} \quad (13)$$

In the NOA algorithm, nutcrackers maintain their current position whenever the quality of their present solution is superior to the newly produced one. This principle is mathematically represented in Equation (14).

$$X_i^{new} = \begin{cases} X_i^{new}, if\ f(X_i^{new}) < f\left(X_i^{old}\right) \\ X_i^{old} \end{cases} \quad (14)$$

To clearly illustrate how this selection principle is implemented in complete procedural steps of the Nutcracker Optimizer are summarized in the following pseudocode as shown in Table 2.

Table 2. Nutcracker Optimization Algorithm

Stage 0: Initialization
Input: Objective function *f(x)*
Number of nutcrackers *N*
Maximum iterations *T*
Lower bound *lb*, Upper bound *ub*
Control parameters (δ, P_{a1}, P_{a2}, P_{rp})
Output: Best solution X_{best}
Best fitness f_{best}

Stage 0: Initialization (Continue.)
Initialize population X_i (i = 1, 2, ... N) randomly within [*lb, ub*]
Evaluate fitness $f_i = f(X_i)$ for all nutcrackers
Determine initial best solution X_{best} and f_{best}
Store initial convergence and population diversity
For t = 1 to T do
  Update adaptive probability P_{a1}
Stage 1: Foraging or Storage Phase
  Generate random number $\phi \in [0,1]$
  If $\phi > P_{a1}$ **then**
    For each nutcracker *i* **do**
      Select three distinct nutcrackers *A,B,C*
      Generate random parameters
      If $t \leq \frac{T}{2}$ **then**
        Update position using population mean and Lévy flight
      Else
        Update position using difference of nutcrackers
      End If
      Apply boundary constraints
    End For
  Else
    For each nutcracker *i* **do**
      Select two distinct nutcrackers *A,B*
      Generate random parameters τ_1, τ_2, τ_3
      If $\tau_1 < \tau_2$ **then**
        Move towards global best using Lévy flight
      Else If $\tau_1 < \tau_3$ **then**
        Perform Gaussian random exploration
      Else
        Shrink solution based on iteration progress
      End If
      Apply boundary constraints
    End For
  End If
  Evaluate fitness of updated population
Stage 2: Reference Point Generation
  For each nutcracker *i* do
    Generate reference points RP_1 and RP_2
    Use random angles, population differences, and probability P_{rp}
    Apply boundary constraints
  End For
Stage 3: Cache Search or Recovery Phase
  Generate random number $\phi_2 \in [0,1]$
  If $\phi_2 > P_2$ **then**
    For each nutcracker *i* **do**
      Update position using recovery strategy
      Move towards global best and reference points
    End For
  Else
    For each nutcracker *i* **do**

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

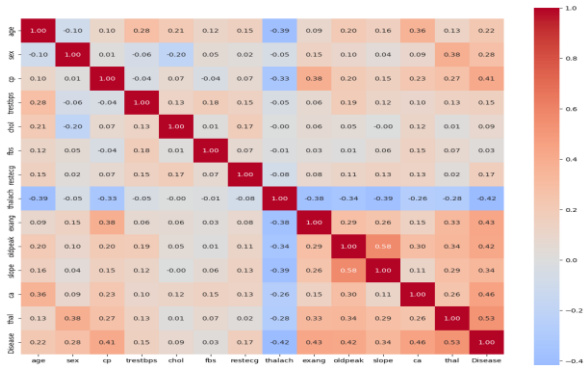
VOL. 11. NO. 4 MAY 2026
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v11i4.7826

Stage 3: Cache Search or Recovery Phase (Continue.)	
47.	Compare current solution with reference points
48.	Replace solution if reference point improves fitness
49.	**End For**
50.	**End If**
51.	Evaluate fitness of updated population
Selection Phase	
52.	**For each** nutcracker i **do**
53.	Select the best solution
54.	Update X_i and f_i
55.	Update global best X_i, f_i if improved
56.	**End For**
57.	Store convergence curve and population diversity
58.	**End For**
59.	**Return** X_{best}, f_{best}

Source: (Research Results, 2026)

RESULTS AND DISCUSSION

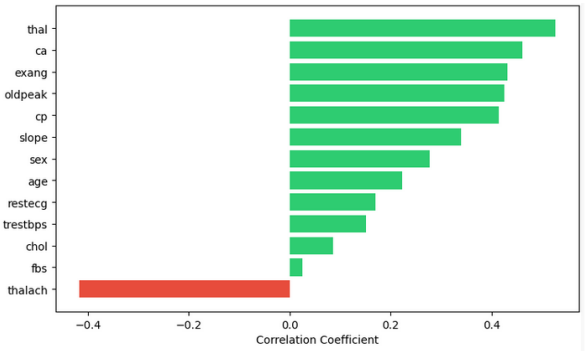
The ability of Support Vector Machine (SVM) models to make accurate predictions can change a lot based on how their hyperparameters are set, especially when working with complicated biomedical data. In medical classification problems such as cardiovascular disease prediction, inappropriate parameter settings can lead to underfitting or overfitting, which may reduce the reliability of the model's decisions. To address this issue, the Nutcracker Optimization Algorithm (NOA) was used in this study to adjust the SVM hyperparameters in a systematic and data-driven manner. Before starting the optimization process, a correlation analysis was carried out to examine how the different features are related to one another and to identify which variables have the greatest influence on the classification results. Understanding these feature relationships is important because strongly correlated variables may contain overlapping information, while weakly correlated yet informative features can contribute unique predictive value. This analysis also helps explain how each feature may affect the model's decision-making process and provides insight into the structure of the dataset used.



Source: (Research Results, 2026)
Figure 3. Feature correlation heatmap

Figure 3 presents a correlation heatmap of all attributes in the Heart Disease dataset, offering a clear visual overview of the interactions among features and serving as a foundation for further analysis of feature relevance and model behavior in the subsequent sections.

As shown in Figure 3, the heatmap displays the correlation between all features in the Heart Disease dataset, providing a clear overview of how strongly and in which direction different variables are related to one another. Notably, variables such as thal, ca, cp, exang, and oldpeak demonstrate stronger connections with the disease label, suggesting their relevance in predicting heart disease, while other features like chol and fbs exhibit weaker relationships. Overall, this visualization supports a better understanding of feature interactions and their potential contribution to prediction accuracy, which is further examined in Figure 4.



Source: (Research Results, 2026)
Figure 4. Feature correlation with target output

The thalach feature shows the strongest link with cardiovascular disease, followed by ca, exang, and oldpeak, as illustrated in Figure 4. This pattern suggests that these variables play a major role in the model's ability to distinguish between people who have the condition and those who do not. This implies that changes in these characteristics may indicate more direct signs of heart disease. Their strong connection to the target variable shows they are important and supports their inclusion in the feature selection and model improvement processes. On the other hand, the thalach feature has a negative relationship with the disease outcome, meaning that individuals with higher maximum heart rates during exercise are more likely to be classified as not having heart disease. This finding matches what is expected in clinical settings, as lower exercise capacity and reduced peak heart rate are often linked to cardiovascular issues. These results highlight the different impacts of various features and show how analyzing

correlations can help in understanding and improving predictive models.

Building on the feature relationship analysis, Table 3 presents how the performance of the SVM model changes before and after applying the Nutcracker Optimization Algorithm. It compares the model's performance using accuracy, precision, recall, and F1-score, which together give a clear picture of how well the model can correctly identify people with cardiovascular disease from those without it. In this study, the Nutcracker Optimization Algorithm was used to adjust the SVM's hyperparameters with the goal of improving prediction results while keeping the same input features unchanged. The analysis focuses only on optimizing the hyperparameters, which helps ensure that any improvement in performance is due to the optimization process itself and not because of changes in the features or the data preparation steps.

Table 3. Model results

Model	Accuracy	Precision	Recall	F1-score
SVM	0.8361	0.7812	0.8929	0.8333
NOA	0.8689	0.8333	0.8929	0.8621
SVM-NOA	0.8852	0.8621	0.8929	0.8772

Source: (Research Results, 2026)

As presented in Table 3, the baseline SVM model achieved an accuracy of 83.61%, with a precision of 78.12%, a recall of 89.29%, and an F1-score of 83.33%. After applying hyperparameter optimization using the Nutcracker Optimization Algorithm, the SVM–NOA model achieved the highest overall performance, improving accuracy to 88.52%, precision to 86.21%, and F1-score to 87.72%, while maintaining the same recall value. These results indicate that NOA effectively enhances the SVM classifier by identifying more suitable hyperparameter configurations without altering the original feature set, thereby improving predictive capability.

While these metric improvements indicate enhanced classification capability, performance values obtained from a single train–test split may still be influenced by sampling variability. To ensure that the observed improvements are robust and not solely attributable to random data partitioning, a statistical validation was performed using bootstrap resampling.

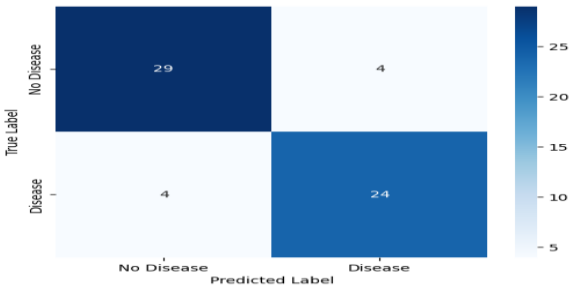
Table 4. Bootstrap confidence intervals for model accuracy

Model	Mean Accuracy	95% Confidence Interval
SVM	0.8359	[0.7377, 0.9180]
SVM-NOA	0.8871	[0.8033, 0.9672]

Source: (Research Results, 2026)

Using bootstrap resampling [35], 95% confidence intervals were estimated for the accuracy of both the baseline SVM and the NOA-optimized SVM models. As shown in Table 4, the NOA-optimized SVM achieves a higher mean accuracy, with its confidence interval consistently shifted toward higher values compared to the baseline model. Although partial overlap exists between the intervals, the improvement remains stable across resampled datasets, indicating that the performance gain is not merely a consequence of random variation. These results provide statistical support for the effectiveness of the Nutcracker Optimization Algorithm in improving SVM-based cardiovascular disease prediction. While the bootstrap analysis confirms that the observed performance improvements are statistically robust.

To better understand how the model performs in predicting outcomes beyond just numbers, a confusion matrix was analyzed for each classifier setup. A confusion matrix [36] is a statistical tool used to evaluate the performance of a classification model by systematically comparing predicted class labels with the corresponding true labels. It summarizes classification outcomes by categorizing predictions into true positives, true negatives, false positives, and false negatives, thereby enabling detailed analysis of model behavior and misclassification errors [37]. Therefore, Figure 5, Figure 6 and Figure 7 shows the confusion matrix comparison between models.



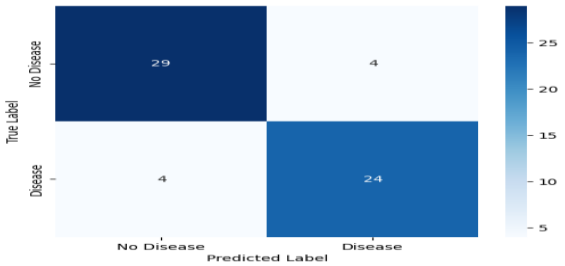
Source: (Research Results, 2026)

Figure 5. Baseline SVM Classifier Confusion Matrix

The confusion matrix for the baseline SVM classifier in Figure 5 shows the model's performance before any optimization. Out of all the samples, 28 were correctly classified as No Disease and 25 as Disease. However, 5 samples were incorrectly labeled as No Disease and 3 were misclassified as Disease. These results match the baseline accuracy of 83.61%, showing that the model performs adequately but still has some errors, especially in the No Disease group. This suggests that the classifier's decision boundary could be improved to enhance its overall performance and generalization ability.

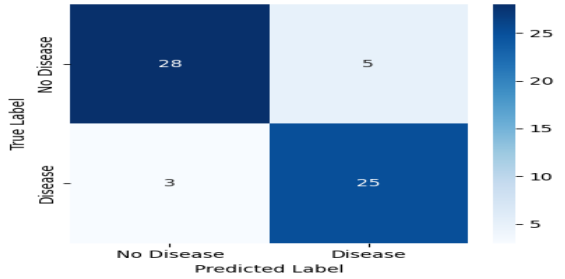
Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

VOL. 11. NO. 4 MAY 2026
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v11i4.7826



Source: (Research Results, 2026)
Figure 6. NOA Classifier Confusion Matrix

The second confusion matrix in Figure 6 shows the performance of the model that was optimized only using the Nutcracker Optimization Algorithm (NOA). When compared to the baseline model, the performance remains similar, with comparable counts of correct and incorrect predictions. This indicates that NOA alone does not significantly alter classification outcomes unless it is directly integrated into the SVM hyperparameter optimization process. This similarity indicates that although NOA improves the ability to search for optimal parameters, its maximum effectiveness might not be achieved without being directly combined with the optimization process of the SVM classifier. Nonetheless, it verifies the algorithm's reliability and its capacity to maintain baseline performance without any decline.



Source: (Research Results, 2026)
Figure 7. SVM-NOA Classifier Confusion Matrix

The final confusion matrix in Figure 7 represents the SVM classifier that was optimized using the Nutcracker Optimization Algorithm, which showed the best overall performance. The number of No Disease cases classified correctly rose to 29, with only 4 errors, while the number of Disease cases correctly identified stayed at 25, with 3 incorrect classifications. Although the numerical improvement appears small, it indicates a significant boost in the model's accuracy and dependability. This shows that applying the Nutcracker Optimization Algorithm directly for tuning SVM hyperparameters enhances the classifier's ability to differentiate between patients with and without cardiovascular disease.

Overall, the results clearly show that combining the Nutcracker Optimization Algorithm with the Support Vector Machine classifier offers noticeable advantages in predicting cardiovascular disease. The improvement in classification performance and the better results seen in the confusion matrices show that fine-tuning the model's parameters leads to more consistent and accurate decision-making.

CONCLUSION

This study demonstrates that appropriate hyperparameter tuning of the Support Vector Machine (SVM) significantly improves its ability to predict cardiovascular disease. When the Nutcracker Optimization Algorithm (NOA) was applied, the model's accuracy increased from 83.61% to 88.52%. Precision also improved from 78.12% to 86.21%, and the F1-score rose from 83.33% to 87.72%, while recall remained stable at 89.29%. Bootstrap analysis confirmed that these improvements were consistent and not due to random data splitting, with the optimized model achieving a higher mean accuracy of 0.8871 compared to 0.8359 for the standard SVM. Correlation analysis identified important clinical features such as *thalach*, *ca*, *exang*, and *oldpeak*, indicating that the model relies on medically meaningful variables when making predictions. In practical terms, these findings suggest that the optimized SVM–NOA model can support clinicians by providing more reliable decision assistance, particularly for early detection and risk assessment of heart disease in real-world healthcare environments. Despite these positive results, the study has several limitations. The findings are based on a single dataset with a limited sample size, which may restrict their generalizability to larger or more diverse populations. Additionally, this limitation may contribute to a lack of detailed analysis regarding model interpretability for clinical use. Future studies could address these issues by evaluating the model on larger and more diverse datasets, incorporating explainable artificial intelligence techniques, and combining NOA with ensemble or deep learning methods to further enhance prediction performance in real-world medical settings.

REFERENCES

[1] A. Timmis et al., "European Society of Cardiology: the 2023 Atlas of Cardiovascular Disease Statistics," European Heart Journal,

vol. 45, no. 38, pp. 4019–4062, Oct. 2024, doi: 10.1093/eurheartj/ehae466.

[2] B. Wei, J. Wu, X. Su, Q. Huang, Y. Liu, and F. Zhang, "Efficient erasure-coded data updates based on file class predictions and hybrid writes," *Comput. Electr. Eng.*, vol. 104, p. 108441, Dec. 2022, doi: 10.1016/j.compeleceng.2022.108441.

[3] G. Alwakid, F. Ul Haq, N. Tariq, M. Humayun, M. Shaheen, and M. Alsadun, "Optimized machine learning framework for cardiovascular disease diagnosis: a novel ethical perspective," *BMC Cardiovasc. Disord.*, vol. 25, no. 1, p. 123, Feb. 2025, doi: 10.1186/s12872-025-04550-w.

[4] D. Mustafa Abdullah and A. Mohsin Abdulazeez, "Machine Learning Applications based on SVM Classification A Review," *Qubahan Acad. J.*, vol. 1, no. 2, pp. 81–90, Apr. 2021, doi: 10.48161/qaj.v1n2a50.

[5] M. O. Oyediran, O. S. Ojo, I. A. Raji, A. E. Adeniyi, and O. J. Aroba, "An optimized support vector machine for lung cancer classification system," *Front. Oncol.*, vol. 14, p. 1408199, Dec. 2024, doi: 10.3389/fonc.2024.1408199.

[6] Wiharto, E. Suryani, and S. Setyawan, "Framework Two-Tier Feature Selection on the Intelligence System Model for Detecting Coronary Heart Disease.," *Ingénierie Systèmes Inf.*, vol. 26, no. 6, p. 541, Dec. 2021, doi: 10.18280/isi.260604.

[7] A. A. Ahmad and H. Polat, "Prediction of Heart Disease Based on Machine Learning Using Jellyfish Optimization Algorithm," *Diagnostics*, vol. 13, no. 14, p. 2392, Jul. 2023, doi: 10.3390/diagnostics13142392.

[8] V. Kumar *et al.*, "A novel nature-inspired nutcracker optimizer algorithm for congestion control in power system transmission lines," *Energy Explor. Exploit.*, vol. 42, no. 6, pp. 2056–2091, Nov. 2024, doi: 10.1177/01445987241253292.

[9] M. Abdel-Basset, R. Mohamed, M. Jameel, and M. Abouhawwash, "Nutcracker optimizer: A novel nature-inspired metaheuristic algorithm for global optimization and engineering design problems," *Knowl.-Based Syst.*, vol. 262, p. 110248, Feb. 2023, doi: 10.1016/j.knosys.2022.110248.

[10] E. I. Elsedimy, S. M. M. AboHashish, and F. Algarni, "New cardiovascular disease prediction approach using support vector machine and quantum-behaved particle swarm optimization," *Multimed. Tools Appl.*, vol. 83, no. 8, pp. 23901–23928, Aug. 2023, doi: 10.1007/s11042-023-16194-z.

[11] N. E. Khalifa, W. Wang, A. A. Mawgoud, and Y.-D. Zhang, "COECG-resnet-GWO-SVM: an optimized COVID-19 electrocardiography classification model based on resnet50, grey wolf optimization and support vector machine," *Multimed. Tools Appl.*, vol. 84, no. 17, pp. 18305–18325, Jul. 2024, doi: 10.1007/s11042-024-19733-4.

[12] P. Agrawal, H. F. Abutarboush, T. Ganesh, and A. W. Mohamed, "Metaheuristic Algorithms on Feature Selection: A Survey of One Decade of Research (2009-2019)," *IEEE Access*, vol. 9, pp. 26766–26791, 2021, doi: 10.1109/ACCESS.2021.3056407.

[13] H. Wu, S. Du, Y. Zhang, Q. Zhang, K. Duan, and Y. Lin, "Threshold Binary Grey Wolf Optimizer Based on Multi-Elite Interaction for Feature Selection," *IEEE Access*, vol. 11, pp. 34332–34348, 2023, doi: 10.1109/ACCESS.2023.3263584.

[14] D. Zhang, Y. Zhang, S. Li, and S. Li, "A novel min–max robust model for post-disaster relief kit assembly and distribution," *Expert Syst. Appl.*, vol. 214, p. 119198, Mar. 2023, doi: 10.1016/j.eswa.2022.119198.

[15] Y.-Q. Cai *et al.*, "Pitfalls in Developing Machine Learning Models for Predicting Cardiovascular Diseases: Challenge and Solutions," *J. Med. Internet Res.*, vol. 26, p. e47645, Jul. 2024, doi: 10.2196/47645.

[16] A. Althnian *et al.*, "Impact of Dataset Size on Classification Performance: An Empirical Evaluation in the Medical Domain," *Appl. Sci.*, vol. 11, no. 2, p. 796, Jan. 2021, doi: 10.3390/app11020796.

[17] R. Bharti, A. Khamparia, M. Shabaz, G. Dhiman, S. Pande, and P. Singh, "Prediction of Heart Disease Using a Combination of Machine Learning and Deep Learning," *Comput. Intell. Neurosci.*, vol. 2021, no. 1, p. 8387680, Jan. 2021, doi: 10.1155/2021/8387680.

[18] T. Liu, A. Krentz, L. Lu, and V. Curcin, "Machine learning based prediction models for cardiovascular disease risk using electronic health records data: systematic review and meta-analysis," *Eur. Heart J. - Digit. Health*, vol. 6, no. 1, pp. 7–22, Jan. 2025, doi: 10.1093/ehjdh/ztae080.

[19] W. A. Rayadhani and M. Rahardi, "Comparative Analysis of Random Forest, SVM, and Naive Bayes for Cardiovascular Disease Prediction," *J. Appl. Inform. Comput.*, vol. 9, no. 6, pp. 3234–3243, Dec. 2025, doi: 10.30871/jaic.v9i6.11451.

[20] H. Zhang and R. Mu, "Refining heart disease prediction accuracy using hybrid machine

VOL. 11. NO. 4 MAY 2026
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v11i4.7826

learning techniques with novel metaheuristic algorithms," *Int. J. Cardiol.*, vol. 416, p. 132506, Dec. 2024, doi: 10.1016/j.ijcard.2024.132506.

[21] D. J. Kalita, V. P. Singh, and V. Kumar, "A novel adaptive optimization framework for SVM hyper-parameters tuning in non-stationary environment: A case study on intrusion detection system," *Expert Syst. Appl.*, vol. 213, p. 119189, Mar. 2023, doi: 10.1016/j.eswa.2022.119189.

[22] S. Malik *et al.*, "Hybrid metaheuristic optimization for detecting and diagnosing noncommunicable diseases," *Sci. Rep.*, vol. 15, no. 1, p. 7816, Mar. 2025, doi: 10.1038/s41598-025-91136-3.

[23] W. Pannakkong, K. Thiwa-Anont, K. Singthong, P. Parthanadee, and J. Buddhakulsomsiri, "Hyperparameter Tuning of Machine Learning Algorithms Using Response Surface Methodology: A Case Study of ANN, SVM, and DBN," *Math. Probl. Eng.*, vol. 2022, pp. 1–17, Jan. 2022, doi: 10.1155/2022/8513719.

[24] S. Srinivasan, S. Gunasekaran, S. K. Mathivanan, B. A. M. M. B, P. Jayagopal, and G. T. Dalu, "An active learning machine technique based prediction of cardiovascular heart disease from UCI-repository database," *Sci. Rep.*, vol. 13, no. 1, p. 13588, Aug. 2023, doi: 10.1038/s41598-023-40717-1.

[25] G. Saranya and A. Pravin, "A novel feature selection approach with integrated feature sensitivity and feature correlation for improved prediction of heart disease," *J. Ambient Intell. Humaniz. Comput.*, vol. 14, no. 9, pp. 12005–12019, Sep. 2023, doi: 10.1007/s12652-022-03750-y.

[26] S. Srinivasan, S. Gunasekaran, S. K. Mathivanan, B. A. M. M. B, P. Jayagopal, and G. T. Dalu, "An active learning machine technique based prediction of cardiovascular heart disease from UCI-repository database," *Sci. Rep.*, vol. 13, no. 1, p. 13588, Aug. 2023, doi: 10.1038/s41598-023-40717-1.

[27] L. O. Joel, W. Doorsamy, and B. S. Paul, "A comparative study of imputation techniques for missing values in healthcare diagnostic datasets," *Int. J. Data Sci. Anal.*, vol. 20, no. 7, pp. 6357–6373, Nov. 2025, doi: 10.1007/s41060-025-00825-9.

[28] J. Li *et al.*, "Comparison of the effects of imputation methods for missing data in predictive modelling of cohort study datasets," *BMC Med. Res. Methodol.*, vol. 24, no. 1, p. 41, Feb. 2024, doi: 10.1186/s12874-024-02173-x.

[29] A. Rácz, D. Bajusz, and K. Héberger, "Effect of Dataset Size and Train/Test Split Ratios in QSAR/QSPR Multiclass Classification," *Molecules*, vol. 26, no. 4, p. 1111, Feb. 2021, doi: 10.3390/molecules26041111.

[30] J. Sadaiyandi, P. Arumugam, A. K. Sangaiah, and C. Zhang, "Stratified Sampling-Based Deep Learning Approach to Increase Prediction Accuracy of Unbalanced Dataset," *Electronics*, vol. 12, no. 21, p. 4423, Oct. 2023, doi: 10.3390/electronics12214423.

[31] A. Roy and S. Chakraborty, "Support vector machine in structural reliability analysis: A review," *Reliab. Eng. Syst. Saf.*, vol. 233, p. 109126, May 2023, doi: 10.1016/j.ress.2023.109126.

[32] D. Mustafa Abdullah and A. Mohsin Abdulazeez, "Machine Learning Applications based on SVM Classification A Review," *Qubahan Acad. J.*, vol. 1, no. 2, pp. 81–90, Apr. 2021, doi: 10.48161/qaj.v1n2a50.

[33] R. Guido, S. Ferrisi, D. Lofaro, and D. Conforti, "An Overview on the Advancements of Support Vector Machine Models in Healthcare Applications: A Review," *Information*, vol. 15, no. 4, p. 235, Apr. 2024, doi: 10.3390/info15040235.

[34] M. Abdel-Basset, R. Mohamed, I. M. Hezam, K. M. Sallam, and I. A. Hameed, "An improved nutcracker optimization algorithm for discrete and continuous optimization problems: Design, comprehensive analysis, and engineering applications," *Heliyon*, vol. 10, no. 17, p. e36678, Sep. 2024, doi: 10.1016/j.heliyon.2024.e36678.

[35] P. Rink and W. Brannath, "Post-selection confidence bounds for prediction performance," *Mach. Learn.*, vol. 114, no. 3, p. 82, Mar. 2025, doi: 10.1007/s10994-024-06632-w.

[36] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Inf. Process. Manag.*, vol. 45, no. 4, pp. 427–437, Jul. 2009, doi: 10.1016/j.ipm.2009.03.002.

[37] J. Terven, D.-M. Cordova-Esparza, J.-A. Romero-González, A. Ramírez-Pedraza, and E. A. Chávez-Urbiola, "A comprehensive survey of loss functions and metrics in deep learning," *Artif. Intell. Rev.*, vol. 58, no. 7, p. 195, Apr. 2025, doi: 10.1007/s10462-025-11198-7.