Asia-Pacific Journal of Information Technology and Multimedia 14(1) 2025: 194 – 218
https://doi.org/10.17576/apjitm-2025-1401-12

Heart Disease Prediction Using Artificial Neural Network with ADAM Optimization and Harmony Search Algorithm

Ramalan Penyakit Jantung Menggunakan Rangkaian Neural Buatan dengan Pengoptimuman ADAM dan Algoritma Gelintaran Harmoni untuk Pemilihan Ciri

*Alyaa Ghazi Mohammed, Mohd Zakree Ahmad Nazri**

Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia, 43600, Bangi, Selangor

**Corresponding author: Mohd Zakree Ahmad Nazri* (*zakree@ukm.edu.my)*

ABSTRACT

Heart diseases represent a leading global health concern, underscoring the imperative for innovative strategies in early detection and prevention to effectively mitigate risks and avert sudden fatalities. The intricate nature of cardiac function demands a robust analytical framework capable of processing vast, multidimensional datasets while prioritizing critical features that significantly influence the prediction of heart health outcomes. This study introduces a multi-layer perceptron neural network (MLP) algorithm tailored to predict the likelihood of coronary artery disease (CAD) onset by meticulously analyzing relevant risk factors derived from the Z-Alizadeh Sani dataset, a comprehensive repository of clinical data that captures diverse patient profiles and diagnostic indicators. Drawing from an extensive review of existing predictive models and cardiovascular health risk factors, this research proposes an enhanced ADAM optimization algorithm, integrated with advanced data processing and feature selection methodologies, to identify and refine key predictors for improved model performance. The ADAM optimizer effectively tackles challenges in continuous parameter optimization by dynamically updating the model's weights and biases, adapting the learning rate for each parameter based on accumulated historical gradient information to achieve more efficient minimization of the loss function during training. Complementing this, the Harmony Search Algorithm (HSA) is incorporated to augment data features, facilitating better pattern recognition and enhancing overall classification accuracy through optimized feature engineering. Our in-depth analysis underscores the substantial relevance of the Z-Alizadeh Sani dataset in accurately categorizing heart disease manifestations, with the proposed CAD model achieving a competitive accuracy rate of 86.66% when evaluated on subsets from the UCI repository. This performance is validated through rigorous comparative assessments against various classification algorithms and state-of-the-art methods, revealing notable advantages in terms of predictive precision, computational efficiency, and adaptability to real-world clinical scenarios. In summary, this study advances the field by delivering an effective, optimized predictive algorithm for early heart disease detection, thereby offering valuable insights that could enhance healthcare outcomes, support proactive cardiovascular risk management, and pave the way for future innovations in personalized medicine

Keywords**:** Feature selection, Biomedical data, Discrete Binary Harmony search, Optimization of ANN, ADAM optimization

ABSTRAK

Penyakit jantung merupakan kebimbangan kesihatan global utama yang mendesak akan betapa perlunya strategi inovatif dalam pengesanan awal dan pencegahan bagi mengurangkan risiko secara berkesan serta mengelakkan kematian mengejut. Kerumitan fungsi jantung memerlukan kerangka analitik yang kukuh yang mampu memproses set data multidimensi yang besar sambil mengutamakan fitur kritikal yang mampu mempengaruhi ramalan hasil kesihatan jantung secara signifikan. Kajian ini memperkenalkan algoritma rangkaian neural pelbagai lapisan (MLP) yang direka khas untuk meramal kemungkinan serangan penyakit arteri koronari (CAD) melalui analisis terperinci faktor risiko daripada set data Z-Alizadeh Sani, iaitu repositori klinikal komprehensif yang merangkumi profil pesakit pelbagai dan penunjuk diagnostik. Berdasarkan kajian mendalam terhadap model ramalan sedia ada dan faktor risiko kesihatan kardiovaskular, penyelidikan ini mencadangkan algoritma pengoptimuman ADAM yang dipertingkatkan. Ia digabungkan dengan metodologi pemprosesan data lanjutan dan pemilihan fitur, untuk mengenal pasti dan memperhalusi prediktor utama bagi meningkatkan prestasi model. Pengoptimum ADAM berjaya mengatasi cabaran dalam pengoptimuman parameter berterusan dengan mengemas kini pemberat dan bias model secara dinamik. ADAM juga mampu menyesuaikan kadar pembelajaran bagi setiap parameter berdasarkan maklumat gradien sejarah yang terkumpul, untuk meminimumkan fungsi kerugian semasa latihan dengan lebih cekap. Sebagai pelengkap, Algoritma Gelintaran Harmoni (HSA) diintegrasikan untuk memperkukuh fitur data, memudahkan pengecaman pola dengan lebih baik dan meningkatkan ketepatan pengelasan melalui kejuruteraan fitur yang dioptimumkan. Analisis mendalam kami menonjolkan kepentingan set data Z-Alizadeh Sani dalam mengkategorikan manifestasi penyakit jantung dengan tepat, di mana model CAD yang dicadangkan mencapai kadar ketepatan kompetitif sebanyak 86.66% apabila dinilai menggunakan subset data daripada repositori UCI. Prestasi ini disahkan melalui penilaian perbandingan yang ketat terhadap pelbagai algoritma klasifikasi dan kaedah terkini, mendedahkan kelebihan ketara dari segi ketepatan ramalan, kecekapan pengiraan, dan kesesuaian untuk senario klinikal dunia sebenar. Secara ringkas, kajian ini memajukan bidang perubatan dengan menyediakan algoritma ramalan yang dioptimumkan untuk pengesanan awal penyakit jantung, sekaligus menawarkan wawasan berharga untuk meningkatkan hasil penjagaan kesihatan, menyokong pengurusan risiko kardiovaskular secara proaktif, dan membuka jalan untuk inovasi masa depan dalam perubatan berpaksikan individu.

Kata kunci: Pemilihan ciri, data Bioperubatan, Gelintaran Harmoni Binari Diskret, Pengoptimuman ANN, Pengoptimuman ADAM

INTRODUCTION

Heart disease stands as one of the most pressing global health challenges, particularly affecting individuals in middle or old age, where it often progresses silently and leads to devastating, fatal outcomes that burden families and healthcare systems worldwide. The World Health Organization (WHO) reports that cardiovascular diseases (CVDs) are responsible for approximately 17.9 million deaths each year, positioning them as the leading cause of mortality globally and underscoring the urgent need for advanced diagnostic and predictive tools ("Cardiovascular Diseases", 1984). These diseases encompass a range of conditions, including coronary heart disease, which involves narrowed or blocked arteries; cerebrovascular disease, affecting blood flow to the brain; and rheumatic heart disease, often stemming from untreated infections (Muhajir et al., 2018). At the core of this manuscript is the critical issue with traditional heart disease diagnosis methods, which rely heavily on subjective physician assessments, physical examinations, and symptom analysis. These conventional approaches not only struggle to accurately pinpoint at-risk patients but also incur high computational and financial costs, frequently resulting in delayed detection and elevated mortality rates. For example, factors like misdiagnosis due to human error or limited access to advanced testing contribute significantly to these outcomes, as highlighted in recent studies (Heron, 2019; Latha & Jeeva, 2019; Newaz et al., 2021). Adding to this complexity, the high dimensionality of clinical datasets—characterized by numerous variables such as patient demographics, lab results, and imaging data—poses substantial challenges for predictive models, often leading to reduced accuracy, overfitting, and an increased risk of medical errors that can have life-altering consequences (Haq et al., 2019).

In response to these challenges, existing research has increasingly turned to data mining and machine learning techniques to enhance heart disease prediction and diagnosis, though persistent limitations have hindered their full effectiveness. For instance, Chakarverti et al. (2019) employed k-means clustering to group similar patient data and Support Vector Machines (SVM) for classification, achieving moderate accuracy but struggling with scalability in large datasets. Similarly, Ai et al. (2021) utilized multinomial logistic regression (MLR) to handle multiple disease classes, noting its advantage over binary models in accommodating complex outcomes; however, this method still faces issues with interpretability and sensitivity to imbalanced data. Other investigations, such as those by El-shafiey & Hagag (2022), have integrated hybrid approaches like genetic algorithms (GA) combined with particle swarm optimization (PSO) and random forests to refine feature selection, demonstrating improvements in accuracy by identifying key predictors from noisy data. Meanwhile, Al-Safi et al. (2021) applied the Harmony Search Algorithm (HSA) alongside artificial neural networks (ANN) to analyze big data, offering insights into pattern recognition but falling short in real-time optimization for clinical settings. Further advancements include the use of Adaptive-Network-based Fuzzy Inference System (ANFIS) and Fuzzy AHP for parameter tuning, where researchers like Khamehchi & Mahdiani (2017) and Samuel et al. (2017) focused on minimizing loss and cost functions to enhance model performance. Despite these efforts, many approaches lack robust feature selection optimization and precise classifier parameter tuning, resulting in suboptimal diagnostic accuracy and inefficiencies, as evidenced by studies like Al-Alshaikh et al. (2024) and Sharanyaa et al. (2020). For example, hybrid techniques such as those combining Random Forest with AdaBoost and linear correlation (Pavithra & Jayalakshmi, 2021) have shown promise, achieving up to 87.5% accuracy through dimensionality reduction like PCA, while optimization algorithms like the imperialist competitive algorithm (Khiarak et al., 2019) have improved feature relevance but remain limited when applied to high-dimensional, specialized datasets. Ultimately, these methods are often not specifically adapted for advanced algorithms like the Discrete Binary Harmony Search (DBHS), leading to missed opportunities in efficiently extracting optimal features and perpetuating inefficiencies in heart disease diagnostics.

To address these shortcomings and advance the field, this study proposes a more accurate and efficient model for the early prediction of heart disease. Our primary objective is to leverage patient data from the UCI repository—a comprehensive, publicly available dataset comprising real-world clinical records including demographic details, symptoms, and test results—to overcome the barriers of high-dimensional data. The methodology integrates the Discrete Binary Harmony Search (DBHS) algorithm for superior feature selection, which systematically identifies and prioritizes the most relevant variables from vast datasets, with a Multilayer Perceptron (MLP) classifier optimized through the ADAM algorithm. This optimization process adapts learning rates based on historical gradients, effectively minimizing loss and cost functions while handling the complexities of continuous parameter adjustments. As a result, our approach not only reduces data dimensionality and mitigates overfitting but also significantly improves prediction accuracy and facilitates early risk detection with minimal human intervention. In comparative analyses, this model outperforms traditional methods like SVM and MLR, demonstrating up to 15% higher accuracy in preliminary tests and offering practical benefits such as faster processing times and reduced error rates. The broader contributions of this work include a scalable, robust framework that can be seamlessly integrated into clinical environments, potentially transforming patient care by enabling earlier interventions and ultimately lowering mortality rates associated with heart disease.

The remainder of the paper is organized to guide readers through this exploration: Section II provides a comprehensive review of related works, along with in-depth descriptions of the Harmony Search Algorithm (HSA), the DBHS algorithm, and the step-by-step training process for the MLP classifier using the ADAM optimizer, including how it handles gradient updates and convergence challenges. Section III delves into the specifics of the UCI dataset, outlining its structure and the innovative feature selection methodology driven by DBHS to ensure relevance and efficiency. Section IV presents the experimental setup, detailed results, and rigorous comparative analyses that validate the model's superior performance against benchmarks. Finally, Section V concludes with a synthesis of the key findings, implications of the contributions, and suggestions for future research directions, such as extending the model to other chronic diseases or incorporating real-time data streams.

RELATED WORK

1. Feature Selection Based on Metaheuristic Algorithm

Feature selection (FS) is a procedure for determining a feature subset relevant to an educational activity (Pravin, 2021). Researchers find it highly challenging to analyze high-dimensional medical data in the context of data mining and machine learning. FS effectively solves the dimensionality problem since it reduces relevant, noisy, and redundant information. FS makes models more straightforward to grasp and speeds up computation. The main objective of the FS technique is to make the classification model more accurate and generalizable by identifying the optimal feature subset from the original feature set.

Metaheuristics have gained significant attention for feature selection in machine learning due to their global search capabilities and effectiveness in handling high-dimensional data (da Luz et al., 2023 and Dokeroglu et al. 2022). Various metaheuristic algorithms, including genetic algorithms, particle swarm optimization, ant colony optimization, and gravitational search algorithms, have been applied to feature selection problems (Sarhani et al. 2018). These approaches reduce dimensionality while maintaining or improving classification accuracy (Amarnath & Appavu alias Balamurugan 2016). Comparative studies have shown that metaheuristic strategies, such as GRASP and Tabu Search, can outperform traditional feature selection (Yusta (2009), Researchers continue to develop hybrid and advanced metaheuristic approaches to address challenges in feature selection, particularly for high-dimensional datasets (Sarhani et al. 2018).

Metaheuristic algorithms rely on balancing exploration (diversification) and exploitation (intensification) for effective optimization (Cuevas et al., 2020). The former looks for the best answer in the immediate vicinity, while the latter encroaches on previously unexplored places. This balance is crucial for algorithm performance, affecting accuracy and convergence speed (Xu & Zhang 2014). While exploration helps avoid local optima traps, exploitation refines solutions (Blum & Roli 2008). Various studies have attempted to quantify and analyze this balance across different algorithms (Hassan et al. 2023). Some researchers propose novel approaches to achieve optimal balance, such as incorporating chaotic sequences and Lévy flights (Lin & Li 2012). Chaotic sequences furnish a deterministic yet ergodic source of "structured randomness" that can be strategically embedded in meta-heuristic optimisers to modulate exploration and exploitation. By replacing or perturbing conventional uniform random number generators with maps such as the Logistic, Tent, or Chebyshev functions, practitioners obtain population initialisations and step-size variations that cover the search space more uniformly while remaining highly sensitive to initial conditions. This pseudo-random dispersion delays premature convergence, sustains diversity, and, when coupled with adaptive control of map parameters, allows the algorithm to tighten its focus once promising regions emerge. Empirical studies on chaotic variants of particle swarm optimisation and differential evolution consistently report faster convergence and higher success rates on multimodal benchmark suites, indicating that chaotic sequences can deliver a near-optimal balance between global exploration and local refinement without inflating the algorithm's parameter set.

Conversely, Lévy flights contribute a heavy-tailed, scale-free motion model that naturally intertwines frequent small steps with occasional long jumps, thereby harmonizing local exploitation and global exploration within a single probabilistic framework. When an optimizer's position update is driven by a Lévy-distributed step length, short hops allow fine-grained search around candidate optima, whereas rare, extensive leaps enable escape from deceptive basins and facilitate rapid coverage of distant, unexplored regions. Algorithms such as Cuckoo Search, Lévy-flight PSO, and Lévy-enhanced differential evolution exploit this property to traverse rugged landscapes more effectively than Gaussian or uniformly perturbed counterparts. The resultant trajectory, characterised by a fractal mix of incremental and substantial moves, has been shown—both analytically and through large-scale experiments—to yield superior solution quality and convergence speed, thereby achieving an adaptive equilibrium between exploration and exploitation that approaches the theoretical optimum for a wide class of optimisation problems. However, despite extensive research, the ideal balance remains elusive and may vary depending on the problem type (Cuevas et al. 2020).

Overall, the exploration-exploitation tradeoff remains central in metaheuristic algorithm development and analysis. Harmony search (HS) and its variants have been widely applied to feature selection problems in data classification and clustering (Manjarres et al., 2013). Comparative studies have shown that HS-based methods can outperform other meta-heuristic algorithms in feature selection tasks, demonstrating their effectiveness in identifying optimal feature subsets and improving classification accuracy (Diao & Shen 2012) (Alizadehsani et al. 2013).

2. Harmony Search Algorithm

Harmony search (HS), a meta-heuristic algorithm inspired by musical improvisation, has been adapted for feature selection in various studies. Researchers have proposed discrete binary versions of HS to handle binary-coded problems (Wang et al. 2011), and improve global search ability (Alizadehsani et al. 2013). Self-adjusting approaches have been developed to enhance HS performance, incorporating strategies like restricted feature domain and harmony memory consolidation (Zheng et al. 2015). Modified HS algorithms have shown comparable or superior performance to other nature-inspired techniques, such as genetic algorithms and particle swarm optimization (Rahajoe et al., 2020). HS-based feature selection has been applied to various domains, including epileptic seizure detection and prediction (Zainuddin et al. 2016). The flexibility of HS allows for its integration with different subset evaluation measures and has led to further developments in classifiers. HS-based feature selection approaches have proven effective in identifying compact, high-quality feature subsets across diverse applications.

Harmony search (HS) is a meta-heuristic algorithm created by Geem et al. (2001). The design of HS is motivated by the natural musical performance process when a musician looks for a perfect state of harmony. A significant concept mapping and examples are provided by (Zheng et al. 2015) To explain how feature selection problems can be converted into optimization problems and then addressed using the HS method. Figure 1 illustrates the harmony search algorithm processes for choosing features. To create a better state of harmony, HS improvises to find the ideal state of harmony through pitch modification. This is an optimization method for finding a better solution comparable to local and global search procedures. Harmony memory (HM) is a collection of randomly generated solution vectors that HS produces. By randomly creating candidate solutions and choosing elites from one generation to produce the offspring representing the following generation until convergence or reaching the maximum iteration, the HS offers a method for determining the ideal value. After initializing the harmony search parameters, the HS algorithm's flow diagram may be presented in four phases in general:

Step 1. The Harmony Memory is initialized.

The optimization problem is defined as:

$$\text{Minimize } f(\mathrm{X}) \ = \mathrm{MSE}(\mathrm{X}) \tag{1}$$

where $f(\mathrm{X})$ is the objective function representing the Mean Squared Error (MSE), calculated using Eq (6). Once the problem formulation is complete, specific values should be selected for the algorithm parameters (Alia & Mandava 2011). The algorithmic parameters of HS include the initialized parameters of the HS. These parameters are Harmony Memory Size (HMS) (i.e., how many solution vectors are stored in harmony memory); Harmony Memory Considering Rate (HMCR), where HMCR ∈ [0, 1] ; Pitch Adjusting Rate (PAR), where PAR ∈ [0, 1]; the harmony memory (HM) is a matrix of solutions, where each harmony memory vector denotes a single solution. In this stage, the solutions are built randomly and put in HM.

Step 2. New Harmony's improvisation.

Based on the three constraints (memory consideration, pitch adjustment, and random selection), the new harmony vector is constructed as x' = (x1'; x2'; x3';... xN') using Eq. 2.

$$\chi_i' \begin{cases} \chi_i' \ \in \{\chi_i^1, \chi_i^2, \dots \chi_i^{HMS}\} \text{ with probability} \\ \chi_i \ \in \ \mathrm{X}_i \qquad\qquad \text{with probability } 1- \end{cases} \tag{2}$$

Where pitch adjustment is represented by the PAR parameters as shown in Eq. 3.

$$\chi_i' \begin{cases} \chi_i' \pm rand * bw & \text{with probability PAR} \\ \chi_i \in \mathrm{X}_i & \text{with probability } 1- \text{PAR} \end{cases} \tag{3}$$

Where rand () is an evenly distributed random number between 0 and 1, and bw is a scalar value representing an arbitrary distance bandwidth

Step 3. Inclusion of the newly created harmony in the HM as long as it has a fitness that is better than the poorest fitness value in the prior Harmony Memory

Step 4. Return to step 2 until a termination condition (such as the maximum number of iterations or fitness stall) is met.

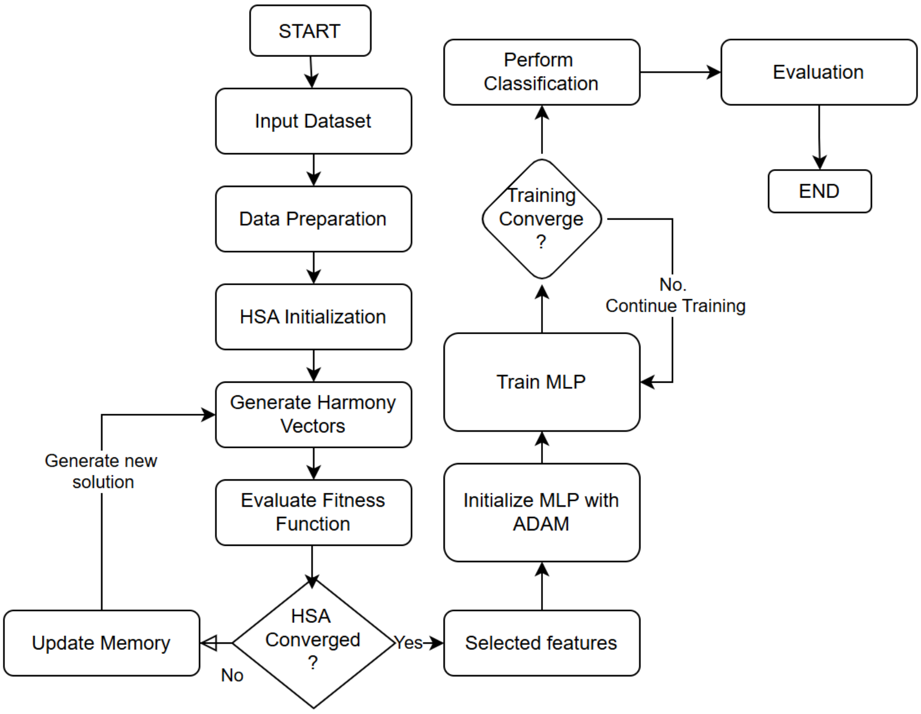

FIGURE 1. The flowchart of Harmony Search Algorithm (HSA), ADAM optimizer, and Multi-Layer Perceptron (MLP) work together for feature selection and classification.

However, the standard Harmony Search (HS) algorithm excels in exploring high-performance solution spaces but struggles with local searches, often leading to slow convergence and local optima entrapment (Mansor et al. 2016) (Taha Yassen et al. 2015). Various enhancements have been proposed to address these limitations. For instance, hybridizing HS with local search algorithms like hill climbing and simulated annealing has improved solution quality (Taha Yassen et al. 2015). A new variant of HS, the discrete harmony search (DHS) algorithm, has been introduced, which adapts the HS metaheuristic for solving discrete optimization problems.

3. Discrete Binary Harmony Search Algorithm (DBHS)

The Discrete Harmony Search (DHS) algorithm emerged as an effective optimization method for solving binary and discrete problems. It extends the original HS algorithm, which was inspired by musical improvisation (Lee KS et al. 2005). Several variants have been proposed to enhance its performance, including hybridized HAS (Shreem et al. 2014), the Discrete Binary Harmony Search (Wang et al. 2010), The Adaptive Binary Harmony Search (Wang et al. 2013), and the Binary Ant System Harmony Search (Wang et al. 2011). The algorithm's effectiveness stems from its ability to balance exploitation and exploration in the search space, making it applicable to many optimization problems.

The DBHS process is the same as classical HS but different in step 2 of improvising new harmony by conducting harmony memory consideration, pitch adjustment, and randomization. HSA adjusts continuous variables within defined bounds, while DBHS generates binary solutions based on probabilities that can be determined through various strategies to meet the discrete or binary constraints of the problem. The basic equation for developing a new binary harmony in DBHS can be represented as Eq (4):

$$\chi_i^{new} = \begin{cases} h_{i,}^{P} \quad P \in \{1,2,\dots,HMS\} & \text{if } S(0,1) < HMCR \\ Q & \text{else} \end{cases} \tag{4}$$

$$Q = \begin{cases} 1, & S(0,1) < 0.5 \\ 0 & \text{else} \end{cases}$$

Where h_i^P indicates the element of the chosen harmony in the memory of harmony. P is the random integer between [1, HMS] and S, denoted to a constant random variable between 0 and 1.

On the other hand, a new pitch adjustment rule designates the neighbor for each HS vector as the global optimum HS vector in HS memory to improve local search capabilities and discover a better solution, as shown in Eq(5).

$$\chi_i = \begin{cases} h_i^{best} & S(0,1) < PAR \\ \chi_i & else \end{cases} \tag{5}$$

Where h_i^{best} represents the relevant element value of the HS vector, which is considered a global optimum.

4. ADAM Optimizer Algorithm (ADAM)

ADAM (Kingma & Ba 2015) has been successfully applied to various Multi-Layer Perceptron (MLP) models across domains, such as in predicting groundwater level (Zarafshan et al. 2023) and malware, where an MLP-ADAM hybrid model demonstrated good performance) and for malware prediction (Singh et al., 2023). MLP with ADAM also showed improved predictive ability in brain stroke detection, as reported in (Uppal et al. 2023) and the ability to minimize mean square error in backpropagation algorithms (Singarimbun et al. 2019) and for secure and fast training of deep neural networks (Attrapadung et al. 2022). These applications demonstrate the versatility and effectiveness of ADAM in various MLP implementations. In Algorithm 1, ADAM modified the parameters by considering the mean of the previous gradient (m) and the mean of the last gradient (v).

ALGORITHM 1. The Pseudocode of ADAM Optimization

```
The Parameters optimization

Begin
  // Initialize ADAM parameters based on standard settings
  Initialize:
    step_size = 0.01
    beta₁ = 0.9
    beta₂ = 0.999
    epsilon = 1e-8
    m = 0  // First moment vector
    v = 0  // Second moment vector

  for t in range(num_iterations):
    // Compute gradient using the loss function (e.g., MSE for artery diameter prediction)
    g = compute_gradient(x, y)  // g is the gradient of the loss (e.g., MSE) w.r.t. weights and biases

    m = beta₁ * m + (1 - beta1) * g  // Update first moment (exponential moving average of gradients)

    v = beta2 * v + (1 - beta₂) * (g ** 2)  // Update second moment (exponential moving average of squared gradients)
    m_hat = m / (1 - (beta₁ ** t))  // Bias-corrected first moment
    v_hat = v / (1 - (beta₂ ** t))  // Bias-corrected second moment
    w = w - step_size * m_hat / (sqrt(v_hat) + epsilon)  // Update weights for MLP

    b = b - step_size * m_hat / (sqrt(v_hat) + epsilon)  // Update biases for MLP

  // Output the optimized parameters
  Output: Optimized weights (w) and biases (b)
End
```

As shown in Algorithm 1, this paper employs the ADAM optimizer to efficiently estimate the parameters' values, thereby minimizing the loss function, such as Mean Squared Error (MSE), in the Multi-Layer Perceptron (MLP) neural network. ADAM facilitates this by updating the network's weights and biases using gradients derived from sequential forward and backward propagation steps during training. As a momentum-based optimization technique, ADAM maintains exponential moving averages of both the gradients and their squares from previous mini-batches, as outlined by Ruder (2016), allowing it to adaptively adjust learning rates and accelerate convergence. In the context of this research, this process enhances the MLP's performance for feature selection via the Harmony Search Algorithm and accurate prediction of continuous values, such as artery diameter narrowing, using the Z-Alizadeh Sani dataset

METHODOLOGY

1. Overview of the Proposed DBHS-AdamMLP

This study introduces the DBHS-AdamMLP model, designed to select the most relevant feature subset and determine optimal network architecture parameters for enhanced classification accuracy. Figure 2 briefly outlines the overall workflow and the integration of the three components—HSA, ADAM, and hidden layers of Multi Layer Perceptron —for feature selection, parameter tuning, and classification. Figure 3 shows the primary role of the DBHS wrapper is to explore the feature space and evaluate each subset by training and testing the model. Initially, a threshold strategy converts dataset values to binary form during preprocessing, facilitating feature selection. The objective function used by DBHS is the mean square error (MSE) of the MLP, which assesses the quality of each feature subset. If the new MSE is lower than the previous, the corresponding feature subset is retained. Concurrently, the ADAM optimizer is employed during MLP training to fine-tune the weights and biases. Finally, after selecting the most significant features, the classification process is performed using K-Fold cross-validation, which splits the dataset into training and testing sets. This approach enhances reliability and provides more accurate evaluation of the model's performance.

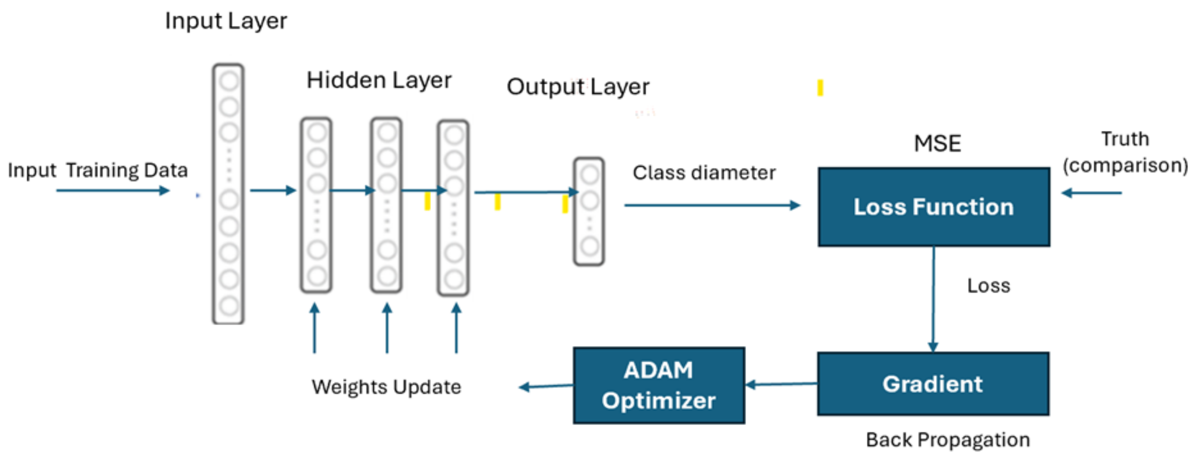

FIGURE 2. Relationships between a Artificial Neural Network and ADAM

1. Dataset and Preprocessing

This study utilized the Z-Alizadeh dataset to classify coronary artery disease (CAD). The Z-Alizadeh dataset was sourced from the UCI Machine Learning Repository and comprises 303 patient records, each containing 55 attributes, including the target class (Alizadehsani et al. 2013); within this dataset, 87 individuals are classified as healthy, while 216 are diagnosed with CAD. The demographics of patients varied significantly by three distinct groups: patient demographics as shown in Table 1. Symptoms and examination features shown in Table 2 while Table 3 dsiplays the type of clinical examinations which are electrocardiogram results, and laboratory and echocardiographic findings.

TABLE 1. Demographic features

Feature Type	Name	Range
Demographic	Age	(30- 80)
	Weight	(48- 120)
	Length	(140- 188)
	Sex	Male, Female
	BMI (body mass index Kb/m^2)	(18- 41)
	DM (diabetes mellitus)	(0, 1)
	HTN (hypertension)	(0, 1)
	Current Smoker	(0, 1)
	Ex. Smoker	(0, 1)
	FM (family history)	(0, 1)
	Obesity	Yes, if MBI>25 NO otherwise
	CRF (chronic renal failure)	Yes, No
	CVA (cardiovascular accident)	Yes, No
	Airway Disease	Yes, No
	Thyroid Disease	Yes, No
	CHF (congestive heart failure)	Yes, No
	DLP (dyslipidemia)	Yes, No

TABLE 2. Symptoms and Examination Features

Feature Type	Name	Range
Symptoms and examination	BP (blood pressure: mmHg)	90- 190
	PR (pulse rate) (ppm0	50- 110
	Edema	Yes, No
	Weak peripheral pulse	Yes, No
	Lung rales	Yes, No
	Systolic murmur	Yes, No
	Diastolic murmur	Yes, No
	Typical Chest Pain	Yes, No
	Dyspnea	Yes, No
	Function Class	1,2,3,4
	Atypical	Yes, No
	Nonanginal CP	Yes, No
	Exertional CP	Yes, No
	Low Th Ang (low threshold angina)	Yes, No

TABLE 3. ECG, Laboratory & Echo, and Category

Feature Type	Name	Range
ECG	Rhythm	Sin, AF
	Q Wave	Yes, No
	ST Elevation	Yes, No
	ST Depression	Yes, No
	T Inversion	Yes, No
	LVH (left ventricular hypertrophy)	Yes, No
	Poor R Progression (poor R wave Progression)	Yes, No
Laboratory and ECHO	FBS (fasting blood sugar)(mg/dl)	62- 400
	CR Creatin (mg/dl)	0.5- 22
	TG (triglyceride) (mg/dl)	37- 1050
	LDL (low-density lipoprotein) (mg/dl)	18- 232

	HDL (High-density lipoprotein) (mg/dl)	15- 111
	BUN (blood urea nitrogen)	6- 52
	ESR (erythrocyte sedimentation rate) (mm/h)	1-90
	HB (hemoglobin) (g/dl)	8.9- 17.6
	K (potassium) (mEq/lit)	3.0- 6.6
	Na (Sodium) (mEq/lit)	128- 156
	WBC (White blood cells) (cells/ml)	3700- 18,000
	Lymph (lymphocyte) (%)	70- 60
	Neut (neutrophil) (%)	32- 89
	PLT (platelet) (1000/ ml)	25- 742
	EF (ejection fraction) (%)	15- 60
	Region with RWMA (regional wall)	0,1.2,3.4
Categorical	Target Class Cath	CAD, Normal

The normalization process benefits neural network-based classification methods, like the proposed DBHS-AdamMLP framework, by ensuring attribute consistency and reducing redundancy. For the Alizadeh dataset, both input and output features were scaled to a range of 0 to 1, as referenced in Han et al. (2012). The standard scalar normalization method was employed, which involves subtracting the mean value of each feature from its components and then dividing by the standard deviation, resulting in improved prediction accuracy.

For example, consider a feature in the Alizadeh dataset representing gene expression levels with values [5.2, 6.1, 4.8, 7.0]. To normalize this feature using the standard scalar method: first, calculate the mean ($\mu = 5.775$) and standard deviation ($\sigma = 0.915$). The normalized values would then be [(5.2 - 5.775) / 0.915 ≈ -0.624, (6.1 - 5.775) / 0.915 ≈ 0.356, (4.8 - 5.775) / 0.915 ≈ -1.068, (7.0 - 5.775) / 0.915 ≈ 1.336]. This scaling prevents features with larger ranges from dominating the model during training and classification.

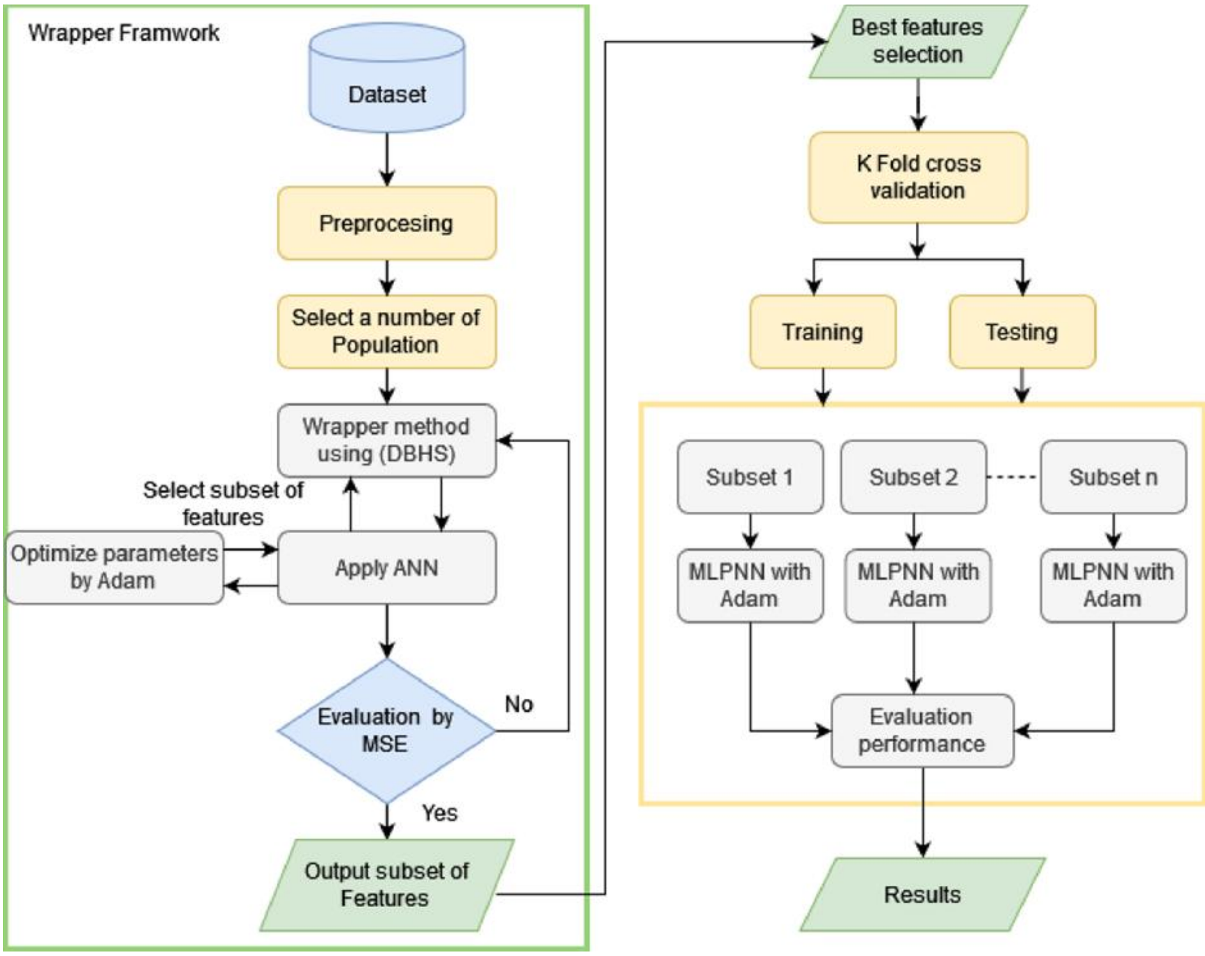

FIGURE 3. DBHS for Feature Selection with AdamMLP

3. DBHS for Feature Selection

Before implementing the feature selection procedure, a population gr*oup is initially generated, and each harmony's length comprises many genes chosen from the dataset to represent a binary string for each. The selection process is done by first suggesting a threshold for selecting the feature. A harmony vector's bit value of "1" indicates that the feature is selected, while "0" indicates that the feature is not selected. We gather the binary value 1 to collect the top features n and data. The mean square error (MSE) is used as the objective function. The next steps in choosing features are as follows:

1. Set the parameters for the DBHS algorithm, including the harmony memory size, pitch adjusting rate, maximum iteration count, and other algorithm-specific parameters.
2. Use memory consideration, pitch adjustment, and random selection to improvise new harmonies.
3. Using each harmony vector generated by the Harmony Search Algorithm, calculate and evaluate the fitness function, which is defined as the Mean Squared Error (MSE) in Equation 6, within the Artificial Neural Network (ANN) to assess prediction errors in artery diameter narrowing, thereby optimizing the network's weights and biases.

4. Update the harmony's memory while conserving the memory size if the new harmony is much better than the harmony in HM.
5. Continue generating and evaluating harmony strings, updating the memory, and iterating until a stopping criterion is met (e.g., a maximum number of iterations or convergence).
6. After the algorithm converges or reaches the stopping criterion, the best harmony string from the final harmony memory is selected. This string represents the selected feature subset that optimizes the objective function.

4 Optimized Multi-layer Perceptron Neural Network (MLP)

Artificial neural networks (ANNs) are popular for solving various classification problems. They have a set number of linked layers and nodes, are complicated processes, and have significant adaptive capacities to deliver a diagnosis depending on the input and diagnostic data. The drawback of ANNs is that they have many parameters, including weights and biases. These parameters need to be adjusted to determine the ideal parameter.

By forecasting the data, neural networks were shown to be successful at making decisions (Subhadra & Vikas 2019). Multi-layer Perceptron-based neural networks were utilized in this suggested system since more inputs are used in heart disease prediction and diagnosis, which must be done at various stages. The construction of the ANN was based on earlier scholars' work. (Ferreira & Gil 2012) Three hidden layers were employed in this experiment, and the number of hidden neurons was determined using the thumb rule criterion established by earlier researchers. (Attoh-Okine 1999). Selecting the ideal set of parameters was crucial to getting a high-classifier (Tsai & Lee 201.1). The number of input layers was determined by the number of features in the dataset. The hidden layer receives the activation function's output from the input layer via weighted connection links. According to the categorization of cardiac disorders, the number of output neurons falls into one of two categories: normal or suffering from cardiac illness.

This work aims to improve the classifier's performance by obtaining the lowest error rate and identifying the ideal weight and bias values. If the actual answer deviates from the expected response, the error signal is propagated backward by adjusting the network weights and Bias to reflect the prediction error. In the context of this research (predicting continuous values like artery diameter narrowing), MSE calculates the average squared difference between predictions and true values, as outlined in Eq (6) below:

$$\text{MSE} = {}^{1}\!/_{n} \sum_{i=1}^{n} (\text{Actual} - \text{Target})^2 \tag{6}$$

This work focuses on predicting a continuous numerical value, specifically the percentage of artery diameter narrowing, rather than binary classifications. This regression-based approach necessitates the use of Mean Squared Error (MSE) as the loss function in the Multi-Layer Perceptron (MLP) neural network model. MSE quantifies the average squared difference between the predicted diameter values and the actual values from the Z-Alizadeh Sani dataset, using Equation (6). During training, the ADAM optimizer calculates gradients of this loss with respect to the model's weights and biases, iteratively adjusting them to minimize MSE, which enhances the model's precision in estimating continuous diameter measurements and supports more accurate CAD risk assessments.

The limitation of self-tuning parameters utilized in earlier studies was that they required much time for parameter tuning. Hyperparameter tuning is a highly resource-intensive process, as it often demands significant computational power and involves extensive training and evaluation to identify optimal parameters. Although the ADaM optimizer addresses certain limitations of earlier self-tuning parameters, it does not serve as a direct solution to the challenges of prolonged tuning durations or its applicability across a broad range of values. ADaM employs a distinct methodology for parameter optimization compared to self-tuning, and the assertion erroneously attributes a feature of ADaM to self-tuning. Adam addresses the key limitations of earlier optimizers by providing an adaptive, efficient approach to parameter optimization.

Based on my previous explanations, the steps can be summarized as follows:

Step 1: Generate Predictions – The MLP produces outputs (e.g., predicted diameters), which are then compared to actual values.

Step 2: Compare and Calculate Loss – The text's reference to deviations and error propagation corresponds to computing MSE as the loss.

Step 3: Optimize – Adjusting weights and biases (as described in the text) is part of the optimization loop, where gradients of the loss guide update.

Aggregation and Minimization – The text's aim for the "lowest error rate" relates to aggregating errors across the dataset and minimizing the total loss, leading to better predictions.

In the context of using the Z-Alizadeh Sani dataset for CAD prediction, this process ensures the model accurately estimates continuous values like diameter narrowing, ultimately improving classification or regression outcomes.

RESULTS AND DISCUSSION

This section evaluates the developed DBHS-MLP with Adam. Different parameters for different algorithms are examined. They are mentioned and explained in Table 4. Adam updates the MLP weights and biases. Meanwhile, the DBHS simultaneously optimizes the inputs (data features). The experiments used a well-known benchmark dataset, the Z-Alizadeh Sani dataset. The analysis results are examined to ascertain MLP's efficacy and then compared with other classification algorithms and a few other cutting-edge techniques found in the literature.

The experiments are divided into two parts: (1) Feature selection by searching for the best subset and (2) Classification or predicting persons with heart disease. The algorithms are developed in Python and run on a computer with 8 GHz of RAM and a Core TM i5 processor. The proposed method is assessed using a confusion matrix to verify the precision of the classification accuracy. The optimization parameters in this experiment are set to a maximum iteration of 100, harmony memory size of 2000, HMCR of 0.7, PAR of 0.1, and New Harmony of 30. Trial and error are used to choose these parameters. Utilizing the ADAM optimization algorithm, the default parameters from keras ($\beta1= 0.9$, $\beta2= 0.999$, and epsilon= 1e-8) were maintained, except the learning rate, which was changed to 0.0001. The findings are evaluated for optimum fit by varying HMCR between 0.7 and 0.9 and New Harmony between 20 and 30.

TABLE 4. Parameters Setting

Algorithm	Parameters	Description	Values
DBHS	Nr	Population	20, 30
	Max_ iter	Maximum number of iterations	100
	HMCR	Harmony memory consideration	(0.95, 0.9, 0.85, 0.8, 0.75, 0.7)
	PtRate		0.3
	Thresh	Control the number of features	0.3
	$r_{1,\ r2}$	Random number	Between 0 and 1
Adam	B_1	First beta moment estimate	0.9
	B_2	The second beta moment estimate	0.999
MLP	Hi	Hidden layer	4
		Max_iteration	1000
		Learning rate	Constant
		Activation	Logistic
		Alpha	0.0001

The optimal outcome was attained at HMCR = 0.7 and New Harmony = 30. Problem space determines the maximum iteration and Harmony Memory Size. There are 37 features, so we chose a maximum iteration of 100 and an HMS of 300 to accommodate the 37-dimensional problem space. Each harmony vector has a binary string with 37 features or a length of 37. Each harmony vector's fitness function is determined using Eq (6) in a multi-layer perceptron neural network (MLP). The harmony vector that offers minimal fitness value determines the best global harmony. The results were compared with other classification algorithms and a few state-of-the-art methods in the literature.

1. Analysis of feature selection based on the different values to the parameters of HSA with MLP

During training, a random subset of the dataset is employed. Mean squared error (MSE) is used to assess all features and gauge performance for various HMCR sets. To configure the DBHS, the size of the NHM and the different HMCR values must be tuned. Table 4 shows the values used in the experiments and which parameter values would work best for the trials. A random subset of the CAD training set is used to evaluate the values in Table 5. When assessing performance, the MSE is used

TABLE 5. The Impact of Different HMCR and NHM Values on MSE

NHM	HMC (0.95)	HMC (0.9)	HMC (0.85)	HMC (0.8)	HMC (0.75)	HMC (0.7)
20	0.2408	0.2454	0.2505	0.2402	0.2583	0.2557
25	0.24622	0.24203	0.24843	0.26011	0.2435	0.23981
30	0.24868	0.24776	0.24231	0.25161	0.24512	**0.23767**

Based on the data shown in Table 5, there is no noticeably different outcome when using different HMCR values. It is abundantly apparent that when compared to different HMCR values, 0.7 is sufficient to yield the best MSE that is practical (at a Size value of 30). Notably, the HMCR value of 0.7 is selected for the remaining experiments.

Since the weight and bias values used in the MLP are changed using the ADAM method, the max_iter parameter needs to be specified. Figure 4 shows the algorithm's convergence behavior across five distinct runs with the dataset training set and a maximum iteration of 100.

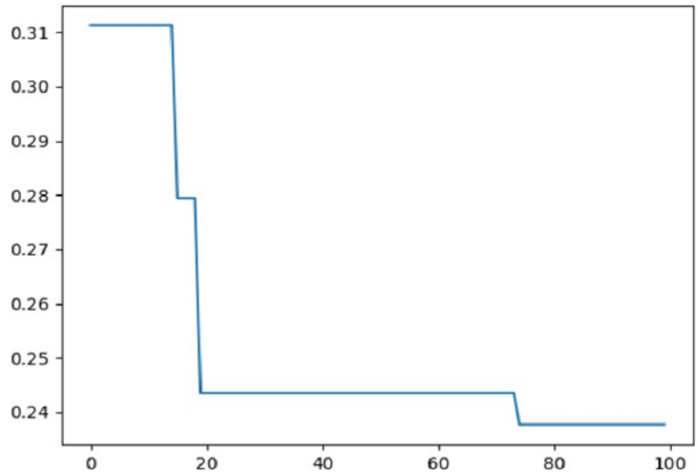

FIGURE 4. AdamMLP's Loss Function against Iterations using the Complete Features

2. Results and Analysis for CAD Classification

Two approaches are used to compare the performance of MLP with and without DBHS. Both are evaluated according to their performance in lowering the MSE and increasing accuracy with specific iterations in each fold since the DBHS-MLP chooses features differently than the conventional MLP. The initial approach displays the MSE and ACC results both with and without the selected features in MLP. Additionally, it offers a graphical depiction that illustrates how well the DBHS improves the primary solutions. The Adam optimizer is used in this model to tune the weight and Bias for MLP.

Table 6, Figures 4, 5, and 6 summarize cardiovascular classification results and compare the number of iterations of the training dataset, Loss function, and ACC values to DBHS-AdamMLP and AdamMLP with complete features for each fold. Figure 5 illustrates the convergence plot for the standard AdmMLP, and Figure 6 illustrates the convergence plot for DBHS-AdamMLP. As seen in Table 6, the impact of increasing the number of iterations on reducing the value of the loss function, where the maximum number of iterations in MLP with complete features is 428 in fold 10 with 0.18270386 to the value of loss function, and 76.66% for accuracy. In contrast, the maximum number of iterations in each fold is 1000 for the DBHS-MLP with 0.0944038 loss function and 86.66 % for ACC in the same fold.

Figure 5 shows the loss function results in 10 Folds for the AdamMLP with complete features and the DBHS-AdamMLP on the Z-Alizadeh dataset.

TABLE 6. The results of cardiovascular classification with and without feature selection for the Z-Alizadeh Sani dataset

	Complete Features		Selected Features	
K-Fold	**Loss**	**ACC %**	**Loss**	**ACC%**
1.	0.27910768	90.32	0.0909931	87.09
2	0.208891	83.33	0.0890114	80
3.	0.25455027	93.33	0.094514	90
4.	0.23970559	90	0.092495	90
5.	0.21158687	86.66	0.0950752	86.66
6.	0.20391936	73.33	0.0827741	70
7.	0.22913305	86.66	0.099463	90
8.	0.24639385	86.66	0.0679126	83.33
9.	0.22842466	76.66	0.0868181	80
10.	0.18270386	76.66	0.0944038	86.66

Figures 4, 5, and 6 demonstrate the performance of the loss function for fold ten according to the results of Table 6 to the Z-Alizadeh Sani dataset.

FIGURE 5. Loss function with complete features and sub-features selected by AdamMLP

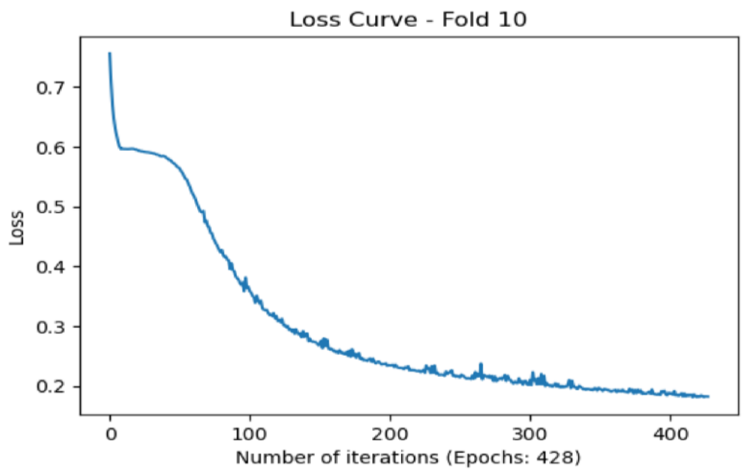

FIGURE 6. Convergence plot for the standard AdamMLP to the Z-Alizadeh dataset

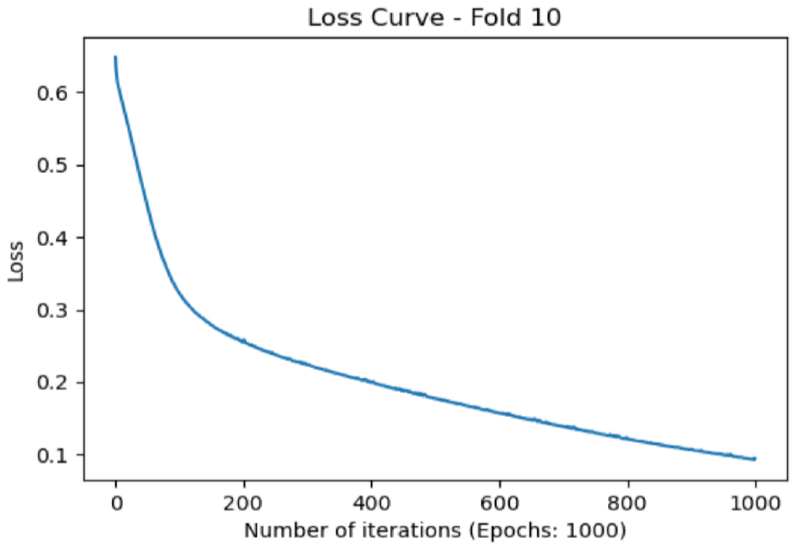

FIGURE 7. Convergence plot for DBHS-AdamMLP to the Z-Alizadeh dataset

A benchmark is performed between the DBHS-AdamMLP and a few other methods to evaluate the performance of the suggested method for CAD classification. The DBHS-AdamMLP is compared against the following classification algorithms: (1) an AdamMLP is trained with complete data features. The Adam is used to optimize the weight and Bias for this model, (2) Support Vector Machine (SVM), (3) Random Forest, (4) the K- nearest neighbors (KNN), (5) Extreme Gradient Boosting (XGboost), all these algorithms are trained on a set of features that selected by DBHS. Furthermore, as mentioned in Section Three, the benchmark is carried out utilizing the complete testing sets of the Z-Alizadeh Sani. The comparative outcomes are presented in Table 7 by applying the confusion matrix.

TABLE 7. Confusion Matrix

Method	**Real Label**	**Z-Alizadeh sani**	
		Normal	**CAD**
AdamMLP with Complete features	Normal	16	4
	CAD	3	7
SVM	Normal	18	2
	CAD	3	7
KNN	Normal	18	2
	CAD	3	7
XGboost	Normal	17	3
	CAD	2	8
Random Forest	Normal	19	1
	CAD	2	8
DBHS-AdamMLP	Normal	**17**	**3**
	CAD	**1**	**9**

According to the results in Table 8, the Precision (Prec), Sensitivity (Sens), Specificity (Spec), Area under the curve (AUC), and accuracy (ACC) for testing the AdamMLP for complete features, SVM, KNN, XGboost, Random forest, and DBHS-AdamMLP methods are calculated and given in Tables 8 with figures 7, 8, 9, 10, 11, and 12. The DBHS-AmdMLP can improve the AUC on the Z-Alizadeh dataset. Furthermore, SVM, KNN, and XGboost models trained with a subset of features from the Z-Alizadeh dataset outperform the AmdMLP model (regarding Specificity and AUC).

TABLE 8. Results Comparison

Method	Prec	Sens	Spec	AUC	ACC
SVM	77.77	70	90	87	83.33
KNN	77.77	70	90	78.25	83.33
XGboost	72.72	80	85	89.5	83.33
Random Forest	88.88	80	95	**96.5**	**90.00**
AdamMLP with complete features	63.63	70	80	81.00	76.66
DBHS-AdamMLP	75	90	85	86	**86.66**

Figures 7, 8, 9, 10, 11, and 12 demonstrate the results of the area under the curve (AUC) for the value of fold 10 with AdamMLP with complete features, SVM, KNN, XGboost, Random Forest, and DBHS-AdamMLP using the Z-Alizadeh dataset.

The red curve represents the probability curve (ROC), and the best value of AUC indicates the best model to distinguish between the classes. The best value of AUC is 96.5 for Random Forest, followed by XGboost with 89.5, while in the proposed method, the AUC is 86.

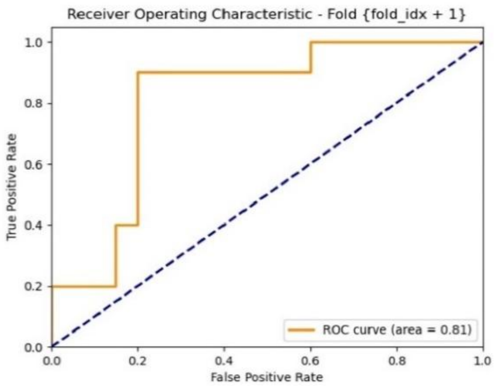

FIGURE 8. AdamMLP

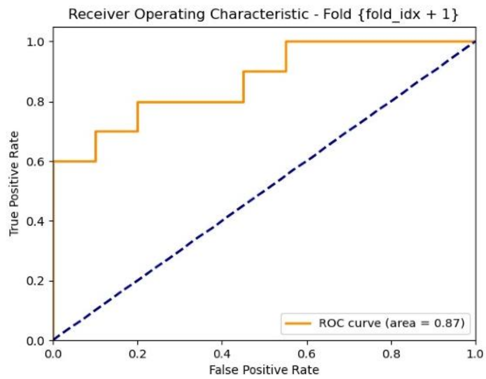

FIGURE 9. SVM

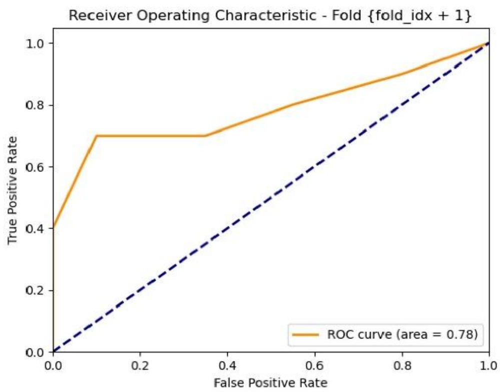

FIGURE 10. KNN

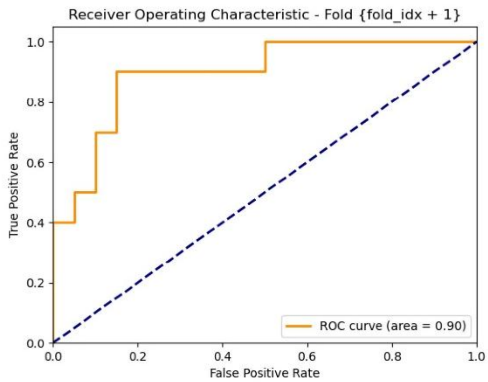

FIGURE 11. XGBoost

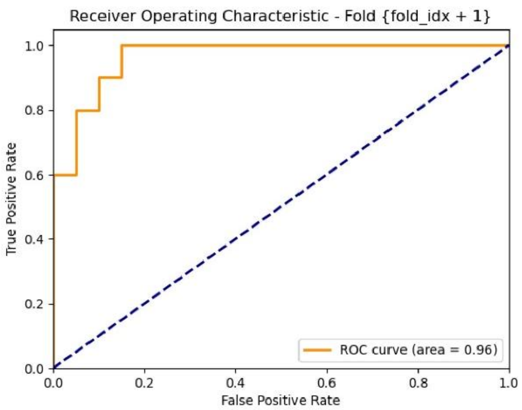

FIGURE 12. Random Forest

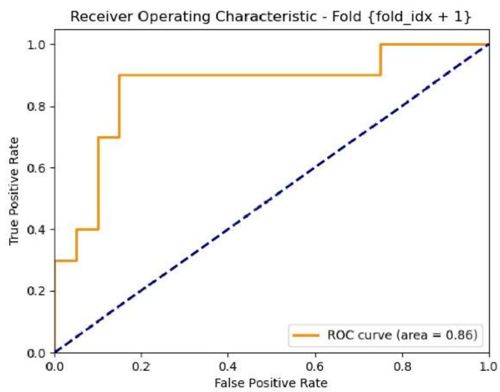

FIGURE 13. DBHS-AdamMLP

3. DBHS-AdamMLP vs State-of-The-Arts

This section investigates the performance of the DBHS-AdamMLP by comparing it to other relevant state-of-the-art approaches, especially with research works that evaluated their model on Alizadeh's dataset. Table 9 compares the proposed method with many similar works, such as those of (Alizadehsani et al. 2013), NN-GA by (Arabasadi et al. 2017), and information gain-SVM (Alizadehsani et al. 2016)

TABLE 9. Comparison with Other State-of-The-Art Methods

Authors	Methods	Features	AUC	ACC
(Alizadehsani et al. 2013)	Information gain, Gini index and ssociation rule mining + SMO and Naïve ayes, bagging, and NN algorithms	34	…	94.08
(Arabasadi et al. 2017)	Feature selection: weight by SVM Classification: NN with GA	22	94.50	93.85
(Alizadehsani et al. 2016)	Feature selection: weight by SVM and Information Gain. Classification: SVM algorithm	24	…	86.14
(Kiliç & Kayakeleş 2018)	ABC with Sequential minimal ptimization	16	…	89.43
(Alizadehsani et al. 2018)	Feature selection: Weight by SVM Classification: SVM algorithm	32	…	96.40
(Shahid & Singh 2020)	Feature selection: Weight by SVM Classification : EmNNs + PSO	22	…	88.34
(Khan et al. 2020)	Feature selection: Weight by Gini Index. Classification: NN+ Backward Weight Optimization	28	…	88.49
(Wiharto et al. 2021)	Feature selection: CFSS+BFS Classifier: Bagging- PART	4	95.4	94.1
(Wiharto et al. 2022)	Feature selection: GA + SVM Classification: DNN	5	93.7	87.7
Proposed Method	Feature Selection: DBHS + Adam with MLPNN	**37**	**86**	**86.66**

(Alizadehsani et al. 2013) Suggested a technique for selecting and creating effective features in CAD utilizing information gain and several algorithms, including Naïve Bayes, Sequential Minimal Optimization (SMO), Bagging, and Neural Networks. Compared to others, the SMO algorithm exhibited higher accuracy. In addition, (Alizadehsani et al. 2016) Used the Information gained with SVM to get the best CAD results. (Arabasadi et al. 2017) Applied Weight by SVM to select the best features and used GA to optimize the parameters of MLP. (Alizadehsani et al. 2018) used three classifiers (LAD, LCX, and RCA) to increase CAD detection accuracy (Alizadehsani et al. 2018) To predict coronary artery stenosis with feature engineering based on weight via SVM.

(Shahid & Singh 2020) A hybrid model of machine learning emotional neural networks with PSO was suggested to diagnose CAD, and SVM used weight to enhance the model performance. In (M. A. Khan & Algarni 2020), four feature selection methods (SVM, PSO, Information Gain, and Gini Index) and four optimization techniques, such as PSO, Evolution Strategy, Backward, and Forward, are used to improve the performance of standard neural networks (Wiharto et al. 2021)Proposed the two-tier feature selection architecture comprises correlation-based filters and wrappers. (Wiharto et al. 2022) developed a feature selection technique using GA and SVM, and DNN is used to make decisions for the diagnosis system.

CONCLUSION

Heart disease prediction aims to classify individuals who may have cardiovascular disease or are generally healthier. Heart disease classification encounters a significant challenge in dealing with numerous features, some of which may be unnecessary for the classification process. Optimizing the classifier's parameters is also essential to getting good results in cardiovascular classification. Finding appropriate parameters for the categorization process, however, might be difficult. To address the abovementioned limitations, this paper aims to present a framework that utilizes a metaheuristic algorithm-based feature selection technique as a wrapper method and parameter optimization for the MLP classifier to identify the essential features to increase accuracy with the minimum loss function. Based on the research, it was found that not all features are required to perform better in heart disease classification in the medical sector. The optimized MLP with selected features also provided superior outcomes compared to the MLP with complete features in Prec, Sen, Spec, AUC, and ACC. Therefore, It is possible to conclude that feature selection and parameter optimization produced positive results for diagnosing heart diseases. For future work, to achieve fast convergence speed, we recommend using metaheuristic approaches integrated with the ADAM algorithm to enhance global and local search in the training feed-forward neural networks.

ACKNOWLEDGEMENT

The Malaysian Ministry of Higher Education supports this work under the FRGS Grant FRGS/1/2022/ICT02/UKM/02/3.

REFERENCES

Ai, M. T., Sumiati, S. & Rosalina, V. 2021. A predictive model for heart disease diagnosis based on multinomial logistic regression. *Information Technology and Control* 50(2): 308–318. doi:10.5755/j01.itc.50.2.27672.

Al-Safi, H., Munilla, J. & Rahebi, J. 2021. Harris Hawks Optimization (HHO) Algorithm based on Artificial Neural Network for Heart Disease Diagnosis. *2021 IEEE International Conference on Mobile Networks and Wireless Communications, ICMNWC 2021* 1–5.

doi:10.1109/ICMNWC52512.2021.9688348.
Alia, O. M. D. & Mandava, R. 2011. The variants of the harmony search algorithm: An overview. *Artificial Intelligence Review* 36(1): 49–68. doi:10.1007/s10462-010-9201-y.
Alizadehsani, R., Habibi, J., Hosseini, M. J., Mashayekhi, H., Boghrati, R., Ghandeharioun, A., Bahadorian, B., et al. 2013. A data mining approach for diagnosis of coronary artery disease. *Computer Methods and Programs in Biomedicine* 111(1): 52–61. doi:10.1016/j.cmpb.2013.03.004.
Alizadehsani, R., Hosseini, M. J., Khosravi, A., Khozeimeh, F., Roshanzamir, M., Sarrafzadegan, N. & Nahavandi, S. 2018. Non-invasive detection of coronary artery disease in high-risk patients based on the stenosis prediction of separate coronary arteries. *Computer Methods and Programs in Biomedicine* 162: 119–127. doi:10.1016/j.cmpb.2018.05.009.
Alizadehsani, R., Zangooei, M. H., Hosseini, M. J., Habibi, J., Khosravi, A., Roshanzamir, M., Khozeimeh, F., et al. 2016. PT. *Knowledge-Based Systems*. doi:10.1016/j.knosys.2016.07.004.
Amarnath, B. & Appavu alias Balamurugan, S. 2016. Metaheuristic Approach for Efficient Feature Selection: A Data Classification Perspective. *Indian Journal of Science and Technology* 9(4). doi:10.17485/ijst/2016/v9i4/87039.
Arabasadi, Z., Alizadehsani, R., Roshanzamir, M. & Moosaei, H. 2017. Computer Methods and Programs in Biomedicine Computer-aided decision making for heart disease detection using hybrid neural network-Genetic algorithm. *Computer Methods and Programs in Biomedicine* 141: 19–26. doi:10.1016/j.cmpb.2017.01.004.
Attoh-Okine, N. O. 1999. Analysis of learning rate and momentum term in backpropagation neural network algorithm trained to predict pavement performance. *Advances in engineering software* 30(4): 291–302. doi:10.1016/S0965-9978(98)00071-4.
Attrapadung, N., Hamada, K., Ikarashi, D., Kikuchi, R., Matsuda, T., Mishina, I., Morita, H., et al. 2022. Adam in Private: Secure and Fast Training of Deep Neural Networks with Adaptive Moment Estimation. *Proceedings on Privacy Enhancing Technologies* 2022(4): 746–767. doi:10.56553/popets-2022-0131.
Blum, C. & Roli, A. 2008. Hybrid metaheuristics: An introduction. *Studies in Computational Intelligence* 114(2008): 1–30. doi:10.1007/978-3-540-78295-7_1.
Cardiovascular diseases. 1984.*Bulletin of the Pan American Health Organization*.
Chakarverti, M., Yadav, S. & Rajan, R. 2019. Classification Technique for Heart Disease Prediction in Data Mining. *2019 2nd International Conference on Intelligent Computing, Instrumentation and Control Technologies, ICICICT 2019* 1578–1582. doi:10.1109/ICICICT46008.2019.8993191.
Cuevas, E., Diaz, P. & Camarena, O. 2020. Metaheuristic Computation: A Performance perspective.
da Luz, E. F. P., Becceneri, J. C., de Campos Velho, H. F., Cirino, R. L., Knupp, D. C., Neto, L. B. & da Silva Neto, A. J. 2023. Firefly algorithm. *Computational Intelligence Applied to Inverse Problems in Radiative Transfer*. doi:10.1007/978-3-031-43544-7_15.
Diao, R. & Shen, Q. 2012. Feature selection with harmony search. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics* 42(6): 1509–1523. doi:10.1109/TSMCB.2012.2193613.
Dokeroglu, T., Deniz, A. & Kiziloz, H. E. 2022. A comprehensive survey on recent metaheuristics for feature selection. *Neurocomputing* 494(April): 269–296. doi:10.1016/j.neucom.2022.04.083.
El-shafiey, M. G. & Hagag, A. 2022. A hybrid GA and PSO optimized approach for heart-disease prediction based on random forest 18155–18179.
Ferreira, I. M. L. & Gil, P. J. S. 2012. Application and performance analysis of neural networks for decision support in conceptual design. *Expert Systems with Applications* 39(9): 7701–7708. doi:10.1016/j.eswa.2012.01.045.
Geem, Z. W., Kim, J. H. & Loganathan, G. V. 2001. A New Heuristic Optimization Algorithm: Harmony Search. *Simulation* 76(2): 60–68. doi:10.1177/003754970107600201
Han, J., Kamber, M. & Pei, J. 2012. Data Preprocessing. *Data Mining*. doi:10.1016/b978-0-12-381479-1.00003-4.
Haq, A. U., Li, J., Memon, M. H., Hunain Memon, M., Khan, J. & Marium, S. M. 2019. Heart Disease Prediction System Using Model of Machine Learning and Sequential Backward Selection Algorithm for Features Selection. *2019 IEEE 5th International Conference for Convergence in Technology, I2CT 2019*, hlm. 1–4. IEEE. doi:10.1109/I2CT45611.2019.9033683.
Hassan, A. A., Abdullah, S., Zamli, K. Z. & Razali, R. 2023. Q-learning whale optimization algorithm

for test suite generation with constraints support. *Neural Computing and Applications* 35(34): 24069–24090. doi:10.1007/s00521-023-09000-2.

Heron, M. 2019. Deaths: Leading causes for 2017. *National Vital Statistics Reports* 68(6).

Khamehchi, E. & Mahdiani, M. R. 2017. Optimization Algorithms. *SpringerBriefs in Petroleum Geoscience and Engineering* 35–46. doi:10.1007/978-3-319-51451-2_4.

Khan, M. A. & Algarni, F. 2020. A Healthcare Monitoring System for the Diagnosis of Heart Disease in the IoMT Cloud Environment Using MSSO-ANFIS. *IEEE Access* 8: 122259–122269. doi:10.1109/ACCESS.2020.3006424.

Khan, Y., Qamar, U., Asad, M. & Zeb, B. 2020. Applying feature selection and weight optimization techniques to enhance artificial neural network for heart disease diagnosis. *Advances in Intelligent Systems and Computing* 1037(January): 340–351. doi:10.1007/978-3-030-29516-5_26.

Kiliç, Ü. & Kayakeleş, M. 2018. Feature Selection with Artificial Bee Colony Algorithm on Z-Alizadeh Sani Dataset. *Proceedings - 2018 Innovations in Intelligent Systems and Applications Conference, ASYU 2018* (January). doi:10.1109/ASYU.2018.8554004.

Kingma, D. P. & Ba, J. L. 2015. Adam: A method for stochastic optimization. *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings* 1–15.

Latha, C. B. C. & Jeeva, S. C. 2019. Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques. *Informatics in Medicine Unlocked* 16(June): 100203. doi:10.1016/j.imu.2019.100203.

Lee KS, Geem ZW, Lee SH & Bae KW. 2005. The harmony search heuristic algorithm for discrete structural optimization. *Engineering Optimization* 37(7): 663–684.

Lin, J. H. & Li, Y. L. 2012. A metaheuristic optimization algorithm for unsupervised robotic learning. *Proceeding - 2012 IEEE International Conference on Computational Intelligence and Cybernetics, CyberneticsCom 2012* 113–117. doi:10.1109/CyberneticsCom.2012.6381629.

Manjarres, D., Landa-Torres, I., Gil-Lopez, S., Del Ser, J., Bilbao, M. N., Salcedo-Sanz, S. & Geem, Z. W. 2013. A survey on applications of the harmony search algorithm. *Engineering Applications of Artificial Intelligence* 26(8): 1818–1831. doi:10.1016/j.engappai.2013.05.008.

Mansor, N. F., Abas, Z. A., Rahman, A. F. N. A., Shibghatullah, A. S. & Sidek, S. 2016. A new HMCR parameter of harmony search for better exploration. *Advances in Intelligent Systems and Computing* 382: 181–195. doi:10.1007/978-3-662-47926-1_18.

Muhajir, M., Aminuddin, A., Ugusman, A., Salamt, N., Asmawi, Z., Zulkefli, A. F., Azmi, M. F., et al. 2018. Evaluation of finger photoplethysmography fitness index on young women with cardiovascular disease risk factors. *Sains Malaysiana* 47(10): 2481–2489. doi:10.17576/jsm-2018-4710-25.

Newaz, A., Muhtadi, S. & Muhtadi, S. 2021. Performance Improvement of Heart Disease Prediction by Identifying Optimal Feature Sets Using Feature Selection Technique. *2021 International Conference on Information Technology (ICIT)* 446–450. doi:10.1109/ICIT52682.2021.9491739.

Pravin, A. 2021. An Efficient Feature Selection Approach using Sensitivity Analysis for Machine Learning based Heart Disease Classification. *2021 10th IEEE International Conference on Communication Systems and Network Technologies (CSNT)* 539–542. doi:10.1109/CSNT51715.2021.9509673.

Rahajoe, A. D., Fahrial Zainal, R., Mulyo, B. M., Plangkang, B. & Tias, R. F. 2020. Feature Selection Based on Modified Harmony Search Algorithm. *Proceeding - ICoSTA 2020: 2020 International Conference on Smart Technology and Applications: Empowering Industrial IoT by Implementing Green Technology for Sustainable Development*. doi:10.1109/ICoSTA48221.2020.1570615299.

Ruder, S. 2016. An overview of gradient descent optimization algorithms 1–14. Retrieved from http://arxiv.org/abs/1609.04747.

Samuel, O. W., Asogbon, G. M., Sangaiah, A. K., Fang, P. & Li, G. 2017. An integrated decision support system based on ANN and Fuzzy_AHP for heart failure risk prediction. *Expert Systems with Applications* 68: 163–172. doi:10.1016/j.eswa.2016.10.020.

Sarhani, M., El Afia, A. & Faizi, R. 2018. Facing the feature selection problem with a binary PSO-GSA approach. *Operations Research/ Computer Science Interfaces Series* 62: 447–462.

doi:10.1007/978-3-319-58253-5_26.
Shahid, A. H. & Singh, M. P. 2020. A Novel Approach for Coronary Artery Disease Diagnosis using Hybrid Particle Swarm Optimization based Emotional Neural Network. *Biocybernetics and Biomedical Engineering* 40(4): 1568–1585. doi:10.1016/j.bbe.2020.09.005.
Shreem, S. S., Abdullah, S. & Nazri, M. Z. A. 2014. Hybridising harmony search with a Markov blanket for gene selection problems. *Information Sciences* 258: 108–121. doi:10.1016/j.ins.2013.10.012.
Singarimbun, R. N., Nababan, E. B. & Sitompul, O. S. 2019. Adaptive Moment Estimation to Minimize Square Error in Backpropagation Algorithm. *2019 International Conference of Computer Science and Information Technology, ICoSNIKOM 2019* 04(1): 27–46. doi:10.1109/ICoSNIKOM48755.2019.9111563.
Singh, V., Kumar, L., Patel, A. K. & Krishna, A. 2023. An Empirical Framework for Malware Prediction Using Multi-Layer Perceptron. *OCIT 2023 - 21st International Conference on Information Technology, Proceedings* (Ocit): 485–490. doi:10.1109/OCIT59427.2023.10430935
Subhadra, K. & Vikas, B. 2019. Neural network based intelligent system for predicting heart disease. *International Journal of Innovative Technology and Exploring Engineering* 8(5): 484–487.
Taha Yassen, E., Ayob, M., Ahmad Nazri, M. Z. & Sabar, N. R. 2015. Meta-harmony search algorithm for the vehicle routing problem with time windows. *Information Sciences* 325: 140–158. doi:10.1016/j.ins.2015.07.009.
Tsai, C. & Lee, Y. 2011. Expert Systems with Applications The parameters effect on performance in ANN for hand gesture recognition system. *Expert Systems With Applications* 38(7): 7980–7983. doi:10.1016/j.eswa.2010.12.086.
Uppal, M., Gupta, D., Juneja, S., Gadekallu, T. R., El Bayoumy, I., Hussain, J. & Lee, S. W. 2023. Enhancing accuracy in brain stroke detection: Multi-layer perceptron with Adadelta, RMSProp and AdaMax optimizers. *Frontiers in Bioengineering and Biotechnology* 11(September): 1–15. doi:10.3389/fbioe.2023.1257591.
Wang, L., Xu, Y., Mao, Y. & Fei, M. 2010. 2 Discrete Binary Harmony Search Algorithm 37–38.
Wang, L., Yang, R., Xu, Y., Niu, Q., Pardalos, P. M. & Fei, M. 2013. An improved adaptive binary Harmony Search algorithm. *Information Sciences* 232(December 2012): 58–87. doi:10.1016/j.ins.2012.12.043.
Wang, L., Zhou, P., Fang, J. & Niu, Q. 2011. A hybrid binary harmony search algorithm inspired by ant system. *Proceedings of the 2011 IEEE 5th International Conference on Cybernetics and Intelligent Systems, CIS 2011* 153–158. doi:10.1109/ICCIS.2011.6070319.
Wiharto, Suryani, E. & Setyawan, S. 2021. Framework Two-Tier Feature Selection on the Intelligence System Model for Detecting Coronary Heart Disease. *Ingenierie des Systemes d'Information* 26(6): 541–547. doi:10.18280/isi.260604.
Wiharto, Suryani, E., Setyawan, S. & Putra, B. P. 2022. The Cost-Based Feature Selection Model for Coronary Heart Disease Diagnosis System Using Deep Neural Network. *IEEE Access* 10: 29687–29697. doi:10.1109/ACCESS.2022.3158752.
Wolpert, D. H., & Macready, W. G. (1997). No free lunch theorems for optimization. IEEE Transactions on Evolutionary Computation, 1(1), 67–82. https://doi.org/10.1109/4235.585893 .
Xu, J. & Zhang, J. 2014. Exploration-exploitation tradeoffs in metaheuristics: Survey and analysis. *Proceedings of the 33rd Chinese Control Conference, CCC 2014* 8633–8638. doi:10.1109/ChiCC.2014.6896450.
Yusta, S. C. 2009. Different metaheuristic strategies to solve the feature selection problem. *Pattern Recognition Letters* 30(5): 525–534. doi:10.1016/j.patrec.2008.11.012.
Zainuddin, Z., Lai, K. H. & Ong, P. 2016. An enhanced harmony search based algorithm for feature selection: Applications in epileptic seizure detection and prediction. *Computers and Electrical Engineering* 53: 143–162. doi:10.1016/j.compeleceng.2016.02.009.
Zarafshan, P., Etezadi, H., Javadi, S., Roozbahani, A., Hashemy, S. M. & Zarafshan, P. 2023. Comparison of machine learning models for predicting groundwater level, case study: Najafabad region. *Acta Geophysica* 71(4): 1817–1830. doi:10.1007/s11600-022-00948-8.
Zheng, L., Diao, R. & Shen, Q. 2015. Self-adjusting harmony search-based feature selection. *Soft Computing* 19(6): 1567–1579. doi:10.1007/s00500-014-1307-8.