



OPEN

Intelligent cardiovascular disease diagnosis using deep learning enhanced neural network with ant colony optimization

Biao Xia^{1✉}, Nisreen Innab², Venkatachalam Kandasamy³, Ali Ahmadian⁴ & Massimiliano Ferrara^{5✉}

To identify patterns in big medical datasets and use Deep Learning and Machine Learning (ML) to reliably diagnose Cardio Vascular Disease (CVD), researchers are currently delving deeply into these fields. Training on large datasets and producing highly accurate validation results is exceedingly difficult. Furthermore, early and precise diagnosis is necessary due to the increased global prevalence of cardiovascular disease (CVD). However, the increasing complexity of healthcare datasets makes it challenging to detect feature connections and produce precise predictions. To address these issues, the Intelligent Cardiovascular Disease Diagnosis based on Ant Colony Optimisation with Enhanced Deep Learning (ICVD-ACOEDL) model was developed. This model employs feature selection (FS) and hyperparameter optimization to diagnose CVD. Applying a min–max scaler, medical data is first consistently prepared. The key feature that sets ICVD-ACOEDL apart is the use of Ant Colony Optimisation (ACO) to select an optimal feature subset, which in turn helps to upgrade the performance of the ensuring deep learning enhanced neural network (DLENN) classifier. The model reforms the hyperparameters of DLENN for CVD classification using Bayesian optimization. Comprehensive evaluations on benchmark medical datasets show that ICVD-ACOEDL exceeds existing techniques, indicating that it could have a significant impact on CVD diagnosis. The model furnishes a workable way to increase CVD classification efficiency and accuracy in real-world medical situations by incorporating ACO for feature selection, min–max scaling for data pre-processing, and Bayesian optimization for hyperparameter tweaking.

Keywords Cardiovascular disease, Ant Colony Optimisation, Min–max scaler, Bayesian optimisation, Hyperparameter

Cardiovascular disease (CVD) is the primary cause of death worldwide¹. Delays in diagnosis cause a great deal of deaths. Early CVD risk prediction has the potential to greatly improve this terrible situation². Cardiovascular disorders (CVDs) include heart failure, stroke, coronary artery disease, and hypertension. CVDs are the leading cause of mortality and disability globally³. Risk factors include lifestyle choices and genetics. Healthcare providers play a crucial role in early detection and preventive education, which includes encouraging healthy living and arranging regular checkups.

Cardiovascular diseases (CVDs) present with a variety of symptoms. Heart failure and coronary artery disease are two illnesses that can cause symptoms like chest pain, dyspnea, tiredness, and irregular heartbeats⁴. On the other hand, symptoms of aneurysms and strokes might include severe headaches, difficulty speaking, blurred vision, and numbness or paralysis on one side of the body. It's crucial to keep in mind that several CVDs have symptoms that are similar to one another, making it challenging to distinguish between them based alone on these markers. For instance, chest pain can be brought on by coronary artery disease as well as heart failure. In addition to being common indicators of cardiovascular diseases (CVDs), exhaustion and shortness of breath can

¹Medical Equipment Department, Changzhou No2 Hospital Nanjing Medical University, Changzhou 213164, Jiangsu, China. ²Department of Computer Science and Information Systems, College of Applied Sciences, AlMaarefa University, 13713 Diriyah, Riyadh, Saudi Arabia. ³Department of Mathematics, Faculty of Science, University of Hradec Kralove, Hradec Kralove, Czech Republic. ⁴Faculty of Engineering and Natural Sciences, Istanbul Okan University, Istanbul, Turkey. ⁵ICRIOS, University Bocconi, Via Röntgen no 1, 20136 Milan, Italy. ✉email: stephenxiabiao@sina.com; massimiliano.ferrara@unibocconi.it

also be indicators of heart failure hypertension, or high blood pressure. A comprehensive medical evaluation involving imaging scans, laboratory testing, and physical examinations is required to appropriately diagnose and treat CVDs. Sometimes, specialized testing is required to establish a diagnosis and decide on the best course of treatment⁵. This emphasizes how important it is to preserve cardiovascular health with a thorough and tailored approach.

Cardiovascular diseases (CVDs) are disorders that affect the heart or blood vessels, including the arteries and veins. Heart disease has become a “second epidemic” in many countries due to its high and rising incidence, surpassing infectious diseases as the leading cause of death and disability worldwide among those over 65⁶. A reduction in death rates depends on early identification. Echocardiography is a useful diagnostic technique for heart disorders. This non-invasive examination provides information about the size, shape, and function of the heart’s chambers and valves by using sound waves to create images of the organ. Crucially, echo can identify heart anomalies in newborns and early children⁷.

With advanced deep-learning neural networks, it is possible to assess and analyze cardiovascular disease (CVD) indicators and post-determined symptoms in an efficient manner. Several clinical datasets are used by methods like Deep Belief Networks, Medical Image Segmentation^{8,9}, and Bayesian Networks to forecast the effects of risk variables. The purpose of this project is to develop a predictive model that uses machine learning algorithms to accurately identify heart disease symptoms in the United Arab Emirates (UAE). The results of this study could improve cardiovascular health in the United Arab Emirates by improving early identification and aggressive care.

Fortunately, a wealth of medical information may be found in patient-provided electronic medical records (EMRs). This knowledge will pave the way for a period when CVD is detected early and treated aggressively, saving many lives. The development of decision support systems (DSS) that utilize patient data, artificial intelligence (AI), and subject-matter expertise is being fueled by the increasing adoption of data in healthcare¹⁰. Employing patient data analysis, critical situation recognition, error detection, and prompt action, these DSSs aim to provide healthcare workers with valuable information and assistance. This is particularly crucial for complex illnesses like cardiovascular disease (CVD), which is influenced by a multitude of factors such as diabetes, hypertension, smoking, and a sedentary lifestyle. Traditional manual analysis of CVD data is laborious and error-prone, which emphasizes the need for more precise and effective methods like AI-powered DSSs¹¹.

Due to the growing significance of huge data in healthcare, artificial intelligence (AI) techniques—particularly those based on deep learning (DL) and machine learning (ML)—are being employed more and more to address difficult problems in diagnosis and prediction¹². These AI algorithms analyze vast amounts of healthcare data, making them more accurate in predicting the presence of disease than manual diagnosis techniques. ML-driven diagnostic models that can predict disease based on individual risk profiles and can integrate into clinical decision support systems are noteworthy among them¹³. Machine learning methods are a useful tool for improving patient outcomes and diagnostic accuracy since they have a lot of potential in cardiovascular care. Machine learning (ML) is essentially based on models that use mathematical optimization and statistical analysis to take in input (text, images, etc.) and produce desired predictions (healthy, unwell, or not ill)¹⁴. Since these models are trained on massive volumes of raw electronic medical records collected from moderately priced wearable devices, they provide accurate and resource-efficient diagnoses of heart illness¹⁵.

Deep learning has emerged as a crucial element in the detection of cardiovascular diseases (CVDs) in recent years¹⁶. These algorithms are highly skilled at identifying patterns and learning from vast amounts of complex data. They are a type of artificial neural network. They are particularly prevalent in medical imaging and diagnostic testing. Deep learning has proven to be especially helpful in the detection and diagnosis of cardiac disease using imaging tests like echocardiograms, MRIs, and angiograms. Its potential impact on enhancing cardiovascular healthcare is seen in the increased accuracy and effectiveness of cardiovascular disease (CVD) diagnosis due to its ability to spot minor trends in such data.

Deep learning algorithms are very skilled at spotting specific patterns and traits in medical images that can point to heart issues. These algorithms can identify abnormalities in heart architecture, blood flow, and tissue texture by analyzing images¹⁷. Cardiovascular data, such as electrocardiograms (ECGs) and patient health records are included in deep learning beyond image analysis. These algorithms use ECG data to predict the probability of cardiac events such as arrhythmias. To estimate the risk of cardiovascular disease based on factors including age, sex, family history, and lifestyle, they also review patient medical data. Although deep learning for the diagnosis of cardiovascular illness is still in its early stages, the preliminary results are promising. As data becomes more available and technology advances, deep learning is poised to become a vital tool in the ongoing battle against heart disease¹⁸.

Particularly in the fields of image analysis and natural language processing, deep learning has shown tremendous promise in the medical field^{19,20}. Deep learning extracts patterns and forecasts from massive datasets employing artificial neural networks. Deep learning has the primary advantage of eliminating the need for tedious human feature engineering by automatically extracting features from raw data. Because of their ability to handle massive datasets and adapt well to novel tasks, these models are perfect for real-world healthcare applications^{21,22}. Feature extraction from deep learning-based data patterns is used in regression analysis, which makes heart disease (CVD) predictions²³. Another study looks at using machine learning algorithms to predict epidemics of chronic illnesses in high disease prevalence areas²⁴. This study employed latent component modeling to predict outcomes to fill in the gaps in the hospital data collected in China between 2013 and 2015. The authors identified which chronic illnesses were prevalent in the area by using a CNN algorithm to extract features from statistical data and unstructured text. Ultimately, they proposed a CNN-based multimodal sickness risk prediction method for both structured and unstructured data. These applications show how versatile and promising deep learning is in a range of healthcare settings.

A statistical model for forecasting continuous outcomes based on predictor variables is called regression analysis. Regression models can be used to evaluate the risk of cardiovascular disease (CVD) based on factors such as age, blood pressure, cholesterol, and other relevant risk markers²⁵. Depending on the kind of data and the prediction job, several regression model types, such as logistic, non-linear, and linear regression, may be used. An essential step in the regression modeling process is a comprehensive evaluation for accuracy and dependability. The dataset is divided into test, validation, and training sets to confirm that the model is generalizable. After that, the model is assessed using the test set. Techniques for cross-validation improve the model's performance estimations even more. Regression models can be used to identify people who are at high risk for cardiovascular disease. These models use characteristics related to cardiovascular risk to help formulate well-informed judgments about therapeutic medicines and preventive measures²⁶.

The identification of cardiovascular disease (CVD) is a difficult task that requires figuring out which model is best for a certain task and managing a large number of features. It has been discovered that the two most important steps in considerably improving the performance of CVD detection systems are feature selection (FS) and hyperparameter tuning. Finding the most pertinent features for the CVD detection task is FS's main goal. It is essential since employing an excessive number of features could cause overfitting, which would reduce the model's applicability to a wider population. The model focuses on the most informative data and produces precise predictions by picking the most important characteristics^{27,28}.

Hyperparameter tuning is the process of adjusting the model's parameters to achieve the best performance achievable. By fine-tuning these hyperparameters, the method performs better since it is more appropriate for the specific CVD detection issue. In summary, the identification of CVD requires both hyperparameter tuning and FS. By applying these strategies, the algorithm can improve its generalization performance, reduce over fitting, and raise accuracy. This paper proposes a novel Intelligent Cardiovascular Disease Diagnosis model (ICVD-ACOEDL) that combines Ant Colony Optimisation with an enhanced deep learning approach. The ICVD-ACOEDL strategy uses the ACO approach to select the optimal features and min-max scaling is employed for preprocessing medical data. Next, a deep learning-enhanced neural network (DLENN) classifier is used to classify CVDs. Bayesian techniques are used to optimize the hyperparameters of this DLENN classifier. ICVD-ACOEDL's efficacy can be assessed with reference datasets from the medical domain.

Numerous noteworthy advancements in the field of cardiovascular disease (CVD) diagnosis are provided by Novel Concepts for Better CVD Diagnosis research, including The Deep Learning Enhanced Neural Network (DLENN) classification, min-max normalization, Ant Colony Optimisation (ACO)-based feature selection, and Bayesian optimization are all combined in the ICVD-ACOEDL Approach. Notably, this combination technique has not yet been documented in the literature for the diagnosis of CVD. An innovative ACO-driven technique dubbed "ACO-based Feature Selection" is proposed to select an optimal subset of features that could improve model performance and reduce computing expenses. Bayesian Optimization for DLENN: A cross-validation Bayesian optimization approach is used to optimize the hyperparameters of the DLENN classifier. Through this optimization, the model's predictive accuracy for yet-to-be-observed data is improved, potentially increasing its overall efficacy in the diagnosis of cardiovascular disease.

Since cardiovascular diseases (CVDs) are one of the main causes of death, they pose a serious threat to world health. Effective disease management and better patient outcomes depend on early and precise diagnosis. The intricacy and size of medical information, however, frequently make it difficult to make an accurate diagnosis of CVDs. Large and complicated datasets are a common source of trouble for traditional machine learning and deep learning models used to diagnose CVD, which leads to subpar predictive accuracy. Furthermore, it takes a lot of time and may not produce the best results to manually tune features and model parameters in these models.

The Intelligent Cardiovascular Disease detection based on Ant Colony Optimisation with Enhanced Deep Learning (ICVD-ACOEDL) model is presented in this research proposal. It uses innovative techniques to improve the detection of cardiovascular disease. This model is novel in that it combines hyperparameter optimization with Bayesian optimization to optimize and expedite the learning process, along with feature selection via Ant Colony Optimisation (ACO).

ICVD-ACOEDL intends to lessen reliance on personal intervention and domain expertise by automating the feature selection and hyperparameter tuning procedures. This will make the system easier to deploy and more effective in a variety of healthcare contexts. This model has the potential to transform the area of medical diagnostics by offering not only increased accuracy in diagnosing cardiovascular disease (CVD) but also scalability to other complicated diseases.

The remaining part of the paper is structured as follows: the first section on the proposed methodology describes the algorithmic model of the data scaling, ACO-based FS technique, and CVD classification using the optimal DLENN approach, among other steps of the suggested model. The initially related studies provide research on deep neural network classifiers for Bayesian optimisation, CVD, and ACO. This paper's last section compares and evaluates the model's overall strength in the performance validation section.

Related works

Several machine learning (ML) techniques for the identification of heart disease (HD) were investigated by Nagavelli et al.²⁹. They looked into four main methods: HD is predicted by Naive Bayes employing a weight system and a probabilistic model based on the Bayes theorem. Different features can have their weights assigned to them to increase prediction accuracy. In order to detect or localize ischemia HD, an automated analytic technique based on the time, data model, and frequency domains evaluates data from these three domains. The Support Vector Machine (SVM) with Extreme Gradient Boosting (XGBoost) approach blends the potent classifier capabilities of SVM with the ensemble learning algorithm XGBoost, which is based on trees. When it comes to HD classification, this combination can reach excellent accuracy.

This enhanced SVM with a duality-optimized strategy for cardiac arrest detection technology uses an improved SVM model that has undergone duality-based methodology optimization to automatically identify cardiac arrest. There are several methods for diagnosing HD, and each has benefits and drawbacks. Which choice is ideal for a given application depends on a number of factors, such as processing power, required precision, and available data. Table 1 provides a summary of all four approaches.

It is crucial to remember that there are other ML approaches being developed for HD diagnosis, of which the study by Nagavelli et al. is only one example. We should anticipate the emergence of increasingly more advanced and precise techniques as this field develops. This study³⁰ investigates two methods for assessing cardiovascular health: (1) using soft sensors to collect real-time data for the extraction of key features and the assessment of cardiovascular risk, such as blood oxygenation, ECG, temperature, and interview information; and (2) using an optimized Artificial Neural Network (OANN) to analyze heart disease, which is the leading cause of death worldwide. This OANN integrates two key functions: Teaching–Learning-Based Optimization (TLBO) optimizes its hyperparameters for improved performance, and Distance Based Misclassified Instance Removal (DBMIR) eliminates misclassified data points to increase model accuracy. The proposed OANN provides a feasible solution for continuous monitoring and risk assessment. It operates in two phases: offline training using previous data and online cardiovascular risk prediction for new patients³¹.

The goal of Ali et al.³² was to determine which machine learning (ML) methods resulted in the best accuracy when it came to heart disease prediction. They looked into and compared the accuracy and effectiveness of several supervised machine-learning methods for HD prediction. Interestingly, they determined feature relevance ratings for every feature that was employed in the algorithms—K-Nearest Neighbours (KNN) and Multilayer Perceptrons (MLP) excluded. Based on these evaluations, they ranked each attribute to determine which ones were most important in producing reliable HD predictions. This study³³ uses health service sensors and the UCI Repository database to create a successful environment for heart disease (HD) prediction. The proposed method uses classifier algorithms for patient data analysis to identify individuals with HD. During the training phase, the classifiers are trained using the benchmark dataset. Subsequently, during the testing phase, patient data is utilized to predict the presence of HD.

The study evaluates the efficiency of multiple classifiers, including Multilayer Perceptron (MLP), J48, Support Vector Machine (SVM), and Logistic Regression (LR), on the benchmark database. Li et al.³⁴ introduced an effective machine learning-based method for evaluating heart disease. Their method merged several classifier approaches (e.g., SVM, LR, ANN, KNN, NB, and DT) with conventional methods of feature selection (e.g., Relief, Least Absolute Shrinkage as well as Selection Operator, Minimal Redundancy Maximal Relevance, and Local Learning) to eliminate inapplicable and unneeded attributes. Atallah and Al-Mousa³⁵ developed an overwhelming ensemble algorithm to predict the incidence of heart disease using readily available medical testing. Through the use of real patient data and the integration of many machine learning algorithms' predictions, this system aimed to increase doctors' confidence and accuracy in their diagnosis. Results may be accurate when comparing this approach to utilizing just one method.

The epidemiology of cardiovascular diseases (CVDs) among adults 65 years of age and older was investigated in a study by Noale et al.³⁶. Their study concentrated on risk factors associated with prevalent causes of disability and death in this age group, such as obesity, lipids and lipoproteins, smoking, diabetes, physical inactivity, high blood pressure, and a higher chance of acute stroke. The findings of Athanasiou et al. strengthen this focus. The use of risk prediction models for cardiovascular disease (CVD) was examined by³⁷ in their study. These models employ several techniques to enhance the interpretability of the cognitive process involved in CVD risk assessment, promote trust in the results, and increase clarity. The XGBoost technique, renowned for its rapid computation speed, robust generalization capability, and superior prediction accuracy, is one of the techniques evaluated. The study looked at how these techniques could enhance understanding and enhance existing models such as the Framingham CVD risk calculator, SCORE, DECODE, and various other approaches. The researchers proposed a new approach to predict CVD risk by constructing a comprehensible model with the help of the XGBoost and SHAP approaches.

Charlton et al.³⁸ provided an overview of wearable photoplethysmography (PPG) and its analysis. A possibility for simple, non-intrusive cardiovascular health tracking has been generated by the growing demand for wearable smart devices with PPG sensors. The PPG signal, which measures arterial blood volume, and an assortment of proposed combinations are examined in this study. The authors list several potential applications for wearable PPG sensors, including the diagnosis of atrial fibrillation, the monitoring of infectious diseases, the assessment of mental stress, and the identification of obstructive sleep apnea. The study essentially emphasizes the wearable PPG devices' potential for everyday cardiovascular health monitoring.

Techniques	Description	Strengths	Weaknesses
Naïve Bayes with a weight system	Predicts HD based on Bayes theorem	Simple to implement	May not capture complex relationships between features
Time domain, data model, and frequency domains based analysis	Automatically analyzes data from different domains	Can identify subtle patterns in the data	Requires expertise in signal processing
SVM with XGBoost	Combines a powerful classifier (SVM) with an ensemble learning algorithm (XGBoost)	High accuracy	Can be computationally expensive
Enhanced SVM with a duality optimized method for cardiac arrest detection	Utilizes an improved SVM model optimized for cardiac arrest detection	High accuracy for cardiac arrest detection	May not be as effective for other types of HD

Table 1. Machine Learning (ML) techniques for the identification of heart disease (HD).

Chieng and Kistler³⁹ examined the benefits of regular tea and coffee use in averting cardiovascular disease (CVD). Depending on their study, two to three cups of coffee per day can help persons with metabolic syndrome. The study also discovered a link between coffee drinking and a lower risk of coronary heart disease. Remarkably, a daily intake of seven cups of coffee is linked to a 9% reduction in the chance of hypertension. Consuming tea and coffee may also help prevent cardiovascular diseases (CVDs), which include arrhythmia, cardiac arrest, stroke, and cardiovascular disease-related death.

The triglyceride-glucose (TyG) index was investigated by Tao et al.⁴⁰ in a review as a potential biomarker for insulin resistance (IR), a risk factor for cardiovascular disease (CVD). The TyG index is a reliable and practical technique for assessing insulin resistance (IR) and predicting conditions such as hypertension and obesity by integrating lipid and sugar elements. Studies show that the TyG index is a more accurate measure of insulin resistance in individuals with and without diabetes than the widely used HOMA-IR. The authors propose that risk classification and CVD outcome forecasts could be enhanced through TyG score optimisation.

Battineni et al.⁴¹ recommended doing a study to assess how well telemedicine technologies work in offering individualized care to prevent cardiovascular disease (CVD) and lessen the need for hospital visits. Telemedicine systems apply precision medicine ideas at every stage of cardiovascular disease (CVD) and offer routine check-ups to patients with CVD. Telemedicine involves the exchange of clinical information via text, audio, and visuals to facilitate safe phone or video consultations between patients and medical professionals. The application of telemedicine to cardiovascular medicine is known as telecardiology, and it offers many benefits for remote patient management. This research highlights the value of telemedicine in helping patients with cardiovascular disease (CVD) receive customized treatment.

Bays et al.⁴² provide a summary of the various risk factors connected to CVD in a paper. Ten key risk factors were identified by the survey: high blood pressure, dyslipidemia, hyperglycemia, obesity, genetics/familial hypercholesterolemia, kidney dysfunction, thrombosis/smoking, and demographics (such as age, race/ethnicity, and sex differences). Numerous risk factors may be present in patients with CVD, therefore preventative strategies need to be diverse. The book emphasizes the role that exercise, healthy eating, and intermittent fasting play in reducing the risk of CVD.

The study highlights how important it is to recognize and control the numerous risk factors connected to cardiovascular disease (CVD) to enhance prevention strategies. Dickson et al.⁴³ investigated the effects of work-related factors on older workers with CVD and their self-care practices using the middle-range theory of self-care. This idea holds that self-care is an organic decision-making process. The health of workers with cardiovascular disease may be negatively impacted by work-related problems such as job stress brought on by disorganized work settings because these conditions increase physiological demands. Important factors that can reduce the risk of CVD at work are examined in the study, including job control, workplace support, work-life balance, and organizational fairness. To overcome obstacles to self-care and encourage older workers with CVD to practice sustainable self-care at work, a multi-level intervention is considered important.

Physical symptoms that may mimic those of other medical conditions are often considered when diagnosing depression. Ellis et al. conducted research on the connection between depression symptoms and disease markers⁴⁴. As part of the author's research, interviews were done with 46 out of 61 suitable groups, primarily older individuals with advanced illnesses and a high frequency of somatic depressive symptoms⁴⁵. Answers to interactive questions on emotions and legal queries about symptoms of depression were provided by the participants (Table 2).

Although the precise relationship between dyspepsia and gastroesophageal reflux disease (GERD) is unknown, Hsu et al.⁴⁶ demonstrate that these conditions frequently coexist in the general population. There were overlaps with "epigastric pain or burning" in 107 persons, while these symptoms were absent in 761. More severe GERD symptoms were shown by those with GERD-D, and they were more commonly linked to GERD alone (IBS) syndrome⁴⁷. Cross-over cardiovascular issues, such palpitations, chest tightness, and dyspnea, which are frequently noticed in well-performing individuals during stressful situations, make it difficult for medical professionals and patients to attribute the cause or co-occurring roles to mental health. Cardiologists and primary care physicians frequently place a higher priority on controlling risk factors and physical symptoms than on addressing mental and emotional issues⁴⁸. A widespread concern over routine activities has been associated to promoting and accelerating cardiovascular disease in psychological discomfort, including tension and agony, even though it does not fulfill the formal diagnostic criteria for generalized anxiety disorder⁴⁹.

Existing models often struggle to manage the high dimensionality and heterogeneity inherent in medical datasets. This complexity can hinder the ability of traditional machine learning and deep learning algorithms to extract meaningful patterns and make accurate predictions. Many current diagnostic models lack robust mechanisms for optimal feature selection, which is critical for improving the accuracy and efficiency of predictions. Manual feature selection is prevalent, which is not only labor-intensive but also prone to human bias and error.

Proposed model

A novel ICVD-ACOEDL method has been created in this study to improve the accuracy of the diagnosis of cardiovascular disease (CVD). Data scaling, feature subset selection based on Ant Colony Optimisation (ACO), Deep Learning Enhanced Neural Networks (DLENN) for categorization, and Bayesian Optimisation (BO) for parameter optimization are all integral parts of the method.

Figure 1 shows the ICVD-ACOEDL method's steps as well as the smooth progression of the diagnostic process. By merging deep learning with DLENN for robust classification and ACO for feature selection, this novel method picks features more effectively. By using Bayesian Optimisation, the overall accuracy of the CVD diagnosis is improved as proper hyperparameter tuning is ensured.

Class	Accuracy (a_b)	Precision (p_n)	Recall (r_l)	F-score (f_s)	G-measure (g_m)
400 epochs					
Normal	94.01	97.12	94.11	94.79	94.06
Diseased	97.04	94.93	97.01	97.03	95.32
Average	95.53	96.33	95.56	95.91	94.69
800 epochs					
Normal	95.11	97.99	95.18	95.99	95.56
Diseased	98.14	95.51	98.12	97.88	96.12
Average	96.63	96.75	96.65	96.94	95.84
1200 epochs					
Normal	88.11	97.99	88.18	93.19	93.16
Diseased	98.44	90.51	98.82	94.98	94.12
Average	93.28	94.25	93.50	94.08	93.64
1600 epochs					
Normal	95.51	96.01	95.38	95.79	95.56
Diseased	96.14	95.81	98.12	96.88	96.42
Average	95.83	95.91	96.75	96.34	95.99
2000 epochs					
Normal	94.01	96.19	94.18	95.29	95.16
Diseased	97.14	95.11	97.21	96.08	96.10
Average	95.58	95.65	95.70	95.69	95.63
2400 epochs					
Normal	99.11	99.89	98.98	99.49	99.56
Diseased	99.94	99.13	99.94	99.58	99.52
Average	99.53	99.51	99.46	99.54	99.54

Table 2. Complete outcome of the CVD categorization using the ICVD-ACOEDL method.



Fig. 1. Workflow of ICVD-ACOEDL algorithm.

Data scaling process

The ICVD-ACOEDL method uses a min-max scaler as its first step in preparing medical data. This scaler converts the original data into a linear form using the Min-Max normalization approach. The maximal (MaxY) and minimal (MinX) values of attribute A dictate the transformation. By utilizing the threshold values N_MinX and N_MaxX, the Min-Max normalization translates the value of attribute A to M' within the given range, which is usually zero to one. This method ensures that the data is standardized and normalized for use in subsequent research phases.

$$M' = \frac{M - MinX}{MaxX - MinA} (N_MaxX - N_MinX) + N_MinX \tag{1}$$

Biases towards qualities with larger magnitudes are less likely when data is normalized. In general, scaled data makes the learning algorithm run more quickly since features with the same scale of gradient descent converge

more quickly. This prevents computation-related numerical instabilities that may result from extremely high or low attribute values.

Algorithmic design of ACO based FS technique

Ants use a team-based foraging strategy to find the simplest pathways to food sources, which served as the model for the ACO algorithm. Conventional ACO-based feature selection techniques frequently depend on feature knowledge that is known ahead of time. By using classifier performance and the length of the chosen feature vector as heuristic information for ACO, on the other hand, the suggested algorithm presents a fresh approach. This change makes the feature selection process more sensitive and adaptive, which encourages the development of a more dynamic and efficient algorithm. A graph with a collection of nodes and edges connecting nodes must be used to represent the problem. The phases of the Ant Colony Optimization Technique (ACO) are shown in Fig. 2.

1. Heuristic Desirability (η): Assists in effective path evaluation, this measure indicates the “goodness” of paths from one node to all associated nodes in the network.
2. Construction of Feasible Solutions: This refers to setting up a system that effectively generates potential solutions within the constraints of the situation.
3. Pheromone Updating Rule: This refers to a technique that updates the pheromone levels on edges. It usually entails choosing the best- n ants and making changes to the paths they have selected. One evaporation rule to control pheromone persistence is included in this.
4. Probabilistic Transition Rule: Determines the likelihood of a path traversal during the exploration phase by controlling the probability of an ant moving from one node in the graph to the next.

The problem is translated into an appropriate form for Ant Colony Optimization (ACO) in the feature selection context. The ACO framework requires a graph representation in which features are represented by nodes and the selection of the next feature is indicated by edges. Finding the best feature subset becomes like an ant moving across a graph and trying to visit as many nodes as possible that satisfy the traversal stopping criterion. This configuration is shown in Fig. 1, where an ant at node ‘a’ has the choice to add the next feature (shown by dotted lines) to its journey (Fig. 3).

The ant chooses feature ‘b’ first, then ‘c’, and lastly ‘d’ in accordance with the probabilistic transition rule. When the current subset {a; b; c; d} reaches ‘d’, it meets the traversal halting requirement with a high enough classification accuracy. As it comes to an end, the ant outputs this feature subset as a potential option for data reduction. The likelihood that an ant at feature ‘i’ will choose to migrate to feature ‘j’ at time ‘t’ is determined by the probabilistic transition rule, which is based on the combined heuristic attractiveness of traversal and edge pheromone levels. The deep learning-enhanced neural network classifier is shown in Fig. 3.

$$C_{ij}^n(t) = \begin{cases} t \frac{[\mu_{ij}(t)]^\vartheta \cdot [\delta_{ij}]^\sigma}{\sum_{k \in f_i^n} [\mu_{ik}(t)]^\vartheta \cdot [\delta_{ik}]^\sigma} & \text{if } j \in f_i^n \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

μ_{ij} is the heuristic desirability for selecting feature j when at feature i in the ACO feature selection process, where k is the number of ants. δ_{ij} is an optional but frequently important component for ensuring excellent algorithm performance. Neighbor nodes of node i that ant n have not yet visited are indicated by the set f_i^n . The proportional relevance of the heuristic information and pheromone value is determined by the parameters β , ϑ and σ , which are all greater than 0 ($\vartheta > 0, \sigma > 0$), the amount of virtual pheromone on edge (i, j) is indicated by $\mu_{ij}(t)$, and the selection of ϑ and σ is experimental.

The first step of the method is to create k ants, each of which starts from a different random feature and is positioned randomly on the graph. An alternative is to start path creation at distinct features by setting the number of ants to equal the number of features. Ants move along edges in a probabilistic manner until a stopping condition is satisfied. After evaluating the resulting subsets, the method stops and outputs the best feature subset if an optimal subset is discovered or if the predetermined number of iterations is reached. The procedure iterates,

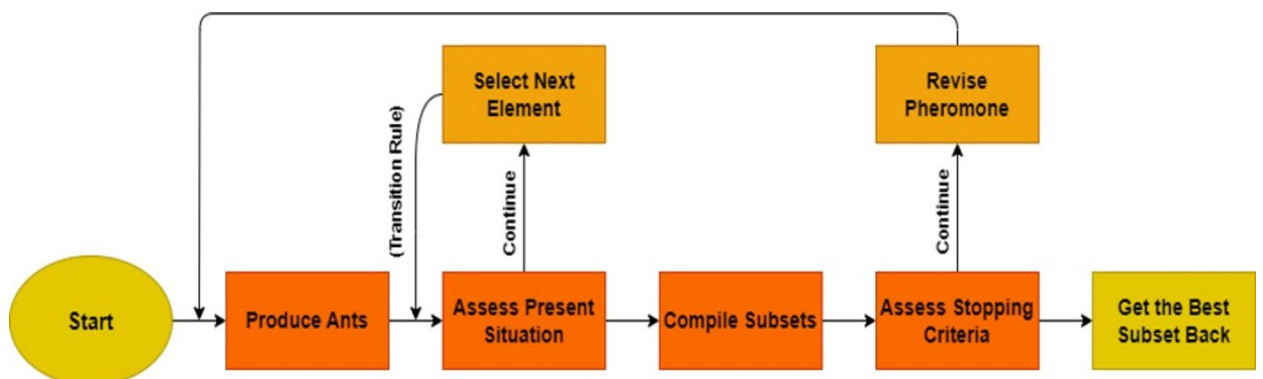


Fig. 2. Flow diagram of Ant Colony Optimization Technique.

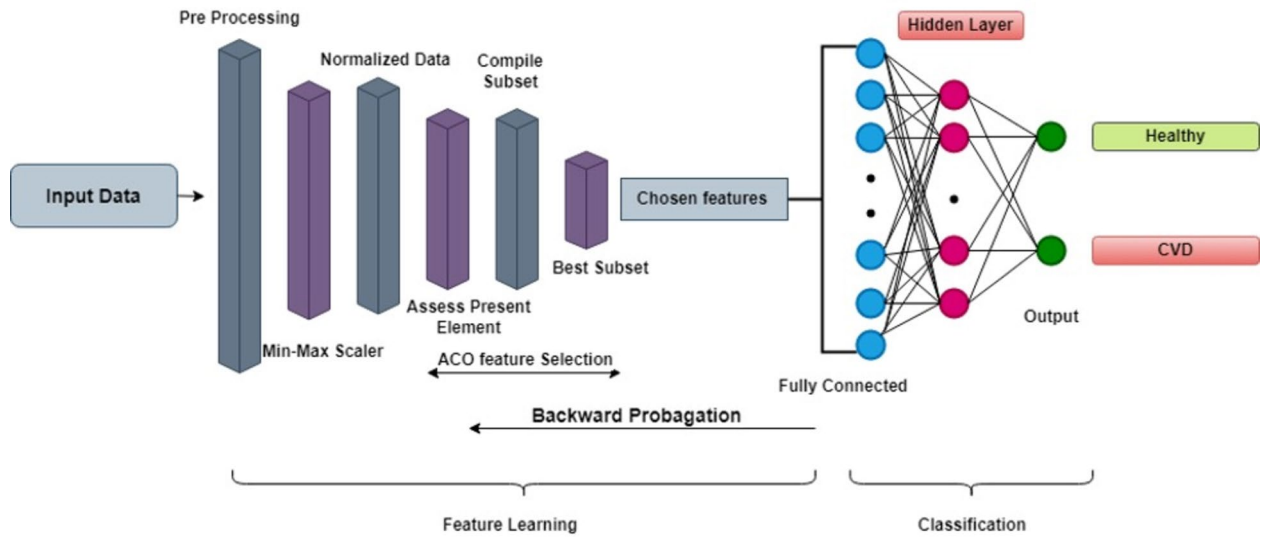


Fig. 3. Model of deep learning enhanced neural network classifier.

updating the pheromone and creating new ants if any of the two conditions is not met. Each edge’s pheromone update is determined by a certain formula:

$$\mu_{ij}(t + 1) = (1 - \emptyset) \cdot \mu_{ij}(t) + \emptyset \cdot \Delta\mu_{ij}(t) \tag{3}$$

where

$$\Delta\mu_{ij}(t) = \sum_{n=1}^m (\partial'(F^n) / |F^n|) \tag{4}$$

If the edge has been traversed, the pheromone update on edge (i, j) takes place; if not, (t) $\Delta\mu_{ij}(t)$ is 0. The rate of pheromone evaporation is represented by the decay constant, \emptyset ($0 \leq \emptyset \leq 1$). F^n represents the feature subset that ant n discovered. The “goodness” measure (∂') of the ant’s feature subset and the subset’s size both influence the pheromone update. By this concept, every ant participates in the pheromone update.

CVD classification using optimal DLENN model

The DLENN classifier is utilized in the study for the categorization of Cardiovascular Disease (CVD). Within the input area of the classification, inputs are routed to different nodes, each of which represents a unique attribute⁵⁰. Weighted averages, assigned at random, are connected to every input. The next layer is called the Hidden Layer (HL), in which each linked input node’s weight vector is added to the input product values by hidden nodes. The weighted values in DL-NN can be optimized by employing methods such as DLENN and KHOA (Kernel Harmony Search Algorithm). To improve the result of the optimizer approach, random weighted values are added, introducing another backpropagation (BP) strategy. This layer implements activation functions and propagates its results to other layers. The algorithmic phases of the DLENN approach include the following, and the weights have a significant impact on the classification result.

Step 1: Give the chosen features the score value and matching weights as follows:

$$U_i = \{U_1, U_2, U_3 \dots U_4\} \tag{5}$$

$$V_i = \{V_1, V_2, V_3 \dots V_n\} \tag{6}$$

Step 2: Add a weight vector (which can be chosen at random) to the input and then run the entire summation.

$$Y = \sum_{i=1}^n U_i V_i \tag{7}$$

here:

V_i —Weight values, Y —Total value, U_i —Value of the input entropy.

Step3: Determine the activation function by utilizing

$$F_j = f\left(\sum_{i=1}^n U_i V_i\right) \tag{8}$$

Step4: Estimate the Hidden Layers’ result as

$$P_i = b_i + \sum H_i V_i \quad (9)$$

here: H_i —values that are changed when AFs are applied, b_i —Bias value, V_i —HLs and the input are both given equal weight.

Step 5: For each layer in DLENN, follow the three previously mentioned processes. In the end, calculate the output layer neuron value by adding the weighted inputs to assess the resultant unit.

$$O_i = b_i + \sum Q_i V_i \quad (10)$$

here: V_i —Weighted of HL, O_i —output unit, Q_i —The creation of the final layer is influenced by the values of the layers.

Step 6: Compare the target value to the network effect. The incorrect data is represented by the difference between these two values. The mathematical expression for this variance is:

$$R_i = T_i - O_i \quad (11)$$

here: T_i —aimed target output, O_i —Classifier present output, R_i —Error signal.

Step 7: The goal value is now being weighted by the resultant unit. As a result, the relative error has a definition. To propagate the mistake backward to each unit in the network, a value δ_i can be determined based on this error.

$$\varphi_i = R_i [f(O_i)] \quad (12)$$

Step 8: Using the Backpropagation (BP) method, the weighted correction was calculated. The expression for this relationship is:

$$\tau v c_i = \omega \varphi_i (U_j) \quad (13)$$

here: φ_i —Error which is distributed from the network, $\tau v c_i$ —Weighted correction, U_j —Input vector, ω —Momentum term.

Finally, the Bayesian Optimization (BO) technique can be used to pick the hyperparameters optimally. BO is a successful approach to solving globally optimized issues⁵¹. Finding an input that minimizes or maximizes the cost of a particular objective function (OF) is the goal of global optimization. Calculating objective functions can be difficult because they can be higher-dimensional, noisy, non-convex, computationally expensive, and non-linear. Using an acquisition task, Bayesian Optimization (BO) effectively explores the Objective Function (OF), also known as the surrogate function, using a probabilistic approach. Bayes' Theorem is the cornerstone of the BO method. The conditional probability of event X given another event Y, or $Q(X | Y)$, is determined by applying Bayes' Theorem and can be written as follows:

$$Q(X|Y) = Q(Y|X) * \frac{Q(X)}{Q(Y)} \quad (14)$$

By eliminating the marginal likelihood $Q(Y)$ from the conditional probability calculation, Bayes' theorem is adjusted to optimize the issue. This yields the following results:

$$Q(X|Y) = Q(Y|X) * Q(X) \quad (15)$$

The marginal probability $Q(X)$ is called the prior probability, the probability is called the inverse conditional probability $Q(X | Y)$, and the posterior probability usually relates to the conditional probability $Q(X | Y)$. Consequently, the following represents an improved form of Bayes' theorem:

$$post = p_b * prior \quad (16)$$

Here, p_b —probability;

The posterior probability is used to direct subsequent sampling in the search space and functions as a function that estimates the Objective Function (OF). Here, the hyperparameters of Convolutional Neural Networks (CNNs) are the search space. Utilizing acquisition functions to model the search space and select the subsequent instance for OF estimation, Bayesian Optimization (BO) is a learning-based search methodology. First, an OF represented by a surrogate function is subjected to a probabilistic procedure in the optimization technique. In this investigation, a Gaussian process system with a Matern 5/2 kernel is used to describe the OF.

The estimate of the Objective Function (OF), represented as $F(x)$, at random instances in the search space (x_1, x_2, \dots, x_n) can be found using those instances. A set of data points, $M = \{x_1, F(x_1), \dots, x_n, F(x_n)\}$, can be formed by sequentially gathering the instances and their accompanying calculations. The number of instances is indicated by n . The previous and posterior functions are derived using this set, represented by the letter M . Equation (16) expresses the posterior function, which is defined as the likelihood of observing the data given the Objective Function (OF).

$$Q(I|J) = Q(J|I) * Q(I) \quad (17)$$

The posterior is then updated once the probability and prior are calculated. Next, as stated in Eq. (17), the acquisition function, represented as L , is optimized on the Gaussian process surrogate function to choose the subsequent instance x_n .

$$x_n = ax_i L(x) J_{1:n-1} \quad (18)$$

In this instance, the Expected Improvement approach was used to complete the acquisition task:

$$L(x) = \varepsilon [\max (f(x) - Q(x^a), 0)] \quad (19)$$

where x^a shows the location of the optimal instance in the search space, $L(x)$ indicates the Objective Function (OF) value for an optimum instance, and ε stands for the expectation operator. The cycle is then repeated until the OF achieves its lowest or the smallest objective identified within the specified runtime, after which the chosen instance is evaluated using the OF. As is customarily mentioned, the Objective Function (OF) has been evaluated a maximum of 30 in this case; each evaluation is used in a single optimized iteration. After reaching a preset threshold, the observed objective is subject to a termination criterion.

Discussion

The ICVD-ACOEDL model has exhibited remarkable performance in classifying Cardiovascular Disease (CVD) over multiple epochs, as evidenced by its accuracy, precision, recall, F-score, and G-measure metrics. This suggests that the model is resilient and can learn from the dataset. The model exhibited consistent superiority over other well-established techniques like SVM, REPTree, ANN, and bagging. This demonstrated the efficacy of combining Bayesian optimization for hyperparameter tuning and Ant Colony Optimization for feature selection. The model performed best in later epochs, indicating that longer training improves the model's diagnostic accuracy. This is an important finding. This ongoing improvement highlights the model's potential for use in clinical settings, where accuracy and dependability are crucial. Nevertheless, in light of these encouraging findings, additional study is necessary to confirm the model's efficacy across a range of demographic and geographic populations and to integrate in clinical trials.

Performance validation

This section tests the Cardiovascular Disease (CVD) classification output of the ICVD-ACOEDL algorithm using a dataset from the Kaggle repository²⁰. There are 629 disease-affected samples and 561 normal samples in the collection. The information about the dataset is given in Table 3. It contains features such as age, sex, chest pain type, resting blood pressure, cholesterol level, fasting blood sugar, resting electrocardiographic results, maximum heart rate, exercise induced angina, ST depression induced by exercise, slope of the peak exercise, number of major vessels, thalium stress result, target. From an array of 14 available features, the Ant Colony Optimisation (ACO) algorithm selected the following features: heart rate, cholesterol, ECG, pain type, and resting blood pressure.

At Epoch 500, the ICVD-ACOEDL approach categorised 602 samples as belonging to the disease class and 522 samples as normal. Additionally, 615 samples were classified as disease class and 550 samples as normal at Epoch 1000 by the ICVD-ACOEDL technique. In Epoch 1500, 616 samples were classified as disease class and 490 samples as normal by the ICVD-ACOEDL approach. In the end, at Epoch 3000, the ACVD-HBOMDL algorithm classified 555 samples as normal and 628 samples as disease class. Table 2 presents detailed findings for the classification of cardiovascular disease (CVD) over several eras using the ICVD-ACOEDL approach.

Figure 4a provides a detailed representation of the classifier findings for the first 400 epochs of the ICVD-ACOEDL technique. The results demonstrate how well the ICVD-ACOEDL method distinguishes between healthy and diseased materials. In normal class, the ICVD-ACOEDL technique has attained a_b of 94.01%, p_n of 97.12%, r_l of 94.11%, f_s of 94.79%, and g_m of 94.06%. Concurrently, on disease class, the ICVD-ACOEDL method has obtained a_b of 97.04%, p_n of 94.93%, r_l of 97.01%, f_s of 97.03%, and g_m of 95.32%.

Figure 4b displays the classifier findings for the first 800 epochs of the ICVD-ACOEDL method. The results demonstrate that the ICVD-ACOEDL method can accurately classify both healthy and sick samples. In normal class, the ACVD-HBOMDL system has achieved a_b of 95.11%, p_n of 97.99%, r_l of 95.18%, f_s of 95.99%, and g_m of 95.56%. Concurrently, on disease class, the ICVD-ACOEDL method has obtained a_b of 98.14%, p_n of 95.51%, r_l of 98.12%, f_s of 97.88%, and g_m of 96.12%.

Figure 4c displays the classifier outcomes for the first 1200 epochs of the ICVD-ACOEDL method. The results demonstrate that the ICVD-ACOEDL method can accurately classify both healthy and diseased data. In normal class, the ACVD-HBOMDL system has achieved a_b of 88.11%, p_n of 97.99%, r_l of 88.18%, f_s of 93.19%, and g_m of 93.16%. Concurrently, on disease class, the ICVD-ACOEDL method has obtained a_b of 98.44%, p_n of 90.51%, r_l of 98.82%, f_s of 94.98%, and g_m of 94.12%.

Figure 4d shows the classifier results for the ICVD-ACOEDL method's first 1600 epochs. The results demonstrate that the ICVD-ACOEDL technique can accurately classify healthy and sick data. In normal class, the ACVD-HBOMDL system has achieved a_b of 95.51%, p_n of 96.01%, r_l of 95.38%, f_s of 95.79%, and g_m of 95.56%. Concurrently, on disease class, the ICVD-ACOEDL method has obtained a_b of 96.14%, p_n of 95.81%, r_l of 98.12%, f_s of 96.88%, and g_m of 96.42%.

Class	No. of instances
Healthy	561
Diseased	629
Total instances	1190

Table 3. Dataset information.

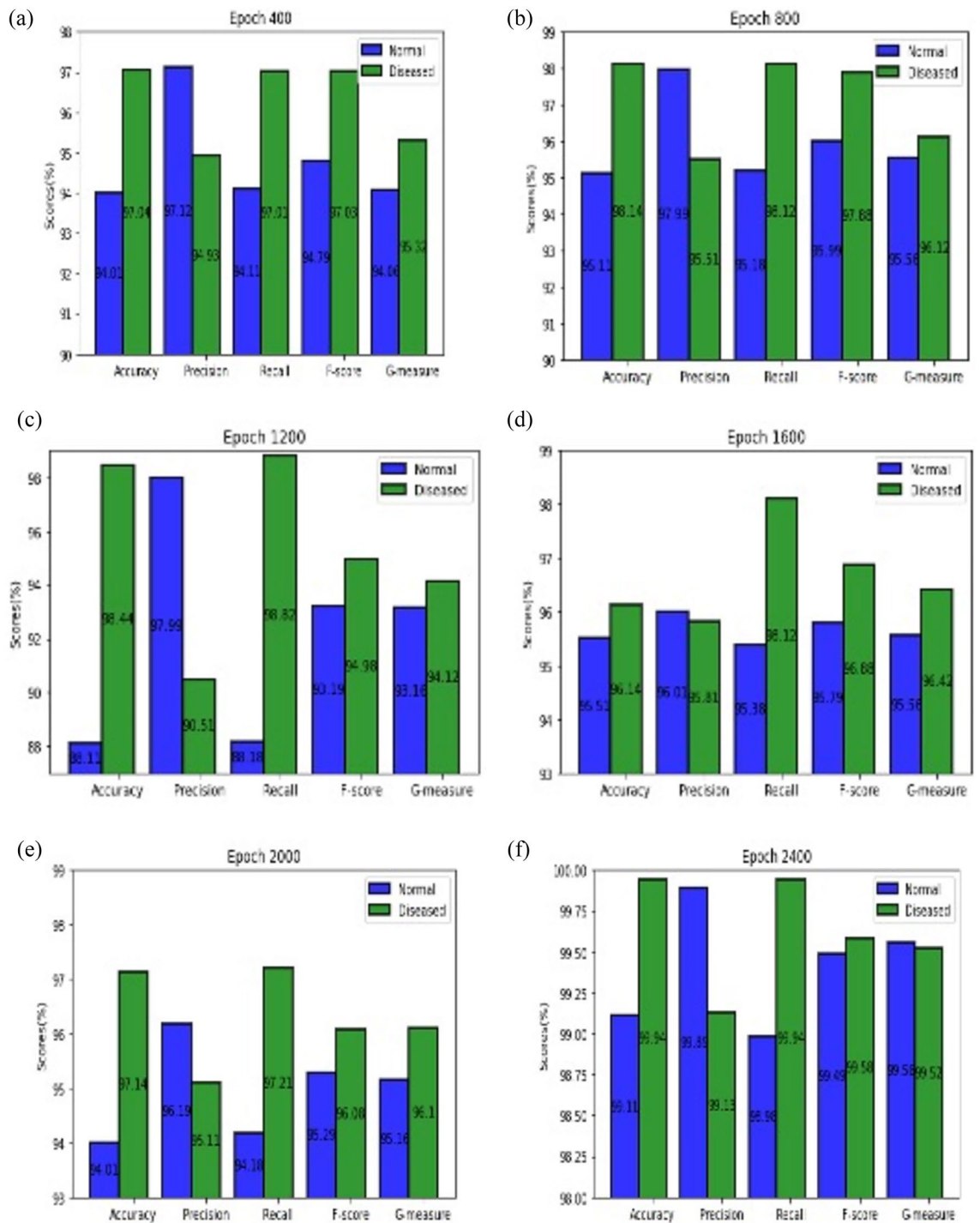


Fig. 4. The ICVD-ACOEDL approach’s algorithm results for the first 400 epochs (a), the initial 800 epochs (b), the first 1200 epochs (c), the first 1600 epochs (d), the first 2000 epochs (e), and the first 2400 epochs (f).

Figure 4e displays the classifier findings for the ICVD-ACOEDL algorithm for the first 2000 epochs. The results demonstrate that the ICVD-ACOEDL method can accurately classify both healthy and sick samples. In normal class, the ACVD-HBOMDL system has achieved a_b of 94.01%, p_n of 96.19%, r_l of 94.18%, f_s of 95.29%, and g_m of 95.16%. Concurrently, on disease class, the ICVD-ACOEDL method has obtained a_b of 97.14%, p_n of 95.11%, r_l of 97.21%, f_s of 96.08%, and g_m of 96.10%.

Figure 4f displays the classifier performance for the first 2400 epochs of the ICVD-ACOEDL method. The results demonstrate that the ICVD-ACOEDL technique can accurately classify healthy and sick data. In normal class, the ACVD-HBOMDL system has achieved a_b of 99.11%, p_n of 99.89%, r_l of 98.98%, f_s of 99.49%, and g_m of 99.56%. Concurrently, on disease class, the ICVD-ACOEDL method has obtained a_b of 99.94%, p_n of 99.13%, r_l of 99.94%, f_s of 99.58%, and g_m of 99.52%.

The ICVD-ACOEDL technique’s TACC (Training Accuracy) and VACC (Validation Accuracy) evaluations are quantified. The results of the study demonstrate that the ICVD-ACOEDL approach works better than the others, with higher TACC and VACC scores. Particularly noteworthy is the maximal TACC achieved by the ICVD-ACOEDL technique. Figure 11 illustrates the evaluation of TLS and VLS for the ICVD-ACOEDL technique in CVD performance. The findings indicate superior performance, with minimal values for both TLS and VLS. Notably, the ICVD-ACOEDL algorithm has yielded the least VLS outcome. Figure 5 analyzes and displays the Cholesterol Level, which is a useful metric for identifying CVD. CVD is characterized by four types of chest pain, including (i) typical angina: chest pain associated with a reduction in the heart’s blood supply, (ii) Atypical angina: pain in the chest unrelated to the heart, (iii) Pain that is not associated with the heart, usually esophageal spasms; (iv) asymptomatic pain is chest pain that does not manifest as a disease. Based on these chest pain types, Fig. 6 predicts and illustrates CVD.

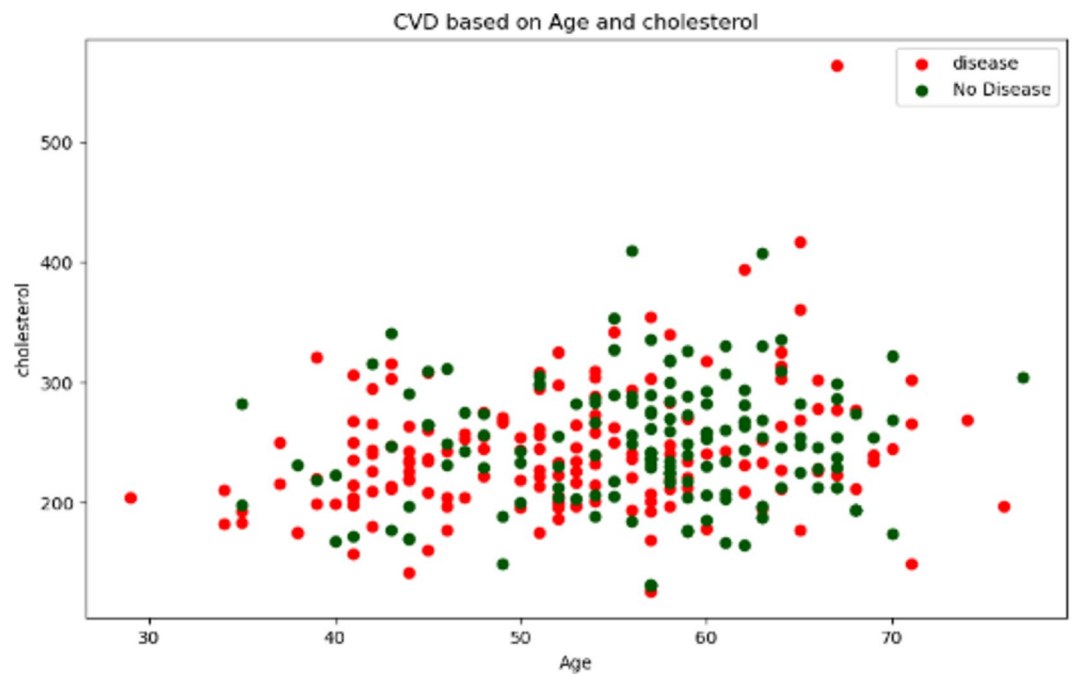


Fig. 5. CVD based on age and cholesterol level.

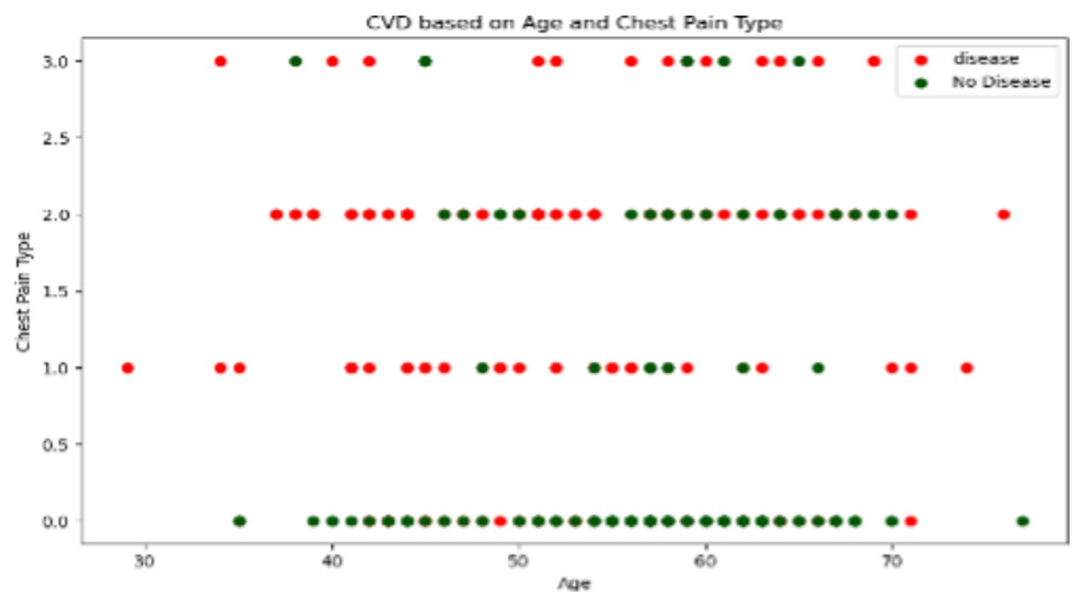


Fig. 6. CVD based on age and chest pain type.

The ICVD-ACOEDL technique’s confusion matrices under Correlation between independent variables have classified 602 samples as disease-related and 522 samples as normal. This could provide insight into which independent variables might influence our target variable and which might not. Figure 7 shows a large table of numbers that illustrates the relationship between each variable and the others. Higher positive values indicate a possible increase in the correlation, while higher negative values indicate a possible decrease in the correlation.

In Fig. 8a, the CVD classification performance of the ICVD-ACOEDL technique is compared with existing models based on a_y ^{52–54}. Results reveal that SVM and REPTree models exhibit lower accuracy at 85.19% and 93.41%, respectively. Meanwhile, ANN and bagging models achieve slightly improved a_y at 95.95% and 95.19%, respectively. Additionally, SC and J48 models achieve close a_y values at 99.16% and 98.11%, respectively. Notably, the ICVD-ACOEDL technique demonstrates superior outcomes with an increased a_y of 99.71%.

In Fig. 8b, the CVD classification results of the ICVD-ACOEDL method are compared with current techniques based on p_n . The findings show that the SVM and REPTree models achieve minimum p_n at 85.15% and 93.32%, respectively. Meanwhile, the bagging and ANN techniques achieve slightly improved p_n at 95.18% and 95.96%, respectively. Additionally, the J48 and SC techniques obtain closer p_n at 98.02% and 99.05%, respectively. Notably, the ICVD-ACOEDL algorithm displays superior outcomes with an improved p_n of 99.44%.

In Fig. 8c, a comparative analysis of CVD classification outcomes for the ICVD-ACOEDL method is presented in terms of r_l . The results indicate that the SVM and REPTree models achieve reduced r_l values of 84.18% and 93.47%, respectively. Meanwhile, the bagging and ANN techniques attain slightly improved r_l values of 94.29% and 95.15%, respectively. Additionally, the SC and J48 methods accomplish closer r_l values of 97.91% and 99.11%, respectively. Lastly, the ICVD-ACOEDL method demonstrates superior outcomes with an increased r_l of 99.71%.

In Fig. 8d, a comparative analysis of CVD classification outcomes for the ICVD-ACOEDL method is presented in terms of f_s . The results indicate that the SVM and REPTree techniques achieve reduced f_s values of 84.29% and 94.78%, respectively. Meanwhile, the bagging and ANN techniques accomplish slightly boosted f_s values of 95.14% and 96.07%, respectively. Moreover, the SC and J48 approaches achieve nearer f_s values of 98.82% and 98.97%, respectively. Finally, the ICVD-ACOEDL algorithm demonstrates superior outcomes with an increased f_s of 99.83%. These results underscore the enhanced CVD classifier performance of the ICVD-ACOEDL technique.

Conclusion

This study has created a novel ICVD-ACOEDL technique that uses feature selection (FS) and hyperparameter tuning procedures to accurately diagnose CVD. The initial step in the methodology is to use a min–max scaler to pre-process medical data. The optimal subset of attributes is found using the ACO method. For CVD classification, the DLNN classifier is employed, and the hyperparameters of the classifier optimal are chosen using Bayesian optimization. Enhancing the process of precisely identifying CVD is the aim of this integrated approach. The experimental results of the ICVD-ACOEDL technique have been validated on a benchmark

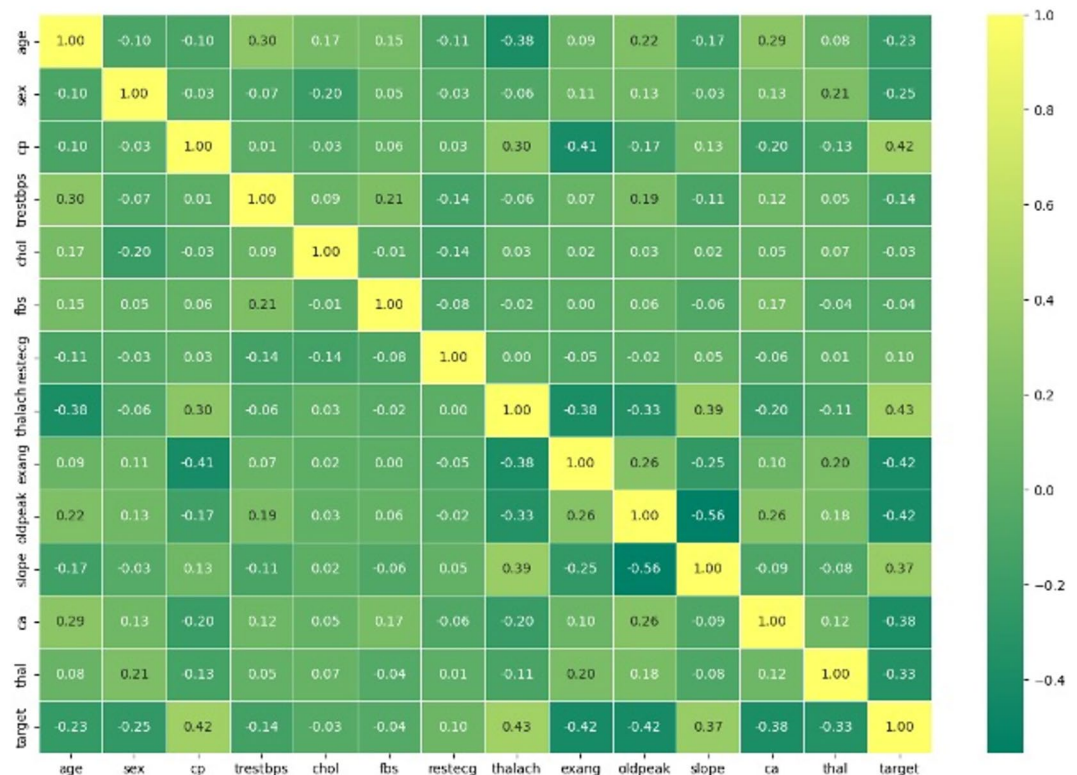


Fig. 7. ICVD-ACOEDL technique’s confusion matrices under Correlation between independent variables.

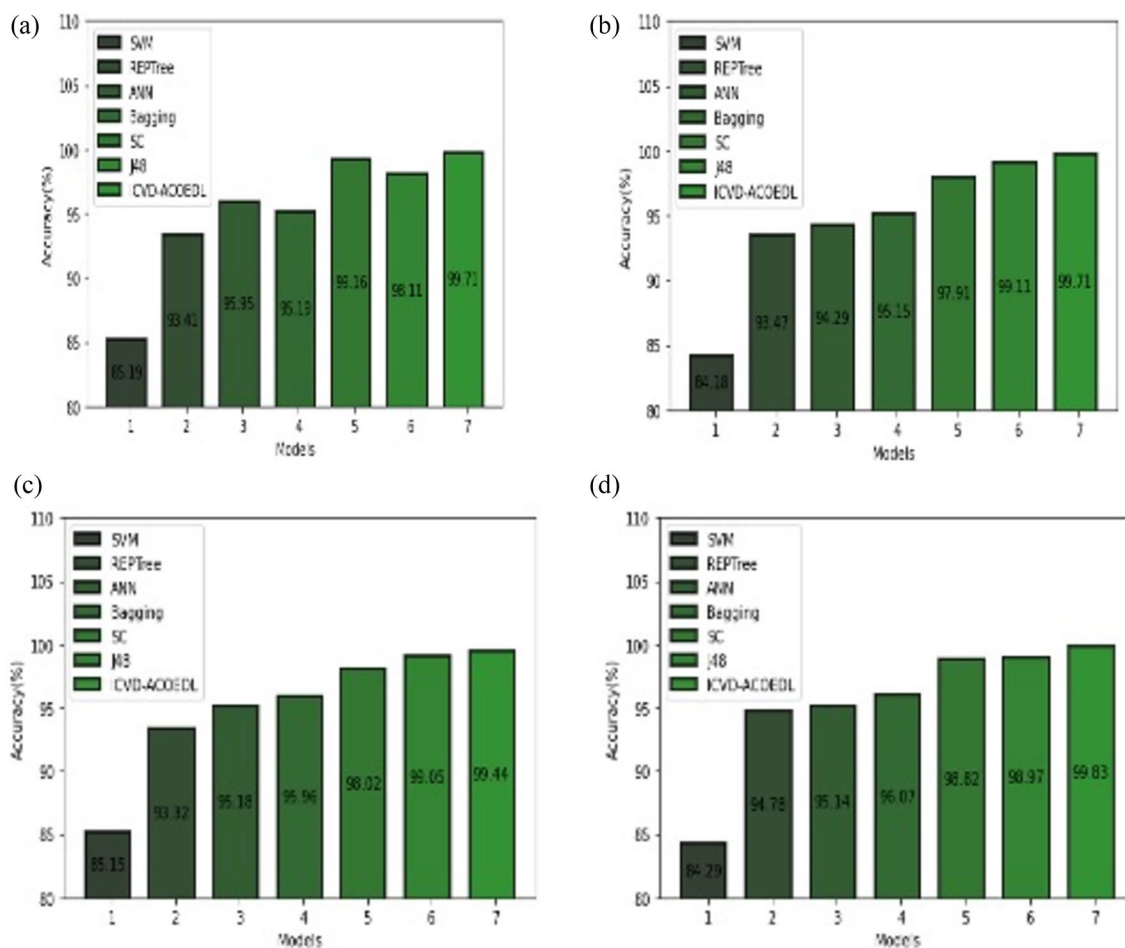


Fig. 8. Accuracy analysis of ICVD-ACOEDL approach with existing algorithms. (a) Accuracy (b) Precision (c) Recall (d) Fscore.

medical dataset, showing significant improvements over earlier techniques, with an enhanced accuracy of 99.71%. Further improvements in the efficiency of classification of the ICVD-ACOEDL method could be achieved by integrating methods for identifying outliers and data clustering.

Data availability

The datasets generated during and/or analysed during the current study are available in the [kaggle] repository, [<https://www.kaggle.com/datasets/sulianova/cardiovascular-disease-dataset>].

Received: 8 February 2024; Accepted: 2 September 2024

Published online: 18 September 2024

References

1. Weberling, L. D., Lossnitzer, D., Frey, N. & André, F. Coronary computed tomography vs. cardiac magnetic resonance imaging in the evaluation of coronary artery disease. *Diagnostics* **13**(1), 125 (2022).
2. P.Wang, Z. Lin, X.Yan, Z. Chen, M. Ding, Y. Song, and L. Meng, "Awearable ECG monitor for deep learning based real-time cardiovascular diseasedetection," 2022, [arXiv:2201.10083](https://arxiv.org/abs/2201.10083).
3. Gao, X. *et al.* Direct oral anticoagulants vs. vitamin K antagonists in atrial fibrillation patients at risk of falling: A meta-analysis. *Front. Cardiovasc. Med.* **9**, 757087 (2022).
4. Swathy, M. & Saruladha, K. 'A comparative study of classification and prediction of cardio-vascular diseases (CVD) using machine learning and deep learning techniques'. *ICT Exp.* **8**(1), 109–116 (2022).
5. Gao, X., Cai, X., Yang, Y., Zhou, Y. & Zhu, W. Diagnostic accuracy of the HAS-BLED bleeding score in VKA- or DOAC-treated patients with atrial fibrillation: A systematic review and meta-analysis. *Front. Cardiovasc. Med.* **8**, 757087 (2021).
6. Bing, P., Liu, Y., Liu, W., Zhou, J. & Zhu, L. Electrocardiogram classification using TSST-based spectrogram and ConViT. *Front. Cardiovasc. Med.* **9**, 983543 (2022).
7. Liu, D., Liu, X., Chen, Z., Zuo, Z., Tang, X., Huang, Q., Arai, T, Magnetically driven soft continuum microrobot for intravascular operations in microscale. *Cyborg Bionic Syst.*, 2022.
8. Yu, Y. *et al.* Targeting loop3 of sclerostin preserves its cardiovascular protective action and promotes bone formation. *Nat. Commun.* **13**(1), 4241 (2022).
9. Fu, Q. *et al.* Sodium intake and the risk of various types of cardiovascular diseases: A Mendelian randomization study. *Front. Nutr.* **10**, 1250509 (2023).

10. Kim, S. *et al.* Bioprinting methods for fabricating in vitro tubular blood vessel models. *Cyborg Bionic Syst.* **4**, 0043 (2023).
11. Kim, G. *et al.* A deep learning-based automatic analysis of cardiovascular borders on chest radiographs of valvular heart disease: Development/external validation. *Eur. Radiol.* **32**(3), 1558–1569 (2022).
12. Dai, Z. *et al.* Role of autophagy induced by arecoline in angiogenesis of oral submucous fibrosis. *Archiv. Oral Biol.* **102**, 7–15 (2019).
13. Malnajjar, M., Khaleel, Abu-Naser, and S. Samy. (2022). *Heart sounds analysis and classification for cardiovascular diseases diagnosis using deep learning*. [Online]. Available: <http://dspace.alazhar.edu.ps/xmlui/handle/123456789/3534>
14. Shrivastava, P. K., Sharma, M., Sharma, P. & Kumar, A. HCBiLSTM: A hybrid model for predicting heart disease using CNN and BiLSTM algorithms. *Meas. Sens.* **25**, 100657 (2023).
15. Huang, L. *et al.* Successful robot-assisted laparoscopic resection of pheochromocytoma in a patient with dilated cardiomyopathy: A case report on extremely high-risk anesthesia management. *Medicine* **102**(41), e35467 (2023).
16. Zhou, Y. *et al.* Sex-specific differences in the association between steps per day and all-cause mortality among a cohort of adult patients from the United States with congestive heart failure. *Heart & Lung* **62**, 175–179 (2023).
17. Mathur, P., Srivastava, S., Xu, X. & Mehta, J. L. Artificial intelligence, machine learning, and cardiovascular disease. *Clin. Med. Insights Cardiol.* **14**, 117954682092740 (2020).
18. Hong, S., Zhou, Y., Shang, J., Xiao, C. & Sun, J. Opportunities and challenges of deep learning methods for electrocardiogram data: A systematic review. *Comput. Biol. Med.* **122**, 103801 (2020).
19. Suganyadevi, S., Seethalakshmi, V. & Balasamy, K. 'A review on deep learning in medical image analysis'. *Int. J. Multimedia Inf. Retr.* **11**(1), 19–38 (2022).
20. Hassan, M. U., Alaliyat, S. & Hameed, I. A. Image generation models from scene graphs and layouts: A comparative analysis. *J. King Saud Univ. Comput. Inf. Sci.* **35**(5), 101543 (2023).
21. Sun, T. *et al.* In vivo liver function reserve assessments in alcoholic liver disease by scalab15e photoacoustic imaging. *Photoacoustics* **34**, 100569 (2023).
22. Yang, C., Sheng, D., Yang, B., Zheng, W., & Liu, C, A dual-domain diffusion model for sparse-view CT reconstruction. *IEEE Signal Processing Letters*, 2024.
23. Lu, S. *et al.* *Surgical instrument posture estimation and tracking based on LSTM* (ICT Express, 2024).
24. Chen, M., Hao, Y., Hwang, K., Wang, L. & Wang, L. 'Disease prediction by machine learning over big data from healthcare communities'. *IEEE Access* **5**, 8869–8879 (2017).
25. Matsushita, K. *et al.* 'The relationship of COVID-19 severity with cardiovascular disease and its traditional risk factors: A systematic review and meta-analysis'. *Glob. Heart* **15**(1), 64 (2020).
26. Siontis, K. C., Noseworthy, P. A., Attia, Z. I. & Friedman, P. A. 'Artificial intelligence-enhanced electrocardiography in cardiovascular disease management'. *Nat. Rev. Cardiol.* **18**(7), 465–478 (2021).
27. W. A. W. A. Bakar, N. L. N. B. Josdi, M. B. Man, and M. A. B. Zuhairi, A review: Heart disease prediction in machine learning & deep learning. in *Proc. 19th IEEE Int. Colloq. Signal Process. Appl. (CSPA)*, 2023, pp 150–155.
28. I. S. Brites, L. M. Silva, J. L. Barbosa, S. J. Rigo, S. D. Correia, and V. R. Leithardt, "Machine learning and IoT applied to cardiovascular diseases identification through heart sounds: A literature review," in *Proc. Int. Conf. Inf. Technol. Syst.* Cham, Switzerland: Springer, 2022, 356–388.
29. Nagavelli, U., Samanta, D. & Chakraborty, P. 'Machine learning technology-based heart disease detection models'. *J. Healthcare Eng.* **2022**, 1–9 (2022).
30. Arpaia, P. *et al.* Conceptual design of a machine learning based wearable soft sensor for non-invasive cardiovascular risk assessment. *Measurement* **169**, 108551 (2021).
31. Selvi, R. T. & Muthulakshmi, I. 'An optimal artificial neural network based big data application for heart disease diagnosis and classification model'. *J. Ambient Intell. Humaniz. Comput.* **12**(6), 6129–6139 (2021).
32. Ali, M. M. *et al.* 'Heart disease prediction using supervised machine learning algorithms: Performance analysis and comparison'. *Comput. Biol. Med.* **136**, 104672 (2021).
33. M. Ganesan and N. Sivakumar, "IoT based heart disease prediction and diagnosis model for healthcare using machine learning models," in *Proc. IEEE Int. Conf. Syst., Comput., Autom. Netw. (ICSCAN)*, 2019, pp 1–5.
34. Li, J. P. *et al.* 'Heart disease identification method using machine learning classification in e-healthcare'. *IEEE Access* **8**, 107562–107582 (2020).
35. R. Atallah and A. Al-Mousa, "Heart disease detection using machine learning majority voting ensemble method," in *Proc. 2nd Int. Conf. New Trends Comput. Sci. (ICTCS)*, 2019, pp 1–6.
36. M. Noale, F. Limongi, and S. Maggi, Epidemiology of cardiovascular diseases in the elderly, *Frailty and Cardiovascular Diseases: Research into an Elderly Population*. 2020, pp 29–38.
37. M. Athanasiou, K. Sfrintzeri, K. Zarkogianni, A. C. Thanopoulou, and K. S. Nikita, An explainable XGBoost-based approach towards assessing the risk of cardiovascular disease in patients with type 2 diabetes mellitus. In *Proc. IEEE 20th Int. Conf. Bioinf. Bioengineering (BIBE)*, 2020, pp 859–864.
38. Charlton, P. H. *et al.* 'Wearable photoplethysmography for cardiovascular monitoring'. *Proc. IEEE* **110**(3), 355–381 (2022).
39. Chieng, D. & Kistler, P. M. 'Coffee and tea on cardiovascular disease (CVD) prevention'. *Trends Cardiovasc. Med.* **32**(7), 399–405 (2022).
40. Tao, L.-C., Xu, J.-N., Wang, T.-T., Hua, F. & Li, J.-J. Triglyceride-glucose index as a marker in cardiovascular diseases: Landscape and limitations. *Cardiovasc. Diabetol.* **21**(1), 1–17 (2022).
41. Battineni, G., Sagaro, G. G., Chintalapudi, N. & Amenta, F. 'The benefits of telemedicine in personalized prevention of cardiovascular diseases (CVD): A systematic review'. *J. Personal. Med.* **11**(7), 658 (2021).
42. Bays, H. E. *et al.* Ten things to know about ten cardiovascular disease risk factors. *Amer. J. Preventive Cardiol.* **5**, 100149 (2021).
43. Dickson, V. V., Jun, J. & Melkus, G. D. 'A mixed methods study describing the self-care practices in an older working population with cardiovascular disease (CVD): Balancing work, life and health'. *Heart Lung* **50**(3), 447–454 (2021).
44. Ellis, G. K., Robinson, J. A. & Crawford, G. B. 'When symptoms of disease overlap with symptoms of depression'. *Austral. Family Phys.* **35**(8), 647–649 (2006).
45. Alhadeethy, N. F. A., Zeki, A. M. & Shah, A. 'Deep learning model for predicting and detecting overlapping symptoms of cardiovascular diseases in hospitals of UAE'. *Turkish J. Comput. Math. Educ. (TURCOMAT)* **12**(14), 5212–5224 (2021).
46. Hsu, C.-S. *et al.* Overlap of dyspepsia in patients with gastroesophageal reflux disease: Impact of clinical, metabolic, and psychosocial characteristics. *Digest. Dis. Sci.* **62**(4), 994–1001 (2017).
47. Chaddha, A., Robinson, E. A., Kline-Rogers, E., Alexandris-Souphis, T. & Rubenfire, M. 'Mental health and cardiovascular disease'. *Amer. J. Med.* **129**(11), 1145–1148 (2016).
48. Goodwin, G. M. 'Depression and associated physical diseases and symptoms. *Dialogues Clin. Neurosci.* **8**(2), 259–265 (2006).
49. Daoulah, A. *et al.* Outcomes of myocardial revascularization in diabetic patients with left main coronary artery disease: A multi-center observational study from three Gulf countries'. *Cardiovasc. Revasc. Med.* **46**, 52–61 (2023).
50. Muthu, B. *et al.* 'A framework for extractive text summarization based on deep learning modified neural network classifier'. *ACM Trans. Asian Low-Resour. Lang. Inf. Process.* **20**(3), 1–20 (2021).
51. Atteia, G., Alhussan, A. & Samee, N. 'BO-ALLCNN: Bayesian based optimized CNN for acute lymphoblastic leukemia detection in microscopic blood smear images'. *Sensors* **22**(15), 5520 (2022).

52. M. Siddhartha. Heart Disease Dataset (Comprehensive) Statlog + Cleveland + Hungary Dataset. Accessed: May 22, 2023. [Online]. Available: <https://www.kaggle.com/datasets/sid321axn/heart-statlog-clevelandhungary-final>
53. Kumar Dubey, A., Choudhary, K. & Sharma, R. 'Predicting heart disease based on influential features with machine learning'. *Intell. Autom. Soft Comput.* **30**(3), 929–943 (2021).
54. Mary, N. *et al.* 'Investigating of classification algorithms for heart disease risk prediction'. *J. Intell. Med. Healthc.* **1**(1), 11–31 (2022).

Author contributions

Biao Xia: Conceptualization, Methodology, Software Sajeela Fatima: Data curation, Supervision, Editing- Original draft preparation Nisreen Innab and Venkatachalam Kandasamy: Data curation, Supervision, Editing- Original draft preparation Ali Ahmadian: Validation, Reviewing Original draft Massimiliano Ferrara: Data curation, Supervision, Reviewing and editing- Original draft preparation.

Funding

Nisreen Innab would like to express sincere gratitude to Almaarefa University, Riyadh, Saudi Arabia, for supporting this research.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to B.X. or M.F.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024