Research Article – Vol. 03, Issue 01, No. 07

Advanced in Engineering and Intelligence Systems

Journal Web Page: https://aeis.bilijipub.com

Cardiovascular Disease Detection through Innovative Imbalanced Learning and AUC Optimization

Karthikeyan palanisamy[1], Krishnaveni Krishnasamy[2, *], Praba Venkadasamy[2]

[1] *School of CS & IT, JAIN (Deemed-to-be University), Bengaluru, India*

[2] *Department of Computer Science, Sri S.Ramasamy Naidu Memorial College, sattur, Virudhunagar, India*

Highlights

- Introduction of ImAUC-PSVM, a novel approach for cardiovascular disease (CVD) detection.
- Integration of AUC maximization into the objective function for efficient handling of imbalanced datasets.
- Theoretical analysis showcasing structural similarity with standard PSVM, ensuring efficacy in handling progressive CVD scenarios.
- Incorporation of a tailored Differential Evolution algorithm for precise navigation of hyperparameter space, enhancing model performance.

Article Info

Received: 18 February 2024
Received in revised: 27 March 2024
Accepted: 30 March 2024
Available online: 30 March 2024

Keywords

Cardiovascular disease
Support vector machine
AUC maximization learning
differential evolution
hyperparameter optimization

Abstract

Cardiovascular diseases (CVDs) are a primary global health concern, impacting the heart and blood vessels extensively. In this paper, we introduce a novel approach named Imbalanced Maximizing-Area Under the Curve (AUC) Proximal Support Vector Machine (ImAUC-PSVM), which harnesses the foundational principles of traditional PSVM for the detection of CVDs. The ImAUC-PSVM method offers several key advantages: 1) It skillfully incorporates AUC maximization directly into the objective function. This integration simplifies the model by reducing the number of parameters needing adjustment, making it particularly effective for handling imbalanced datasets through an efficient training process; 2) Theoretical analysis demonstrates that ImAUC-PSVM retains the same structural solution as standard PSVM. This similarity means it inherits PSVM's benefits, particularly in addressing progressive CVD scenarios with rapid incremental updates. Furthermore, we have incorporated a tailored Differential Evolution (DE) algorithm designed to navigate the complex hyperparameter space with finesse. The performance of this model was rigorously evaluated using comprehensive data from a medical survey conducted in 2012, which included an extensive cohort of 26,002 athletes. Critical parameters such as height, weight, age, gender, blood pressure, and resting heart rate were meticulously documented. The empirical results, benchmarked against established performance metrics, underscore the model's exceptional accuracy, solidifying its role as a reliable tool for CVD detection. This approach advances cardiovascular diagnostics and offers a scalable and adaptable solution, potentially influencing the broader landscape of healthcare analytics and patient care.

1. Introduction

Studies by the World Health Organization indicate that CVDs are the primary cause of death globally, accounting for roughly 17.9 million fatalities each year. Key factors contributing to this include smoking, being overweight, high blood pressure, and lack of physical activity. Although regular physical activity can mitigate these dangers, athletes who undergo rigorous and frequent training or participate in competitions might still face risks due to the strenuous nature of their sports. In this situation, sports medicine experts are vital, managing the well-being of athletes through collecting biomedical and personal data and performing electrocardiogram (ECG) tests. The outcomes of ECG screenings help in classifying individuals

* Corresponding Author: Krishnaveni Krishnasamy
Email: Kkrishnaveni@srnmcollege.ac.in

as either at risk or not. Athletes deemed at risk could be subject to limitations in their participation in sports and may require further health evaluations. There's a noticeable skew in the distribution of individuals in these categories, with a majority in the "N" category, indicating no risk factors, in contrast to the smaller "P" category for those with detected risks. Incorrect classification as false negatives can lead to severe, occasionally fatal outcomes. Conversely, false positives result in additional health examinations and a temporary halt in sports activities[1].

In the realm of medical diagnostics, the process of classification is crucial, facilitating the separation of binary outcomes and guiding clinical decision-making. Numerous machine learning algorithms have been integrated into diagnostic systems to improve the evaluation of data, health risk determination, and diagnostic precision[2], [3]. For example, Salam and McGrath [4] employed machine-learning methods to detect skin conditions in dermatology. Similarly, Campbell et al. [5]used a kernel-based technique to diagnose a rare illness. However, the dataset used needed to be proportionate to the illness's actual occurrence in the broader population. Wong et al. [6]utilized Bayesian networks to analyze seven years of medical records, focusing on epidemiological research with potential uses in anti-terrorism efforts. Fontaine et al. [7]explored data mining methods for examining clinical neurological disorders. Sacchi et al. [8]implemented a Naive Bayes model enhanced with resampling to predict glaucoma, adapting to the limitations of the dataset at hand. Recommendations have been made to compare different classification techniques to identify the most effective method in medical statistical analysis. However, there is still a noticeable research gap in utilizing large datasets for diseases with low occurrence rates, particularly in high-stress conditions. There are also ongoing discussions about cost-effective healthcare practices and the prudent use of medical tests.

The omnipresent dilemma of biased category proportions within machine learning frameworks profoundly influences their efficiency and broad applicability[9], [10]. This complication surfaces when the frequency of examples among various groups is disproportionately allocated, often tilting in favor of the more ubiquitous group while disregarding the scarcer ones. Strategies like arbitrary amplification/reduction sampling and the SMOTE [11] are employed to counteract this disparity. Yet, these tactics could be more complex, encompassing the potential eradication of critical data and the jeopardy of inducing the framework to become overly specialized. Despite SMOTE's widespread adoption, it fails to offer empirical assurances for exactitude, especially within kernel-dependent paradigms[12]. Consequently, differential penalty methodologies have been innovated. These strategies impose divergent error tariffs on separate groups. However, crafting a suitable penalty framework can be complex owing to the erratic nature of penalty distribution[13]. It is essential to recognize that specific indicators, such as the AUC, display a higher susceptibility to category bias than mere precision metrics. A forward-looking approach to navigating biased instruction involves the construction of classifiers to amplify the AUC[14]. This tactic handles category bias, striving for an optimal AUC, and usually demands fewer parameters than differential penalty solutions. This system holds particular merit in cardiovascular disease contexts, where the practicality and relevance of the predictive framework are of utmost importance[15]. The agility to adapt to sequential and instantaneous instruction is paramount, permitting the predictive framework to conform to novel evolutions and surfacing data trends. The PSVM emerges as a formidable classifier in this setting, as demonstrated by its efficacious implementation in various instantaneous instruction scenarios[16], [17].

PSVM shines in processing vast information clusters effortlessly, a pivotal characteristic for flexible or continuous learning spheres where data perpetually undergoes updates. PSVM's capacity for enabling quick, incremental enhancements without the need to overhaul the entire information assembly with every new contribution makes it exceptionally apt for fluid and dynamic data realms. Additionally, its sturdiness in the face of fluctuations and skill in deciphering complex data webs amplify its relevance across a wide range of rigorous and intricate educational scenarios. Such flexibility and computational dexterity designate PSVM as the go-to solution for scenarios necessitating swift and meticulous evaluative decisions. Furthermore, PSVM's structure facilitates seamless integration with varied optimization tactics, enhancing flexibility and operational effectiveness in numerous real-world implementations.

A significant hurdle for machine learning models is their dependence on fine-tuning hyperparameters. Various approaches have been investigated for selecting hyperparameters, including genetic and grid search algorithms[18], [19]. Grid search proves to be a proper technique when working with a small set of parameters and their possible values. Nonetheless, its inherent constraints limit its applicability in more sophisticated machine-learning scenarios. Conversely, genetic algorithms demonstrate proficiency in managing extensive and intricate sets of parameters[20]. Yet, current versions of genetic algorithms for hyperparameter tuning in machine

learning necessitate that hyperparameter be intrinsically discrete or converted into discrete forms before optimization[21]. In this regard, DE [22], [23] emerges as a potent alternative to genetic algorithms. DE is a population-based, evolutionary algorithm tailored explicitly for optimizing search realms encompassing discrete and continuous components. DE adeptly merges the population-based framework characteristic of genetic algorithms with the adaptive mutation mechanisms typical in evolutionary strategies[24], [25].

We introduce a new strategy, named ImAUC-PSVM, for CVD to address skewed class distribution and fine-tuning of hyperparameters. This method builds on the core advantages of the classic PSVM for tasks involving sentiment analysis, incorporating AUC to improve parameter adjustment and solve class imbalances. Conceptually, ImAUC-PSVM maintains the structural essence of PSVM, ensuring rapid adaptability to evolving challenges in CVD. Furthermore, the DE algorithm is utilized to optimize the hyperparameters of our newly developed model. We evaluate the effectiveness of our model using data from a comprehensive medical survey conducted in 2012, which included a participant pool of 26,002 athletes.

The key contributions of this research are:

- A novel educational approach centered on PSVM has been developed. This strategy distinctively integrates the AUC evaluation measure into its core objective to address class disparity. By embedding the AUC measure directly into the educational algorithm, this methodology offers a sophisticated and effective resolution to the widespread problem of category imbalance in facet-term extraction endeavors.
- This study employs the DE algorithm to optimize hyperparameters in the model. This integration is crucial in enhancing the model's performance and ability to adapt to diverse datasets.

The organization of this document is as follows: Section 2 presents a review of relevant literature; Section 3 outlines the developed method for detecting CVD; Section 4 elaborates on the experimental results; and Section 5 wraps up the study with a summary of the main findings.

2. Related work

The significance of data mining in numerous healthcare sectors is increasingly prominent, including tasks such as segmenting health-related images[26], detailed analysis of patient histories[27], and crafting diagnostic approaches for ailments like hepatic cancer [28] and Interstitial Lung Diseases (ILD)[29]. Its key strength lies in the processing and examining complex clinical data from the real world, which is vital for precise disease detection and prognosis. Analyzing Electronic Health Record (EHR) data in clinical informatics is critical, leveraging the capabilities of artificial intelligence and statistical methodologies. Particularly in cardiology, the impact of machine learning is notable. It is utilized for CVD forecasting [30] and categorizing blood pressure using the K-nearest neighbor (KNN) method[31]. SVM classifiers have gained traction, notably in forecasting coronary artery conditions [32] and identifying anomalies in heart valves via cardiac acoustics analysis[33]. The Naive Bayes technique is another efficient approach to predicting cardiac issues. The study by Pattekari and Parveen [34] demonstrates its efficiency when integrated with other data mining techniques. Shah and colleagues [35] enhanced the Naive Bayes approach for cardiac ailment detection by adding pattern recognition[36], though it may face challenges with datasets lacking features. Learning vector quantization (LVQ) is also explored. Chen et al. [37]created a cardiac disorder prediction system using LVQ, attaining an 80% accuracy rate in ROC curve assessments. The fusion of text and data mining is anticipated to propel further advancements in cardiac disease prognosis. Esfahani and Ghazanfari [38] validated the efficacy of a multi-classifier framework in CVD forecasting, benefiting from the diverse characteristics of training data and minimizing training duration. Bashir et al. [39]implemented machine learning classifiers for early detection of cardiac diseases, observing that precise detection correlates with the absence of disease. At the same time, accuracy pertains to correctly identifying high-risk patients. Additionally, Bashir et al. [40]introduced a composite scoring technique for diagnosing cardiovascular diseases, attaining an elevated average effectiveness (83%) across four standard datasets from the UCI repository, outperforming other systems and classifiers.

Incorporating deep learning into CVD research has significantly expanded the potential for accurate predictions and diagnostic processes. This advanced method is particularly beneficial in dealing with the complexities and variability found in real-world healthcare datasets. In this regard, Mohan et al. [41]demonstrated the power of deep learning by creating a novel multi-task deep and wide neural network (MT-DWNN) designed to predict severe events during hospitalization. This model was extensively evaluated using a large dataset spanning 18 years, which included 35,101 heart failure admissions and 2,478 renal failure cases at the Chinese PLA General Hospital. The MT-DWNN showed exceptional

effectiveness, especially in predicting renal issues in patients with heart failure, highlighting its proficiency in handling complex medical data and producing reliable predictions. Arslan and Karhan [42] made a notable contribution in this field by developing two sophisticated deep neural networks, each tailored for assessing risks associated with coronary heart disease. They addressed data inconsistency in real-world datasets by implementing a unique method for assembling training data. This approach divided the original dataset into segments, each characterized by broad and skewed distributions. This division was achieved using variational autoencoders, an artificial neural network that supports unsupervised learning of intricate data distributions. The next step in their process involved training separate classifiers on these distinct segments. This strategy aimed to enhance the accuracy of their predictive models. Using different classifiers for varied data segments enabled a more detailed understanding of the dataset's diverse characteristics. This data processing method led to a more robust and precise prediction system, which is crucial in the case of coronary heart disease, where early and accurate detection is critical to effective treatment and management.

Despite the progress made in CVD research by the current methodologies, they frequently encounter issues related to skewed class distribution and sensitivity to hyperparameter configurations, which hampers their effectiveness in CVD-focused applications. Recognizing these obstacles, our research introduces an innovative approach that combines the robustness of PSVM with the adaptive attributes of the DE algorithm, tailored for hyperparameter optimization. Our objective is to contribute a sophisticated and reliable tool to the arsenal of CVD solutions that is not only conceptually novel but also practically effective in various real-world situations.

3. The proposed method

This research introduces an innovative method for addressing heart-related conditions by focusing on direct improvement of the AUC, meticulously designed to overcome the issue of uneven class representation. Our intrigue with PSVM led to creating an AUC optimization method influenced by the PSVM structure, distinguished by three primary benefits. Firstly, in contrast to conventional SVMs that necessitate addressing a quadratic programming (QP) dilemma, hence incurring significant processing demands, PSVM introduces a solidly convex optimization puzzle with a simple, direct solution. This results in lower processing requirements while still delivering impressive precision in forecasts. Furthermore, prior studies [43], [44] have illustrated PSVM's capability to assimilate fresh data within a continuous feed framework. By circumventing the arduous task of repeated comprehensive matrix recalculations, PSVM enables effective and efficient adaptation within variable environments. Lastly, PSVM's configuration naturally eases the delineation of its association with the AUC index, particularly when the interim loss function effectively mirrors it, thus fostering the advancement of approaches dedicated to AUC enhancement.

The AUC is broadly acknowledged as a more effective metric than accuracy for evaluating performance on imbalanced datasets[45]. This is attributed to AUC's resilience to imbalanced class distributions, rendering it a more dependable metric. From a probabilistic perspective, AUC(s) symbolizes the mean probability that a randomly chosen positive instance (x+) from the dominant class and a negative instance (x−) from the less represented class will be correctly ordered by the classifier, with x+ receiving a higher score than x−. This concept captures the mathematical expectation, shedding light on the classifier's average capability to differentiate between the two classes. The formula for AUC(s) is [46]:

$$AUC(s) = E_{\boldsymbol{x}^+ \sim D^+} E_{\boldsymbol{x}^- \sim D^-} \mathbf{1}(\boldsymbol{s}(\boldsymbol{x}^+) > \boldsymbol{s}(\boldsymbol{x}^-)) \quad (1)$$

Here, D^+ signifies the distribution for the majority class and D^- for the minority class. The term $s(\cdot)$ is the scoring function used by the classifier. For example, a standard formulation for $s(\boldsymbol{x})$ could be $s(\boldsymbol{x}) = \binom{\varphi(\boldsymbol{x})}{1}\mathrm{w}'$, where $\varphi(\boldsymbol{x})$ is a feature transformation and w' is the weight vector of the classifier.

E signifies the statistical mean, $1(\cdot)$ represents the characteristic function, assigning a value of 1 when its criterion is satisfied, and 0 otherwise. Owing to the abrupt nature of the characteristic function $1(\cdot)$, it is replaced with a smooth and loss $(1 - \acute{\boldsymbol{w}}^T(\binom{\varphi(\boldsymbol{x}^+)}{1} - \binom{\varphi(\boldsymbol{x}^-)}{1}))^2$. This modification facilitates the formulation of an experiential variant, labeled as $R_{AUC}(s)$, for $AUC(\sigma)$ employing the chosen scoring mechanism [47].

$$\begin{aligned} R_{AUC}(s) &= \frac{1}{N^+}\sum_{i=1}^{N^+}\left(\acute{\boldsymbol{w}}^T\left(\boldsymbol{c}^+ - \binom{\varphi(\boldsymbol{x}_i^+)}{1}\right)\right)^2 + \frac{1}{N^-}\sum_{j=1}^{N^-}\left(\acute{\boldsymbol{w}}^T\left(\binom{\varphi(\boldsymbol{x}_j^-)}{1} - \boldsymbol{c}^-\right)\right)^2 \\ &+ \frac{1}{2N^+}\sum_{i=1}^{N^+}\left(\acute{\boldsymbol{w}}^T\left(1 - (\binom{\varphi(\boldsymbol{x}_i^+)}{1} - \boldsymbol{c}^-)\right)\right)^2 \\ &+ \frac{1}{2N^-}\sum_{j=1}^{N^-}\left(\acute{\boldsymbol{w}}^T\left(1 - (\boldsymbol{c}^+ - \binom{\varphi(\boldsymbol{x}_j^-)}{1})\right)\right)^2 \end{aligned} \quad (2)$$

$$\boldsymbol{c}^+ = E_{\boldsymbol{x}^+ \sim D^+}\begin{pmatrix}\varphi(\boldsymbol{x}^+)\\ 1\end{pmatrix} \quad (3)$$

$$\boldsymbol{c}^- = E_{\boldsymbol{x}^- \sim D^-}\begin{pmatrix}\varphi(\boldsymbol{x}^-)\\ 1\end{pmatrix} \quad (4)$$

$$N = N^+ + N^- \quad (5)$$

Utilizing the earlier mentioned surrogate loss function converts the goal of maximizing AUC(s) into a corresponding objective of minimizing RAUC(s)[48]. Similarly, in references [49] and[50], where AUC(s) is included in their objective functions, we seamlessly incorporate RAUC in our accurate function formulation, leading to the following equation:

$$\min_{\acute{w}} \frac{1}{2}\acute{w}^T\acute{w} + \gamma R_{AUC}(s) \tag{6}$$

In this context, γ signifies the selected regularization parameter. By merging Equation 2 with Equation 6, we establish the optimization task for our predictive model as outlined below:

$$\begin{aligned}\min_{\acute{w}} \frac{1}{2}\acute{w}^T\Bigg(& I \\ & + \frac{2\gamma}{N^+}\sum_{i=1}^{N^+}\left(c^+ - \begin{pmatrix}\varphi(x_i^+)\\ 1\end{pmatrix}\right)\left(c^+ - \begin{pmatrix}\varphi(x_i^+)\\ 1\end{pmatrix}\right)^T \\ & + \left(\frac{2\gamma}{N^-}\sum_{j=1}^{N^-}\left(\begin{pmatrix}\varphi(x_j^+)\\ 1\end{pmatrix} - c^-\right)\left(\begin{pmatrix}\varphi(x_j^+)\\ 1\end{pmatrix} - c^-\right)^T\right)\acute{w} \\ & + \frac{\gamma}{2N^+}\sum_{i=1}^{N^+}(1 - \acute{w}^T(\begin{pmatrix}\varphi(x_i^+)\\ 1\end{pmatrix} - c^-)))^2 \\ & + \frac{\gamma}{2N^-}\sum_{j=1}^{N^-}(1 - \acute{w}^T(c^+ - \begin{pmatrix}\varphi(x_j^-)\\ 1\end{pmatrix}))^2\end{aligned} \tag{7}$$

With the following notation[51]:

$$\begin{aligned}A &= I + \frac{2\gamma}{N^-}\sum_{j=1}^{N^-}\left(\begin{pmatrix}\varphi(x_j^-)\\ 1\end{pmatrix} - c^-\right)\left(\begin{pmatrix}\varphi(x_j^-)\\ 1\end{pmatrix} - c^-\right)^T \\ &+ \frac{2\gamma}{N^+}\sum_{i=1}^{N^+}\left(c^+ - \begin{pmatrix}\varphi(x_i^+)\\ 1\end{pmatrix}\right)\left(c^+ - \begin{pmatrix}\varphi(x_i^+)\\ 1\end{pmatrix}\right)^T\end{aligned} \tag{8}$$

And the mathematical transformation[52]:

$$\begin{aligned}\varphi(x)^T\acute{\varphi}(x_i) &= \begin{cases}\sqrt{\frac{1}{2p}}\begin{pmatrix}\varphi(x)\\ 1\end{pmatrix}^T\left(\begin{pmatrix}\varphi(x_i)\\ 1\end{pmatrix} - \begin{pmatrix}\frac{\varphi(x_1^-) + \cdots + \varphi(x_{N^-}^-)}{N^-}\\ 1\end{pmatrix}\right) & .y_i = +1\\ \sqrt{\frac{1}{2\,(1-p)}}\begin{pmatrix}\varphi(x)\\ 1\end{pmatrix}^T\left(\begin{pmatrix}\varphi(x_i)\\ 1\end{pmatrix} - \begin{pmatrix}\frac{\varphi(x_1^+) + \cdots + \varphi(x_{N^-}^+)}{N^+}\\ 1\end{pmatrix}\right) & .y_i = -1\end{cases} \\ &= \begin{cases}\sqrt{\frac{1}{2p}}\left(K(x.x_i) - \frac{K(x.x_1^-) + \cdots + K(x.x_N^-)}{N^-}\right) & .y_i = +1\\ \sqrt{\frac{1}{2\,(1-p)}}\left(K(x.x_i) - \frac{K(x.x_1^+) + \cdots + K(x.x_N^+)}{N^+}\right) & .y_i = -1\end{cases}\end{aligned} \tag{16}$$

Within our schema, the variable x is processed by the PSVM decision-maker. The verdict of the PSVM decision-

$$\begin{aligned}&\left(\begin{pmatrix}\acute{\varphi}(x_i)\\ 0\end{pmatrix}.\acute{y}_i\right) \\ &= \begin{cases}\left(\sqrt{\frac{1}{2p}}\left(\begin{pmatrix}\varphi(x_i)\\ 1\end{pmatrix} - c^-\right).\sqrt{\frac{1}{2p}}\right) & .y_i = +1\\ \left(\sqrt{\frac{1}{2\,(1-p)}}\left(\begin{pmatrix}\varphi(x_i)\\ 1\end{pmatrix} - c^+\right).-\sqrt{\frac{1}{2\,(1-p)}}\right) & .y_i = -1\end{cases}\end{aligned} \tag{9}$$

$$p = \frac{N^+}{N^+ + N^-} \tag{10}$$

Where Equation 7 can then be expressed as:

$$\min_{\acute{w}} \frac{1}{2}\acute{w}^T A\,\acute{w} + \frac{\gamma}{2N}\sum_{i=1}^{N}(\acute{y}_i - \acute{w}^T\begin{pmatrix}\acute{\varphi}(x_i)\\ 0\end{pmatrix})^2 \tag{11}$$

The resolution is effortlessly derived as follows:

$$\acute{w} = \acute{Z}^T(\acute{Z}\acute{Z}^T + \frac{N}{\gamma}A)^{-1}\acute{Y} \tag{12}$$

$$\acute{Z} = [(\varphi'(x_1)^T.0); (\varphi'(x_2)^T.0).\ldots.(\varphi'(x_N)^T.0)] \tag{13}$$

$$\acute{Y} = [y_1'.y_2'.\ldots.y_N'] \tag{14}$$

It's observed that both $\acute{Z}$ and A can be computed in advance. When $\varphi(x_i)$ remains elusive, we assemble the kernel array K utilizing the training set. Via the eigenvalue decomposition of K, that is, $K = (\lambda^{\frac{1}{2}}\mathbf{p})^T(\lambda^{\frac{1}{2}}\mathbf{p})$, we deduce a resolution for $\varphi(x_i)$, $(i = 1,\ldots,N)$,, and subsequently ascertain $\varphi'(x_i)$, $\acute{Z}$, and A. Hence, for any evaluation instance x, its judgment function is delineated as[53]:

$$\begin{aligned}f(x) &= \begin{pmatrix}\varphi(x)\\ 1\end{pmatrix}^T\acute{Z}^T\left(\acute{Z}\acute{Z}^T + \frac{N}{\gamma}A\right)^{-1}\acute{Y} \\ &= \begin{pmatrix}\varphi(x)\\ 1\end{pmatrix}^T\left(\begin{pmatrix}\acute{\varphi}(x_1)\\ 0\end{pmatrix}.\begin{pmatrix}\acute{\varphi}(x_2)\\ 0\end{pmatrix}.\ldots.\begin{pmatrix}\acute{\varphi}(x_N)\\ 0\end{pmatrix}\right)\left(\acute{Z}\acute{Z}^T + \frac{N}{\gamma}A\right)^{-1}\acute{Y} \\ &= (\varphi(x)^T\acute{\varphi}(x_1).\varphi(x)^T\acute{\varphi}(x_2).\ldots.\varphi(x)^T\acute{\varphi}(x_N))\left(\acute{Z}\acute{Z}^T + \frac{N}{\gamma}A\right)^{-1}\acute{Y}\end{aligned} \tag{15}$$

Each $\varphi(x)^T\acute{\varphi}(x_i)(i = 1,\ldots,N)$ is computed employing Equation 16, embracing the kernel equation[54].

maker, indicating if a competitor is susceptible or not, is adjudicated by Equation 15. The entire learning procedure of the model is comprehensively detailed in Algorithm 1.

Algorithm 1: Pseudo-code of the proposed model

//Input:

γ: a balancing hyperparameter

//Kernel Matrix Formulation and Eigenvalue Analysis:

Formulate the kernel matrix $\boldsymbol{K}$ for the training data utilizing the given kernel method

Conduct eigenvalue analysis to ascertain the solution for $\varphi(\boldsymbol{x}_i)$, where i is from 1 to N.

//Matrix formulation

Construct the matrix $\boldsymbol{A}$ utilizing Equation 8, $\varphi'(\boldsymbol{x}_i)$, where i covers from 1 to N, applying Equation 9, and $\acute{\boldsymbol{Z}}$.

//Output:

Ascertain the decision mechanism for a new sample $\boldsymbol{x}$, referred to as $\boldsymbol{f}(\boldsymbol{x})$, employing Equation 15

Table 1. The parameters used for optimizing hyperparameters.

Hyperparameter	Range of parameters	Type of value
Size of batch	16 to 256	Integer
Number of epochs	32 to 512	Integer
γ (Balance hyperparameter)	0 to 1	Continuous
Kernel type	[Linear, Polynomial, RBF, Sigmoid]	Categorical
Feature scaling method	[None, Standardization, Normalization]	Categorical
Regularization Parameter (C)	0.01 to 100	Continuous

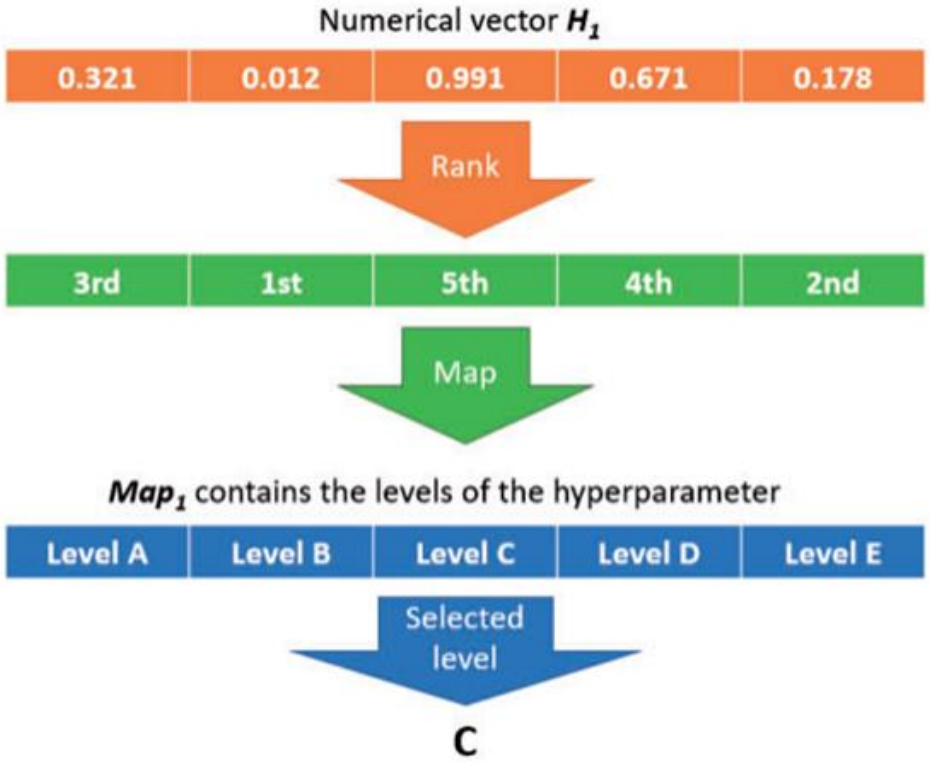

Fig. 1. Overview of applying the random critical method to convert numerical vectors into designated hyperparameter levels. This involves processes of modifying and combining these vectors. Subsequently, a mapping function transforms these vectors into a tailored set of hyperparameters, fine-tuned for model customization. This methodology facilitates evaluating and selecting numerical vectors based on their relevance and efficacy.

3.1.Hyperparameter optimization

Optimizing hyperparameters is critical in machine learning, acting as a key element that significantly improves model efficacy. Skillful adjustments of hyperparameters, including the learning rate and batch size, can enhance a model's prediction accuracy and optimize its training efficiency [55]. An optimal hyperparameter configuration addresses issues like overfitting and underfitting, promoting a robust generalization of the model to new data. Considering the intensive computational needs of deep learning models, effective hyperparameter optimization leads to more efficient use of resources, saving time and reducing costs.

Table 1 details the hyperparameters targeted for optimization in this study. We assigned each hyperparameter a plausible range of values determined by consulting ranges recommended in existing machine-learning literature related to CVDs. These predetermined ranges were then applied as limits during the execution of the DE algorithm.

Our investigation utilizes the Random Key technique, initially conceived for evolutionary computation by Bean, for fine-tuning model settings. This method relies on a coding mechanism, employing an array of T numeric arrays, each with D attributes, labeled as $p_1, p_2, \ldots, p_T$, which form a group. Every variety within this group symbolizes a prospective answer linked to a series of model settings through a stochastic fundamental translation function. The research aims at refining C settings, with C being six, as depicted in Table 1. Each setting c (from 1 to C) includes D_c spots. D_c receives the value 1 for scalar settings, rendering the aggregate attribute D the cumulative of all D_c figures. Each array, p_i, is divided into C sections, with every division holding D_c spots tied to disparate setting values. For nominal settings, this technique associates a segment of the numeric array (D_c dimensional) with a specific array, MAP_c, enumerating the alternatives for the c^{th} setting. This association is done by ordering the components in each division, where the premier component's ranking acts as a pointer to determine the appropriate value from MAP_c. The merit of this approach lies in its harmony with evolutionary procedures such as alteration, hybridization, and choice. It is applied directly to the numerical vector p_i, yielding results that can be consistently interpreted as combinations of both categorical and continuous hyperparameter values.

This method is illustrated in the context of the 'number of layers' hyperparameter (D_c=5), as shown in Figure 1. The random key comprises a series of natural numbers sorted by rank, crucial in aligning with a preset array of choices. Over time, this results in more effective decisions ascending in the key while less effective ones descend. This systematic organization aids in methodically prioritizing choices from most to least effectual, creating an orderly framework for the DE algorithm to navigate and assess.

4. Experimental results

This section outlines the data collection, elaborating on its attributes and extent as applied in our investigation. Subsequently, we delineate the benchmarks, clarifying the standards and evaluations utilized to gauge the efficacy of our frameworks. The part wraps up with the disclosure of outcomes, underscoring the principal insights from our scrutiny and framework assessments and deliberating their importance within our study aims.

4.1.Dataset

In 2012, the Zagreb Clinic for Work and Sports Health embarked on an extensive research endeavor, examining a large dataset comprising 26,002 medical examinations. These assessments were pivotal for athletes seeking clearance for participation in competitive sporting events. The gathered data encompassed vital health metrics such as sex, age, stature, body mass, resting pulse rate, arterial pressure, and baseline ECG readings. The results from these evaluations were predominantly classified into 'N' and 'P.' The 'N' category, representing 91.2% of the cases, typically reflected normal or non-critical findings, indicating that the athletes were in good health and fit for involvement in competitive sports. On the other hand, the 'P' category, constituting 8.8% of the cases, potentially indicated specific medical concerns or conditions that required further clinical assessment or intervention. This comprehensive data compilation was crucial in ensuring athletes' welfare and physical readiness. It provided valuable insights into the standard health profiles and physiological benchmarks of individuals engaged in competitive sports. The breadth and depth of this research make it a significant resource for studies in the field of sports medicine and the well-being of athletes [56].

In the dataset, individuals were identified as either being at risk, with 6,507 samples or not at risk, comprising 633 samples. We designated 70% of the total samples, amounting to 10,200, for training purposes and set aside the remainder for validation.

4.2. Metrics

Our study carefully selected essential performance metrics, including Accuracy, F-measure, and G-means, for their relevance and comprehensive evaluation capabilities in the context of imbalanced datasets like those encountered in CVD detection [57]. Accuracy, while a

standard metric, offers a straightforward initial assessment of overall model performance by calculating the proportion of correctly predicted instances among the total. However, given the imbalanced nature of our datasets, where the prevalence of one class significantly outweighs the other, relying solely on Accuracy might be misleading. This is because a model could achieve high Accuracy by predominantly predicting the majority class while failing to adequately identify the minority class instances, which are often of greater clinical significance in CVD detection.

To address this limitation and ensure a more balanced evaluation, we incorporated the F-measure, which harmonizes the Precision and Recall through their harmonic mean. This metric is particularly valuable in our study as it provides a more nuanced view of the model's ability to correctly identify positive (disease-present) cases, balancing its precision against its recall. This balance is crucial in medical diagnostics, where the cost of false negatives (failing to detect a disease) can be far more consequential than false positives [9].

Moreover, we employed the G-means metric, which evaluates model performance by considering the geometric mean of sensitivity (actual positive rate) and specificity (true negative rate). This metric is especially pertinent in imbalanced dataset scenarios, as it ensures that the model's performance is not biased towards the majority class and maintains a robust detection rate for the minority class, which, in the context of CVD detection, represents the actual CVD cases [12].

The Accuracy, F-measure, and G-means metrics are defined as follows:

- Accuracy is computed as the sum of true positives (TP) and true negatives (TN) divided by the overall number of samples, as shown in the equation:

$$\text{Accuracy} = \frac{TP + TN}{Total\ number\ of\ samples} \quad (17)$$

- F-measure provides a balance between Precision and Recall, using their harmonic mean:

$$\text{F} - \text{measure} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recal}} \quad (18)$$

Where:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (19)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (20)$$

- G-means is the geometric mean of Recall and Specificity, offering a balanced metric for imbalanced datasets:

$$\text{G} - \text{means} = \sqrt{\text{Recall} \times Specificity} \quad (21)$$

Where:

$$Specificity = \frac{TN}{TN + FP} \quad (22)$$

4.3. Baseline methods

Our innovative model, engineered for seamless operation on a high-capacity computing system with a 64-bit Windows Operating System, is equipped with substantial resources, including 32 GB of RAM and a high-performance 64 GB Graphics Processing Unit (GPU). This powerful setup is crucial for accommodating the extensive computational demands of the model.

During its initial phase of training and assessment, our model showcased exceptional efficacy, outperforming a suite of six diverse machine learning algorithms[58]:

- Support Vector Machine (SVM)[59]: SVM is a prominent technique in supervised learning, predominantly utilized for classification challenges. It operates by identifying an optimal hyperplane that effectively separates different class labels in a given dataset. The core strategy of SVM is to maximize the margin between the nearest data points of other classes, ensuring clear class distinction.
- Naïve Bayes Classifier: This algorithm is rooted in probabilistic theory, applying Bayes' theorem with the fundamental presumption of feature independence. Naïve Bayes excels in handling vast datasets and is notably effective in categorization tasks in text analysis, such as email spam filtering, due to its simplicity and speed.
- K-Nearest Neighbors (KNN)[60]: KNN is a straightforward, non-parametric method used in classification and regression. It classifies a new data point based on the predominant categories among its 'K' closest neighbors in the dataset, where 'K' is a user-defined number. The simplicity of KNN lies in its direct approach of considering the nearest data points for decision-making.
- Random Forest [61]: This ensemble learning technique is primarily used for classification and regression tasks. Random Forest constructs numerous decision trees during the training process and integrates their outputs. For classification, it takes the mode of the classes from the trees, and for regression, the average of the predictions, leading to improved accuracy and robustness against overfitting.
- Logistic Regression[62]: Contrary to what its name might imply, logistic regression is a classification algorithm, not a regression model. It's particularly adept at binary classification

problems, estimating the probability that a given input belongs to a particular class. The model utilizes a logistic function to generate outputs ranging from 0 to 1, indicative of class probabilities.

- Decision Tree[63]: This model operates similarly to a flowchart, with internal nodes representing tests on features, branches showing the results of these tests, and leaf nodes indicating class labels. Decision Trees are versatile, being applicable for both classification and regression tasks. They are intuitive, allowing for easy visualization and interpretation of the decision-making process.

Moreover, we have developed a variant of our proposed model, i.e., Proposed w/o Optimization, which ignores the DE algorithm for hyperparameter optimization.

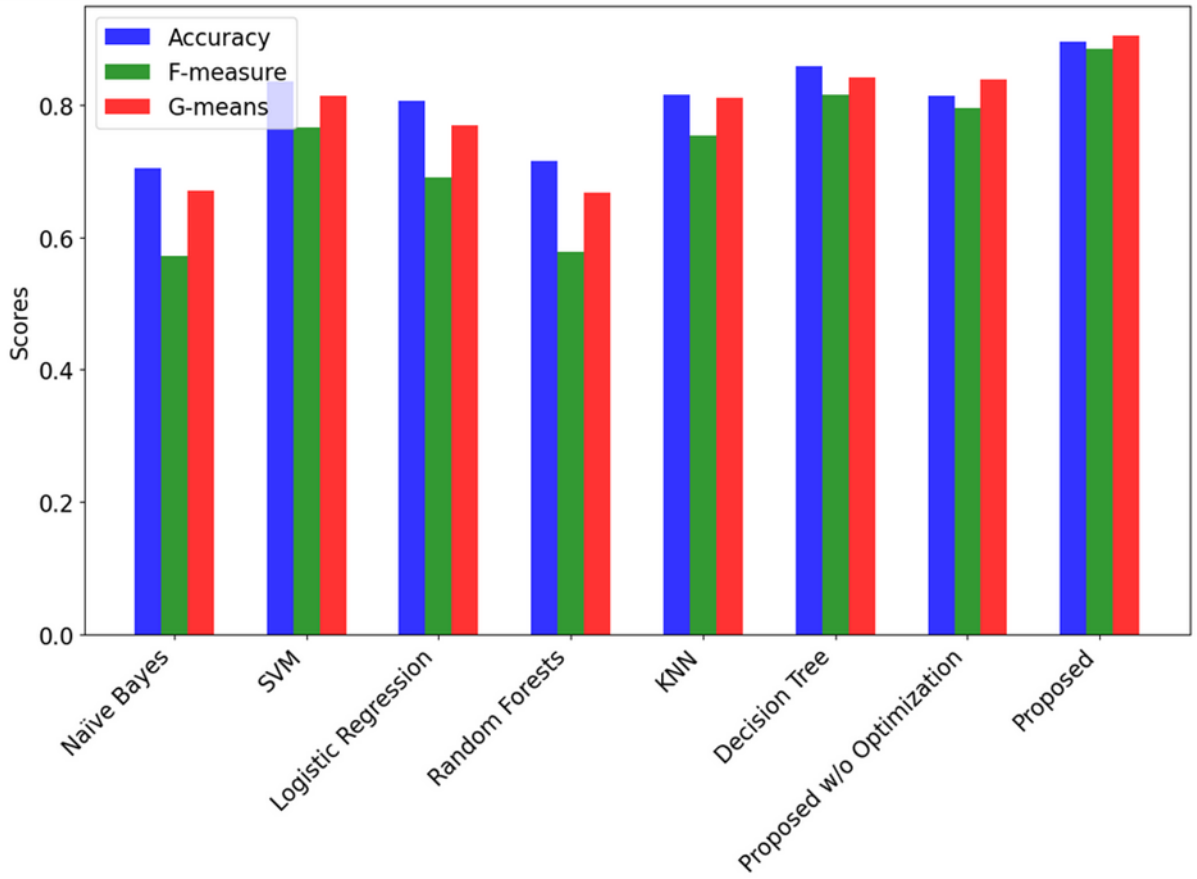

Fig. 2. Comparative evaluation of various models' performance indicators.

4.4. *Main results*

The insights gleaned from Figure 2 emphatically validate the heightened efficacy of our novel model when juxtaposed against conventional machine learning methodologies. It manifests pronounced advancements in pivotal performance indicators such as the F-measure and geometric mean. These metrics are vital for gauging a model's comprehensive effectiveness, particularly in balancing recall and precision. Notably, our model demonstrates a marked reduction in error margins, outperforming standard models by 9% in F-measure and 7% in geometric mean. This underscores its enhanced predictive accuracy and steadfast reliability under diverse scenarios.

A detailed analysis indicates that while traditional approaches like Naïve Bayes, KNN, and Logistic Regression achieve notable accuracy, they fall short in precision and recall compared to our innovative model. This aspect is especially critical in scenarios involving imbalanced datasets. For instance, Naïve Bayes, despite its accuracy of 0.830, lacks the precision exhibited by our model, potentially leading to a higher occurrence of false positives. Similarly, despite its significant accuracy, recall, and F-measure relative to other baseline models, the Decision Tree model does not reach the comprehensive proficiency demonstrated by our proposed model. Its diminished geometric mean suggests a difficulty in handling datasets with skewed distributions.

Further, when contrasting our model with the version without optimization (Proposed w/o Optimization), there is a conspicuous 15% dip in error rates in favor of the optimized version. This stark difference accentuates the critical role of the sophisticated Differential Evolution (DE) technique in elevating the model's overall performance. The implementation of DE is presumably instrumental in fine-tuning hyperparameter optimization, thereby augmenting the model's capability to discern complex patterns within the dataset. This enhancement is a testament to the value added by meticulous optimization in developing machine learning models.

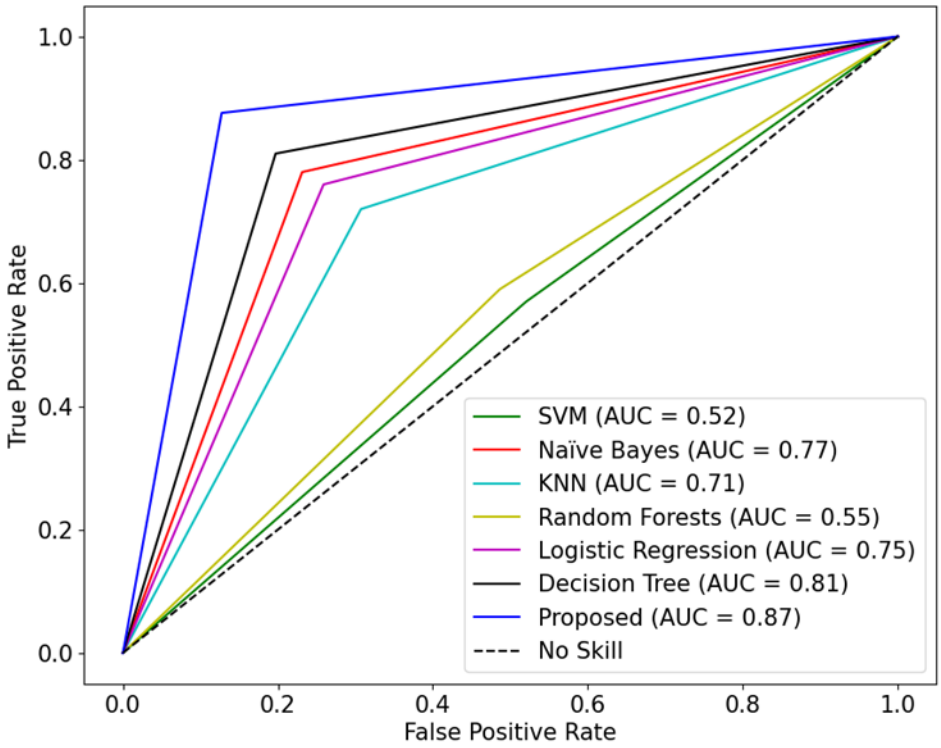

Fig. 3. The overlapping of roc curves across different machine learning techniques highlights the AUC scores, including the proposed model.

Illustrated in Figure 3 are the Receiver Operating Characteristic (ROC) curves for various machine learning techniques, as detailed in Figure 2. These curves delineate the relationship between each method's actual positive rate (TPR) or sensitivity and the false positive rate (FPR), essentially one minus specificity. The Area Under the Curve (AUC) is a pivotal measure of the classifiers' cumulative proficiency. The AUC values paint a picture of the differential predictive prowess inherent in each method. A classifier with an AUC approaching 1 indicates its elevated accuracy in distinguishing between classes, whereas an AUC hovering near 0.5 implies a performance level comparable to random guessing.

The graph highlights that our advanced model outstrips its counterparts, achieving the highest AUC and confirming its superior capability in categorizing positive and negative instances precisely. This is followed by the decision tree and logistic regression methods, which exhibit respectable AUC scores, signifying their commendable classification abilities. In contrast, methodologies like SVM and random forests manifest somewhat intermediate AUC scores. While these scores exceed the baseline threshold of 0.5, they imply a comparative deficiency in these methods' capacity to discriminate between classes, especially against the backdrop of the more productive models.

The observed variance in AUC scores can be attributed to many factors, ranging from the intrinsic characteristics of the data, the complexity and peculiarities of the individual models, to the fundamental limitations inherent in the algorithms themselves. Such differences underscore the necessity of careful model selection and customization by the specific demands and nuances of the dataset. Understanding these distinctions is crucial in machine learning, as it aids in tailoring solutions that are optimally aligned with the desired outcomes and the inherent challenges of the data.

To verify the dependability of our algorithm and avert the risk of excessively tailoring it to the training set at the expense of its ability to perform well on unseen data, Figure 4 offers a visual depiction. This illustration delineates the progression of the hinge loss values for both the training and validation sets throughout the learning phase. The

methodology entails calculating the loss after each forward propagation throughout the training phase, with a subsequent backward propagation to adjust the algorithm's parameters after every training cycle. The loss corresponding to the validation set is determined following each training cycle through forward propagation without modifying the algorithm's parameters. The ideal outcome is a downward trend in the loss metrics for training and validation, leveling off at a minimal value, which would signify the algorithm's adeptness at acquiring knowledge and its proficiency in applying it to unfamiliar data. In contrast, a situation where the training loss diminishes steadily but the validation loss starts to increase would signal overfitting. This would mean that the algorithm needs to become more specialized to the quirks and irrelevant details of the training set, potentially impairing its effectiveness on new data. Hence, vigilant observation of the loss trends during training and validation is crucial to ensure the algorithm strikes a proper balance between learning effectively from the training set and maintaining its generalization capacity.

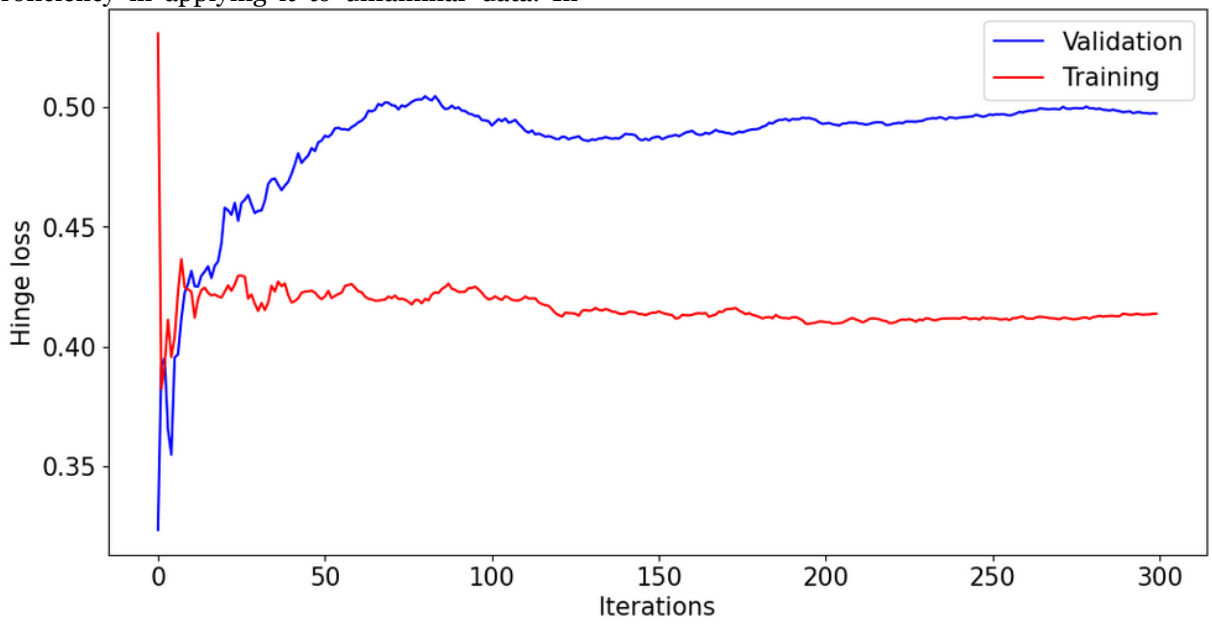

Fig. 4. Visual depiction of hinge loss trends for training and validation sets in the model.

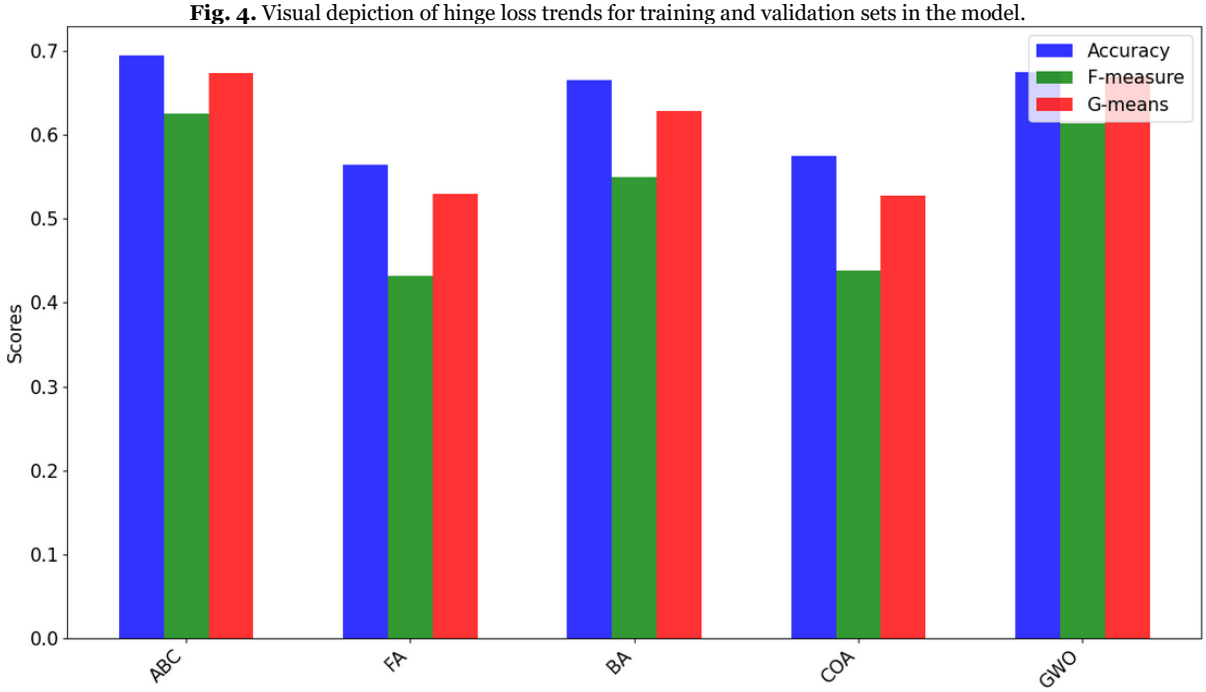

Fig. 5. The outcomes of diverse metaheuristic algorithms.

4.4.1. Analyze of the DE algorithm

In our forthcoming study, we aim to evaluate the performance of our DE algorithm relative to various other metaheuristic optimization techniques. The primary goal is to apply diverse metaheuristic strategies for hyperparameter optimization while keeping all other aspects of the model constant. We have included six algorithms in this comparison: Artificial Bee Colony (ABC) [21], Firefly Algorithm (FA)[64], Bat Algorithm (BA)[65], Cuckoo Optimization Algorithm (COA)[66], and Grey Wolf Optimization (GWO)[67]. The results of this comparative analysis are presented in Figure 5. The data indicates that the DE method achieves a remarkable 22% decrease in error rates compared to the ABC algorithm, highlighting its enhanced efficiency. Moreover, the ABC algorithm demonstrates superior performance over other algorithms like FA, GWO, and BA, yielding more advantageous outcomes. These findings suggest that our developed method surpasses the existing algorithms in accuracy and robustness.

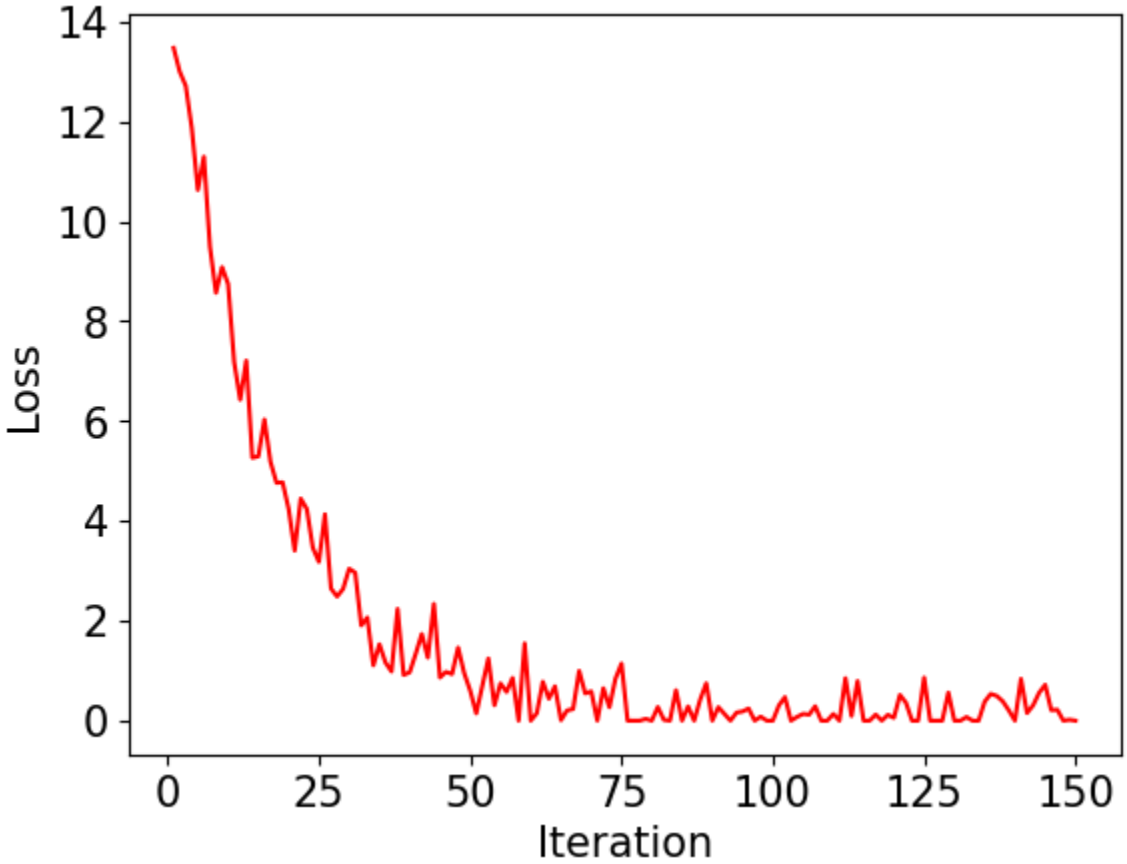

Fig. 6. Progression of the objective function across consecutive iterations using the DE algorithm.

Figure 6 visually presents the progression of the objective function during consecutive iterations within the DE process. The x-axis displays the number of iterations or generations, while the y-axis shows the corresponding values of the objective function. This layout provides a clear insight into the functional mechanisms of the algorithm. A detailed analysis of Figure 6 uncovers significant patterns. The initial phases exhibit marked variability in the objective function values, underscoring the DE algorithm's exploration stage. During this period, the algorithm conducts an extensive search across the solution landscape, aiming to avoid premature settling at local optima and to pinpoint areas of potential. As the process moves forward, a discernible trend toward stabilization becomes apparent. The variations in the objective function values start to decrease, indicating a more focused search by the algorithm. This phase involves capitalizing on the most promising solutions identified thus far and enhancing the search's depth to achieve the optimal outcome. Monitoring for any extended phases of stagnation in the objective function values is crucial, as these might signal that the algorithm has plateaued at a local maximum. Such instances may require modifying the algorithm's parameters or integrating it with additional methods to effectively boost its ability to explore and navigate the solution space.

4.5. Discussion

This article used the ImAUC-PSVM method and the DE algorithm for hyperparameter optimization to detect CVDs.

Several critical considerations underpin the choice of the ImAUC-PSVM for detecting CVDs. Firstly, CVDs present a unique challenge due to their complex nature and the often-imbalanced nature of clinical datasets, where non-occurrence cases significantly outnumber instances of disease occurrence. The ImAUC-PSVM directly addresses this imbalance by incorporating AUC maximization into the objective function, a metric particularly adept at handling skewed data distributions. This integration is pivotal as AUC provides a more nuanced assessment of model performance across different threshold settings, making it more reliable in clinical settings where false negatives can have profound implications. Secondly, by maintaining the structural integrity of the standard PSVM, ImAUC-PSVM ensures the retention of PSVM's inherent benefits. This includes the model's capacity to efficiently process incremental updates, a feature crucial for managing the dynamic progression of CVDs. The ability to handle rapid updates allows for more agile and timely responses in clinical applications, enhancing the model's practical utility. Thus, selecting ImAUC-PSVM is a strategic decision to optimize model performance for cardiovascular disease diagnosis's specific intricacies and challenges.

Incorporating the DE algorithm into our framework is a strategic choice to enhance the model's performance in navigating complex hyperparameter spaces. The nature of ImAUC-PSVM, with its nuanced approach to handling imbalanced datasets and its sophisticated model structure, necessitates precise hyperparameter tuning to achieve optimal performance. DE is known for its efficiency in exploring and exploiting diverse solution spaces, making it an ideal choice for our model. Its ability to simultaneously consider multiple potential solutions and evolve them through generations allows for a more comprehensive search of the hyperparameter space than traditional optimization methods. This capability is precious in medical applications like CVD detection, where the accuracy and robustness of the model can have significant implications on patient outcomes. Furthermore, DE's flexibility in adapting to various types of objective functions and constraints aligns well with the complex nature of healthcare data. By integrating DE, we aim to refine the model's performance, ensuring that it achieves high accuracy and maintains consistency and reliability across various scenarios, thereby enhancing its applicability in the dynamic field of healthcare analytics and patient care.

The theoretical implications of our research are significant, considering the pervasive challenge of CVD affecting the heart and vascular system worldwide. Our introduction of the ImAUC-PSVM leverages the core principles of conventional PSVM to enhance CVD detection. The essence of ImAUC-PSVM lies in its adept integration of AUC maximization into its objective function, streamlining the model by minimizing the need for extensive parameter tuning. This aspect mainly benefits imbalanced data sets, promoting an effective training regimen. Theoretically, ImAUC-PSVM mirrors the structural solution of traditional PSVM, inheriting its advantages, especially in dynamic CVD conditions that necessitate quick, incremental updates. Our methodological innovation extends to applying a customized DE algorithm for hyperparameter optimization, precisely navigating the intricate parameter landscape. Tested on a robust medical dataset from a 2012 survey involving 26,002 participants, including critical health metrics, the ImAUC-PSVM demonstrated remarkable accuracy in CVD identification. These findings validate the model's efficacy and hint at its potential to reshape the domain of healthcare analytics and patient management by providing a scalable and adaptable diagnostic tool.

The limitations of the proposed model are as follows:

- Data Dependency and Generalizability: One limitation of the ImAUC-PSVM model is its reliance on the specific characteristics of the dataset it was trained on. Although the model was validated using a comprehensive medical survey of 26,002 athletes, this dataset predominantly represents a particular population segment. Athletes generally have different physiological profiles than the general population, including non-athletes and older people, potentially limiting the model's generalizability[68]. Their model may perform less effectively when applied to broader, more diverse populations. To mitigate this, further validation with datasets encompassing a more comprehensive range of demographic and health characteristics is necessary to confirm the model's efficacy across various groups[69].
- Complexity in Hyperparameter Optimization: Integrating the DE algorithm, while beneficial for navigating complex hyperparameter spaces, also introduces complexity [23]. DE relies on a fine-tuned balance of exploration and exploitation to optimize hyperparameters, which can be time-consuming and computationally expensive. If not carefully managed, this could lead to longer

training times and increased computational resources, especially in large-scale applications. A potential solution is exploring other optimization algorithms that offer a better balance between performance and computational efficiency or implementing parallel computing techniques to expedite the optimization process [24].

- Sensitivity to Imbalanced Data: While the ImAUC-PSVM is designed to handle imbalanced datasets, its performance is inherently tied to the extent of this imbalance. In cases of extreme imbalance, where positive examples (CVD cases) are vastly outnumbered, the model may still struggle to identify these rare events accurately. This could lead to higher false negatives, particularly critical in medical diagnostics. One approach to address this might involve implementing additional techniques like synthetic data generation (e.g., SMOTE) [13] to artificially balance the dataset, thereby improving the model's ability to learn from underrepresented classes.
- Responsiveness to Rapid Incremental Updates: While adept at handling progressive CVD scenarios with rapid incremental updates, the model's design may face challenges in real-time responsiveness. In real-world healthcare settings, where data is continuously evolving, and immediate decision-making is often required, the model's ability to quickly adapt to new information is crucial [3]. Any lag in integrating and responding to new data can impact its practical utility. Enhancing the model's incremental learning capabilities through online learning algorithms or stream processing techniques could be explored to ensure the model remains agile and responsive in dynamic clinical environments [2].

5. Conclusion

This study presents a unique approach for learning from imbalanced data called the ImAUC-PSVM, which builds upon the conventional PSVM framework to detect CVDs. The ImAUC-PSVM approach comes with several notable strengths: 1) It adeptly embeds AUC maximization into its core objective function, streamlining the model by minimizing the necessity for extensive parameter tuning, thus enhancing its suitability for imbalanced data sets with a more efficient training methodology, 2) Our theoretical analysis indicates that ImAUC-PSVM upholds the fundamental solution architecture of the classic PSVM. This attribute allows it to leverage the advantages of PSVM, particularly useful in rapidly evolving CVD cases requiring quick incremental updates. Additionally, we have integrated a customized Differential Evolution (DE) algorithm, expertly crafted to traverse the intricate hyperparameter landscape effectively. We thoroughly tested the model's efficacy using a comprehensive 2012 medical study dataset involving a large group of 26,002 athletes. Vital metrics like height, weight, age, gender, blood pressure, and resting heart rate were carefully recorded. This methodology marks a significant step forward in cardiovascular diagnostics and presents a scalable and flexible solution that could have far-reaching impacts on healthcare analytics and patient treatment strategies.

Building on the promising foundation laid by our novel ImAUC-PSVM approach for CVD detection, our future research directions are geared towards further enhancing the algorithm's robustness and exploring its applicability across a broader range of medical conditions and healthcare scenarios. Initially, we plan to focus on algorithmic refinements, particularly optimizing the DE algorithm to improve hyperparameter selection efficiency, potentially integrating adaptive or machine learning-based hyperparameter tuning methods over the next two years. Concurrently, we aim to extend the application of ImAUC-PSVM to other imbalanced datasets within healthcare, such as rare diseases or conditions with subtle symptomatic expressions, over the next three to five years.

Another pivotal direction involves incorporating multi-modal data sources, including genomic, proteomic, and lifestyle data, to enrich the model's predictive capabilities. This multidimensional expansion, planned for the next five to seven years, seeks to harness the full spectrum of available patient data, offering a more holistic approach to disease prediction and management. Additionally, we will explore integrating our model into real-time healthcare monitoring systems, leveraging wearable technology and IoT devices to facilitate continuous health status assessment and early intervention.

Parallel to these technical enhancements, we anticipate collaborative research with interdisciplinary teams, including clinicians, biomedical engineers, and data scientists, to ensure that our developments align with practical healthcare needs and ethical standards. This collaborative approach will also open avenues for applying our model beyond cardiovascular health, potentially transforming predictive analytics in various medical domains. Through this structured yet adaptable roadmap, we aim to solidify the role of ImAUC-PSVM in advancing healthcare analytics and patient care, ultimately

contributing to the broader goal of personalized and preventive medicine.

REFERENCES

[1] I. Wronka, E. Suliga, and R. Pawliñska-Chmara, "Evaluation of lifestyle of underweight, normal weight and overweight young women," *Coll Antropol*, vol. 37, no. 2, pp. 359–365, 2013.

[2] L. Hong *et al.*, "GAN-LSTM-3D: An efficient method for lung tumour 3D reconstruction enhanced by attention-based LSTM," *CAAI Trans Intell Technol*, 2023.

[3] H. Zareiamand, A. Darroudi, I. Mohammadi, S. V. Moravvej, S. Danaei, and R. Alizadehsani, "Cardiac magnetic resonance imaging (cmri) applications in patients with chest pain in the emergency department: a narrative review," *Diagnostics*, vol. 13, no. 16, p. 2667, 2023.

[4] A. Salam and J. A. McGrath, "Diagnosis by numbers: defining skin disease pathogenesis through collated gene signatures," *Journal of Investigative Dermatology*, vol. 135, no. 1, pp. 17–19, 2015.

[5] C. Campbell and K. Bennett, "A linear programming approach to novelty detection," *Adv Neural Inf Process Syst*, vol. 13, 2000.

[6] W.-K. Wong, A. W. Moore, G. F. Cooper, and M. M. Wagner, "Bayesian network anomaly pattern detection for disease outbreaks," in *Proceedings of the 20th international conference on machine learning (ICML-03)*, 2003, pp. 808–815.

[7] J.-F. Fontaine, J. Priller, E. Spruth, C. Perez-Iratxeta, and M. A. Andrade-Navarro, "Assessment of curated phenotype mining in neuropsychiatric disorder literature," *Methods*, vol. 74, pp. 90–96, 2015.

[8] L. Sacchi, A. Tucker, S. Counsell, D. Garway-Heath, and S. Swift, "Improving predictive models of glaucoma severity by incorporating quality indicators," *Artif Intell Med*, vol. 60, no. 2, pp. 103–112, 2014.

[9] S. V. Moravvej, S. J. Mousavirad, M. H. Moghadam, and M. Saadatmand, "An LSTM-based plagiarism detection via attention mechanism and a population-based approach for pre-training parameters with imbalanced classes," in *Neural Information Processing: 28th International Conference, ICONIP 2021, Sanur, Bali, Indonesia, December 8–12, 2021, Proceedings, Part III 28*, Springer, 2021, pp. 690–701.

[10] M. S. Sartakhti, M. J. M. Kahaki, S. V. Moravvej, M. javadi Joortani, and A. Bagheri, "Persian language model based on BiLSTM model on COVID-19 corpus," in *2021 5th International Conference on Pattern Recognition and Image Analysis (IPRIA)*, IEEE, 2021, pp. 1–5.

[11] N. V Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: synthetic minority over-sampling technique," *Journal of artificial intelligence research*, vol. 16, pp. 321–357, 2002.

[12] S. Taherinavid, S. V. Moravvej, Y.-L. Chen, J. Yang, C. S. Ku, and L. Y. Por, "Automatic Transportation Mode Classification Using a Deep Reinforcement Learning Approach With Smartphone Sensors," *IEEE Access*, 2023.

[13] S. V. Moravvej *et al.*, "RLMD-PA: A reinforcement learning-based myocarditis diagnosis combined with a population-based algorithm for pretraining weights," *Contrast Media Mol Imaging*, vol. 2022, 2022.

[14] S. V. Moravvej, M. Joodaki, M. J. M. Kahaki, and M. S. Sartakhti, "A method based on an attention mechanism to measure the similarity of two sentences," in *2021 7th International Conference on Web Research (ICWR)*, IEEE, 2021, pp. 238–242.

[15] S. Danaei *et al.*, "Myocarditis diagnosis: a method using mutual learning-based abc and reinforcement learning," in *2022 IEEE 22nd International Symposium on Computational Intelligence and Informatics and 8th IEEE International Conference on Recent Achievements in Mechatronics, Automation, Computer Science and Robotics (CINTI-MACRo)*, IEEE, 2022, pp. 265–270.

[16] S. Pang, L. Zhu, G. Chen, A. Sarrafzadeh, T. Ban, and D. Inoue, "Dynamic class imbalance learning for incremental LPSVM," *Neural Networks*, vol. 44, pp. 87–100, 2013.

[17] S. V Moravvej, M. J. Maleki Kahaki, M. Salimi Sartakhti, and M. Joodaki, "Efficient GAN-based method for extractive summarization," *Journal of Electrical and Computer Engineering Innovations (JECEI)*, vol. 10, no. 2, pp. 287–298, 2022.

[18] J. Crossa, J. W. R. Martini, and D. Gianola, "Deep kernel and deep learning for genome-based prediction of single traits in multienvironment breeding trials," *Front Genet*, vol. 10, p. 498428, 2019.

[19] S. V. Moravvej, A. Mirzaei, and M. Safayani, "Biomedical text summarization using conditional generative adversarial network (CGAN)," *arXiv preprint arXiv:2110.11870*, 2021.

[20] S. Vakilian, S. V. Moravvej, and A. Fanian, "Using the cuckoo algorithm to optimizing the response time and energy consumption cost of fog nodes by considering collaboration in the fog layer," in *2021 5th International Conference on Internet of Things and Applications (IoT)*, IEEE, 2021, pp. 1–5.

[21] S. Vakilian, S. V. Moravvej, and A. Fanian, "Using the artificial bee colony (ABC) algorithm in collaboration with the fog nodes in the Internet of Things three-layer architecture," in *2021 29th Iranian Conference on Electrical Engineering (ICEE)*, IEEE, 2021, pp. 509–513.

[22] K. V Price, "Differential evolution," in *Handbook of optimization: From classical to modern approach*, Springer, 2013, pp. 187–214.

[23] S. V. Moravvej, S. J. Mousavirad, D. Oliva, G. Schaefer, and Z. Sobhaninia, "An improved de algorithm to optimise the learning process of a bert-based plagiarism detection model," in *2022 IEEE Congress on Evolutionary Computation (CEC)*, IEEE, 2022, pp. 1–7.

[24] S. V. Moravvej, S. J. Mousavirad, D. Oliva, and F. Mohammadi, "A novel plagiarism detection approach combining bert-based word embedding, attention-based lstms and an improved differential evolution algorithm," *arXiv preprint arXiv:2305.02374*, 2023.

[25] H. Gharagozlou, J. Mohammadzadeh, A. Bastanfard, and S. S. Ghidary, "RLAS-BIABC: A

reinforcement learning-based answer selection using the bert model boosted by an improved ABC algorithm," *Comput Intell Neurosci*, vol. 2022, 2022.
[26] G. Wang *et al.*, "Interactive medical image segmentation using deep learning with image-specific fine tuning," *IEEE Trans Med Imaging*, vol. 37, no. 7, pp. 1562–1573, 2018.
[27] J. Zhang, K. Kowsari, J. H. Harrison, J. M. Lobo, and L. E. Barnes, "Patient2vec: A personalized interpretable deep representation of the longitudinal electronic health record," *IEEE Access*, vol. 6, pp. 65333–65346, 2018.
[28] H.-C. Shin *et al.*, "Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning," *IEEE Trans Med Imaging*, vol. 35, no. 5, pp. 1285–1298, 2016.
[29] R. M. Ghoniem, "A novel bio-inspired deep learning approach for liver cancer diagnosis," *Information*, vol. 11, no. 2, p. 80, 2020.
[30] S. Nikhar and A. M. Karandikar, "Prediction of heart disease using machine learning algorithms," *International Journal of Advanced Engineering, Management and Science*, vol. 2, no. 6, p. 239484, 2016.
[31] M. Abdar, W. Książek, U. R. Acharya, R.-S. Tan, V. Makarenkov, and P. Pławiak, "A new machine learning technique for an accurate diagnosis of coronary artery disease," *Comput Methods Programs Biomed*, vol. 179, p. 104992, 2019.
[32] I. Maglogiannis, E. Loukis, E. Zafiropoulos, and A. Stasis, "Support vectors machine-based identification of heart valve diseases using heart sounds," *Comput Methods Programs Biomed*, vol. 95, no. 1, pp. 47–61, 2009.
[33] H. Tjahjadi and K. Ramli, "Noninvasive blood pressure classification based on photoplethysmography using k-nearest neighbors algorithm: a feasibility study," *Information*, vol. 11, no. 2, p. 93, 2020.
[34] S. A. Pattekari and A. Parveen, "Prediction system for heart disease using Naïve Bayes," *International Journal of Advanced Computer and Mathematical Sciences*, vol. 3, no. 3, pp. 290–294, 2012.
[35] D. Shah, S. Patel, and S. K. Bharti, "Heart disease prediction using machine learning techniques," *SN Comput Sci*, vol. 1, no. 6, p. 345, 2020.
[36] S. V. Moravvej, M. J. M. Kahaki, M. S. Sartakhti, and A. Mirzaei, "A method based on attention mechanism using bidirectional long-short term memory (BLSTM) for question answering," in *2021 29th Iranian Conference on Electrical Engineering (ICEE)*, IEEE, 2021, pp. 460–464.
[37] A. H. Chen, S.-Y. Huang, P.-S. Hong, C.-H. Cheng, and E.-J. Lin, "HDPS: Heart disease prediction system," in *2011 computing in Cardiology*, IEEE, 2011, pp. 557–560.
[38] H. A. Esfahani and M. Ghazanfari, "Cardiovascular disease detection using a new ensemble classifier," in *2017 IEEE 4th international conference on knowledge-based engineering and innovation (KBEI)*, IEEE, 2017, pp. 1011–1014.
[39] S. Bashir, U. Qamar, and F. H. Khan, "IntelliHealth: a medical decision support application using a novel weighted multi-layer classifier ensemble framework," *J Biomed Inform*, vol. 59, pp. 185–200, 2016.
[40] S. Bashir, A. A. Almazroi, S. Ashfaq, A. A. Almazroi, and F. H. Khan, "A knowledge-based clinical decision support system utilizing an intelligent ensemble voting scheme for improved cardiovascular disease prediction," *IEEE Access*, vol. 9, pp. 130805–130822, 2021.
[41] S. Mohan, C. Thirumalai, and G. Srivastava, "Effective heart disease prediction using hybrid machine learning techniques," *IEEE access*, vol. 7, pp. 81542–81554, 2019.
[42] Ö. Arslan and M. Karhan, "Effect of Hilbert-Huang transform on classification of PCG signals using machine learning," *Journal of King Saud University-Computer and Information Sciences*, vol. 34, no. 10, pp. 9915–9925, 2022.
[43] Z. Zhu, X. Zhu, Y. Guo, Y. Ye, and X. Xue, "Inverse matrix-free incremental proximal support vector machine," *Decis Support Syst*, vol. 53, no. 3, pp. 395–405, 2012.
[44] S.-Y. Lo, P. Oza, and V. M. Patel, "Adversarially robust one-class novelty detection," *IEEE Trans Pattern Anal Mach Intell*, vol. 45, no. 4, pp. 4167–4179, 2022.
[45] D. Chicco and G. Jurman, "The Matthews correlation coefficient (MCC) should replace the ROC AUC as the standard metric for assessing binary classification," *BioData Min*, vol. 16, no. 1, p. 4, 2023.
[46] M. Huang *et al.*, "Auc-oriented graph neural network for fraud detection," in *Proceedings of the ACM web conference 2022*, 2022, pp. 1311–1321.
[47] A. M. Carrington *et al.*, "Deep ROC analysis and AUC as balanced average accuracy, for improved classifier selection, audit and explanation," *IEEE Trans Pattern Anal Mach Intell*, vol. 45, no. 1, pp. 329–341, 2022.
[48] G. Wang, K. W. Wong, and J. Lu, "AUC-based extreme learning machines for supervised and semi-supervised imbalanced classification," *IEEE Trans Syst Man Cybern Syst*, vol. 51, no. 12, pp. 7919–7930, 2020.
[49] G. Wang, J. Y.-C. Teoh, J. Lu, and K.-S. Choi, "Least squares support vector machines with fast leave-one-out AUC optimization on imbalanced prostate cancer data," *International Journal of Machine Learning and Cybernetics*, vol. 11, no. 8, pp. 1909–1922, 2020.
[50] P. Zhao, S. C. H. Hoi, R. Jin, and T. Yang, "Online AUC maximization," 2011.
[51] T. Yang and Y. Ying, "AUC maximization in the era of big data and AI: A survey," *ACM Comput Surv*, vol. 55, no. 8, pp. 1–37, 2022.
[52] V. Cherkassky and Y. Ma, "Practical selection of SVM parameters and noise estimation for SVM regression," *Neural networks*, vol. 17, no. 1, pp. 113–126, 2004.
[53] J. Zhou, S. Huang, M. Wang, and Y. Qiu, "Performance evaluation of hybrid GA–SVM and GWO–SVM models to predict earthquake-induced liquefaction potential of soil: a multi-dataset investigation," *Eng Comput*, pp. 1–19, 2022.
[54] S. Huang, N. Cai, P. P. Pacheco, S. Narrandes, Y. Wang, and W. Xu, "Applications of support vector machine (SVM) learning in cancer genomics," *Cancer Genomics Proteomics*, vol. 15, no. 1, pp. 41–51, 2018.
[55] J. Han, C. Gondro, K. Reid, and J. P. Steibel,

"Heuristic hyperparameter optimization of deep learning models for genomic prediction," *G3*, vol. 11, no. 7, p. jkab032, 2021.
[56] D. Barbieri *et al.*, "Predicting cardiovascular risk in athletes: resampling improves classification performance," *Int J Environ Res Public Health*, vol. 17, no. 21, p. 7923, 2020.
[57] H. Ahmad, B. Kasasbeh, B. AL-Dabaybah, and E. Rawashdeh, "EFN-SMOTE: An effective oversampling technique for credit card fraud detection by utilizing noise filtering and fuzzy c-means clustering," *International Journal of Data and Network Science*, vol. 7, no. 3, pp. 1025–1032, 2023.
[58] M. A. de Almeida, "DATA MINING: DETERMINAC AO DE AGRUPAMENTOS EM GRANDES BASES DE DADOS," *Universidade Federal do Rio de Janeiro*, 2013.
[59] S. Goyal, "Effective software defect prediction using support vector machines (SVMs)," *International Journal of System Assurance Engineering and Management*, vol. 13, no. 2, pp. 681–696, 2022.
[60] A. J. Gallego, J. R. Rico-Juan, and J. J. Valero-Mas, "Efficient k-nearest neighbor search based on clustering and adaptive k values," *Pattern Recognit*, vol. 122, p. 108356, 2022.
[61] J. Hu and S. Szymczak, "A review on longitudinal data analysis with random forest," *Brief Bioinform*, vol. 24, no. 2, p. bbad002, 2023.
[62] E. C. Zabor, C. A. Reddy, R. D. Tendulkar, and S. Patil, "Logistic regression in clinical studies," *Int J Radiat Oncol Biol Phys*, vol. 112, no. 2, pp. 271–277, 2022.
[63] C. S. Lee, P. Y. S. Cheang, and M. Moslehpour, "Predictive analytics in business analytics: decision tree," *Advances in Decision Sciences*, vol. 26, no. 1, pp. 1–29, 2022.
[64] X.-S. Yang, "Firefly algorithm, stochastic test functions and design optimisation," *International journal of bio-inspired computation*, vol. 2, no. 2, pp. 78–84, 2010.
[65] X.-S. Yang, "A new metaheuristic bat-inspired algorithm," in *Nature inspired cooperative strategies for optimization (NICSO 2010)*, Springer, 2010, pp. 65–74.
[66] X.-S. Yang and S. Deb, "Cuckoo search via Lévy flights," in *2009 World congress on nature & biologically inspired computing (NaBIC)*, Ieee, 2009, pp. 210–214.
[67] S. Mirjalili, S. M. Mirjalili, and A. Lewis, "Grey wolf optimizer," *Advances in engineering software*, vol. 69, pp. 46–61, 2014.
[68] F. Peier, M. Mouthon, M. De Pretto, and J. N. Chabwine, "Response to experimental cold-induced pain discloses a resistant category among endurance athletes, with a distinct profile of pain-related behavior and GABAergic EEG markers: a case–control preliminary study," *Front Neurosci*, vol. 17, p. 1287233, 2024.
[69] A. Bostani *et al.*, "MLP-RL-CRD: diagnosis of cardiovascular risk in athletes using a reinforcement learning-based multilayer perceptron," *Physiol Meas*, vol. 44, no. 12, p. 125012, 2023.