

# A Focal Loss-based Multi-layer Perceptron for Diagnosis of Cardiovascular Risk in Athletes

Chuan Yang\*

School of Public Utility Management, Chongqing Vocational College of Transportation, Chongqing 402247, China

**Abstract**—Cardiovascular diseases (CVDs) are a prevalent cause of heart failure around the world. This research was required in order to investigate potential approaches to treating the disease. The article presents a focal loss (FL)-based multi-layer perceptron called MLP-FL-CRD to diagnose cardiovascular risk in athletes. In 2012, 26,002 athletes were measured for their height, weight, age, sex, blood pressure, and pulse rate in a medical exam that had electrocardiography at rest. Outcomes were negative for the largest majority, leading to class imbalance. Training on imbalanced data hurts classifier performance. To address this, the study proposes a training approach based on focal loss, which effectively emphasizes minority class examples. Focal loss softens the influence of simplistic samples, enabling the model to concentrate on more intricate examples. It is useful in circumstances when there is a substantial class imbalance. Additionally, the paper highlights a challenge in the training phase, often characterized by the use of gradient-based learning methods like backpropagation. These methods exhibit several disadvantages, including sensitivity to initialization. The paper recommends the implementation of a mutual learning-based artificial bee colony (ML-ABC). This approach adjusts the primary weight by substituting the food resource candidate, which is selected due to superior fitness, with one based on a mutual learning factor between two individuals. The sample obtains great outcomes, outperforming other machine learning samples. Optimal values for important parameters are identified for the model based on experiments on the study dataset. Ablation studies that exclude FL and ML-ABC of the sample confirm the additive effect of, which is not negative and dependent, these factors on the sample's efficiency.

**Keywords**—Cardiovascular diseases; multi-layer perceptron; focal loss; artificial bee colony; imbalanced classification

## I. INTRODUCTION

As per the World Health Organization's discoveries, cardiovascular diseases are responsible for the most deaths across the globe, leading to 17.9 million fatalities every year [1]. Smoking, obesity, high blood pressure, and physical inactivity are the most significant risk elements. Even though the risk of cardiovascular diseases can be minimized through regular exercise, athletes competing or training with intensity and frequency remain at risk due to the intensity and frequency of their activities. Athletes are consistently monitored by sports physicians, who gain biomedical and personal information, along with executing electrocardiography screenings (ECG). According to the ECG results, people are classified as either a risk factor or not a risk factor. Those deemed at risk may not gain medical clearance for participation in sports and will be subject to further assessments. There is an unequal ratio between the two categories, as the N class contains a larger

percentage of individuals, while the more remarkable P class is not very common. Generally speaking, a false negative (FN) may have severe consequences, potentially including a human fatality. In contrast, a false positive (FP) would lead to additional medical examinations and temporarily halt sports practice [2].

Classification is indeed a method of machine learning that can expect binary classification values, such as P or N, and it could have great use in the health world, particularly in diagnostics. Several machine learning techniques have already been implemented in medical diagnosis systems to assist decision-making. These techniques may estimate health risks from huge datasets. Wong et al. [3] utilized Bayesian networks to uncover past disease outbreaks and achieved remarkable outcomes on actual data taken from a database that contained seven years of medical records from an emergency region. The dataset was collected from unwell patients in a hospital, indicating that the research was aimed at epidemiological investigation rather than diagnostic purposes. This strategy could also be implemented in counterterrorism to discover biological attacks. Campbell et al. [4] utilized a kernel-based approach that effectively identified a rare disease from medical data. However, the size of the dataset was limited, and the proportion of noteworthy instances in the test set was considerably greater than the prevalence of the disease among the overall population. Fontaine et al. [5] investigated methods of data mining to enhance the clinical assessment of patients with brain disorders. Salam and McGrath [6] conducted a machine-learning strategy for dermatology. In this study, a classifier for multiple diseases improved the identification of skin infections. Sacchi et al. [7] applied a Naive Bayes classifier for glaucoma prediction. Because of the small and unbalanced dataset, resampling and bootstrapping were employed to train their model. Numerous authors [8, 9] have suggested comparing different classification methods in medical statistics to identify the significant benefits of one approach compared to others. There is still a dearth of studies using large datasets in regions where diseases have a low prevalence, but individuals may face increased risks due to heightened levels of pressure or stress. Moreover, there is a continuous debate about the necessity for affordable and efficient healthcare, as well as the more cautious utilization of medical tests.

In machine learning approaches, the strategy for extracting features is inflexible, resulting in poor generalization ability, rising time, and low precision [10]. With the emergence of profound learning methods in numerous usages [11], numerous investigators have employed those for categorization [12].

Deep learning can accurately learn high-level features due to its layered structure. Multi-layer perceptron is a universal approximation initially developed for nonlinear XOR and has been implemented effectively for diverse combinatory enhancement issues ever since [13]. MLP is widely utilized for a variety of tasks, including information processing, pattern recognition, classification issues, image processing, linear and nonlinear optimization issues, and real data prediction. MLP functions as a universal approximation where input signals propagate forward. The processing node related to a human neuron is the fundamental component of the ANN method. Each processing node gets a collection of intake deals, adds them, and after that, passes this sum through an activation function that determines the node's output value. In MLP, nodes comprise fully interconnected layers, with the exception that nodes within the same layer are not interconnected [14].

Medical classification faces significant challenges due to data imbalance, which can significantly lower performance because there are far more negative instances [15, 16]. Over-sampling, under-sampling, or a compound of both are used in the data-level strategy to mitigate the negative effects of imbalanced classification [17, 18]. Approaches at the algorithmic level give the minority class more weight [19]. Additionally, profound learning techniques can be used to solve the categorization balancing problem [20]. Huang et al. [21] formulate a process to identify distinguishing features of imbalanced data while upholding inter-cluster and inter-class margins. Yan et al. [22] suggested a technique using the bootstrapping method to balance data of convolutional networks across mini-batches.

Deep models are widely used in natural language processing, computer vision, and medical image analysis [23, 24]. The best solution for optimizing the neural network can be selected from a population of created models using population-based training. This approach is less likely to get stuck in local optima than traditional training methods [25, 26]. Indeed, a simple evolutionary algorithm was found to rival stochastic gradient descent for neural network training. Jaderberg et al. [27] applied population-based training to state-of-the-art models of deep RL, machine translation, and generative adversarial networks and demonstrated consistent improvements in accuracy, training time, and stability. In [28] and [29], effective training of the weights of neural networks was achieved using a differential evolution-based strategy and Artificial Bee Colony (ABC), respectively. The ABC algorithm can be improved by the mutual learning-based ABC [30], which changes the algorithm to use mutual learning between two selected position parameters instead of choosing the candidate food source with the highest fitness [31].

In response to the challenges outlined, this article introduces an FL-based MLP, designated as MLP-FL-CRD, for diagnosing cardiovascular risk in athletes within an imbalanced dataset. The proposed MLP-FL-CRD model contains an MLP for sick and healthy inputs that employ FL to class imbalance. An ML-ABC is utilized for weight initialization to identify a promising area within the study environment for initiating the BP method in the model. This process involves continuous evaluation of the understanding of athleticism, achieved through shared knowledge among current and nearby nutrition

resources, to develop a more valuable nutrition resource. The MLP-FL-CRD model is assessed on a dataset comprising medical examinations of 26,002 athletes, demonstrating its superiority over other approaches. Furthermore, ablation studies are conducted to evaluate the relative contributions of the FL and pre-training strategies. Various alternative model component options, e.g., evolutionary algorithms and loss functions, are tested and compared in a series of experiments on the same dataset to investigate how to obtain the best results.

The main contributions of the proposed model are as follows:

- **Innovative Use of FL for Class Imbalance:** The model employs focal loss to effectively manage the class imbalance in the dataset. This approach is particularly beneficial in circumstances with substantial class disparities, as it reduces the impact of simpler, more common examples (negative outcomes in this case). This allows the model to focus more on complex, minority class examples (positive cases), which are crucial for accurate diagnosis in medical settings.
- **Identification of Optimal Model Parameters:** The research has also led to the identification of optimal values for key parameters in the model, which is crucial for its application in real-world scenarios. This ensures that the model can be fine-tuned for maximum efficiency and accuracy in diagnosing cardiovascular risk.
- **Enhanced Training Approach with ML-ABC:** The model introduces a ML-ABC for the training phase. This method innovatively addresses the sensitivity to initial weight settings inherent in gradient-based learning methods like backpropagation. By using ML-ABC, the model adapts the primary weights based on mutual learning factors between two individuals, leading to a more efficient training process and potentially better performance.

The organization of this paper is as follows. Section I is about the Introduction, Section II delves into related work. Method is proposed in Section III. Section IV presents the experimental results and proposes ideas for further endeavors and Section V concludes the conclusion.

## II. RELATED WORK

In the past few years, the health sector has witnessed notable progress in data analysis and the application of machine learning methods. These approaches have been extensively embraced and proven effective across a range of medical applications, especially within cardiac medicine. The burgeoning growth of medical data affords scholars a unique chance to devise and evaluate novel algorithms in this domain. CVDs continue to be a predominant cause of death in less developed countries [32, 33], and pinpointing risk elements and preliminary indicators of such illnesses is now a crucial research endeavor. The adoption of data analysis and machine learning strategies in this realm could substantially contribute to the timely detection and deterrence of cardiac diseases.

To date, research employing machine learning techniques has been advanced for CVD. The research conducted by Narain et al. [34] aimed to develop a sophisticated, machine-learning-driven CVD prognosis model to enhance the accuracy of the established Framingham risk score (FRS). Utilizing data from 689 subjects exhibiting CVD symptoms and a validation cohort from the Framingham study, the newly proposed model, employing a quantum neural network to discern CVD patterns, underwent experimental validation and comparison with the FRS. The model's proficiency in predicting CVD risk was ascertained to be 98.57%, substantially surpassing the 19.22% accuracy of the FRS and other current methods. The findings suggest the model could serve as a valuable asset for medical professionals in predicting CVD risk, thus aiding in the formulation of superior treatment strategies and promoting prompt diagnosis. Shah et al. [35] sought to construct a cardiovascular disease prediction model using machine learning tools. The data, comprising 303 records and 17 attributes, were sourced from the Cleveland heart disease database available in the UCI machine learning repository. The team employed several supervised classification techniques, including naive Bayes, decision tree, random forest, and k-nearest neighbor (KKN). The study's outcomes revealed that the KKN model demonstrated the highest predictive accuracy, reaching 90.8%, underscoring the promise of machine learning methods in forecasting cardiovascular disease and the importance of model and technique selection for best results. Drod et al. [36] embarked on a study to pinpoint crucial CVD risk factors among patients suffering from metabolic-associated fatty liver disease (MAFLD), utilizing machine learning methodologies. They performed blood biochemistry analyses and subclinical atherosclerosis assessments on a cohort of 191 individuals diagnosed with MAFLD. The research team crafted a model that incorporated ML techniques, including a multiple logistic regression classifier, univariate feature ranking, and principal component analysis (PCA). This model aimed to identify patients at heightened risk for CVD. The findings highlighted hypercholesterolemia, plaque scores, and the duration of diabetes as the most critical clinical indicators. Employing the ML approach, the study was able to effectively distinguish 85.11% of patients at high risk and 79.17% of those at low risk for CVD, achieving an Area Under the Curve (AUC) of 0.87. This underscores the efficacy of ML tools in identifying high-risk MAFLD patients for CVD using basic patient data. In research by Alotalibi et al. [37] the aim was to explore the effectiveness of various machine learning (ML) methodologies in forecasting heart failure incidents. Utilizing patient data sourced from the Cleveland Clinic Foundation, the research applied a variety of ML algorithms, including decision tree, logistic regression, random forest, naive Bayes, and support vector machine (SVM). Predictive models were developed using a rigorous 10-fold cross-validation approach. Among these, the decision tree algorithm was identified as the top performer, achieving a remarkable prediction accuracy of 93.19%, closely followed by

SVM with an accuracy of 92.30%. This study highlights the significant potential of ML strategies in predicting heart failure, particularly underlining the decision tree algorithm's efficacy, making it a prime candidate for further research endeavors. Comparing various algorithms, Hasan and Bao [38] undertook a research project focused on identifying the most efficient feature selection technique for forecasting cardiovascular diseases. The study initially evaluated three well-known feature selection strategies—filter, wrapper, and embedded methods—and generated feature subsets using a standard "True" condition in a Boolean framework. This dual-phase selection procedure was then applied across various models such as random forest, support vector classifier (SVC), k-nearest neighbors, naive Bayes, and XGBoost, to assess their predictive accuracy and establish the best performing model. The study used the artificial neural network (ANN) as a reference point for these comparisons. Results from the investigation highlighted that the XGBoost classifier, combined with the wrapper feature selection method, was the most effective in predicting cardiovascular diseases. XGBoost recorded a 73.74% accuracy rate, closely trailed by the SVC with 73.18% and the ANN with 73.20%.

Research in the realm of cardiovascular diseases utilizing deep learning techniques has yielded promising results. Mohan et al. [39] developed a Multi-Task Deep and Wide Neural Network (MT-DWNN) designed for simultaneous multiple tasks, aiming to predict critical incidents during hospitalizations. This algorithm was evaluated using an extensive dataset covering 18 years, encompassing 35,101 instances of hospital admissions due to heart failure and 2,478 cases of renal failure at the Chinese PLA General Hospital. The MT-DWNN's ability to forecast renal complications (with an AUC of 0.9393) outperformed conventional approaches, surpassing the AUC scores of standalone deep neural networks (0.9370), the random forest model (0.9360), and logistic regression (which scored below 0.9233). The results of these experiments indicate that the MT-DWNN is highly effective in predicting renal issues in heart failure patients. In a separate study, Arslan and Karhan [40] introduced a duo of sophisticated deep neural networks, specifically aimed at accurately predicting the risk of coronary heart disease. Common prediction models often struggle with the inconsistencies typical in many real-world datasets. To overcome this, the researchers proposed a unique approach for assembling training data by dividing the original dataset into two parts: one with a general distribution and the other with a significant bias. They employed a two-phase data processing technique, wherein variable autoencoders initially split the training data into these two distinct categories. Following this, two separate deep neural network classifiers are trained on these datasets. The efficacy of this method was evident, as it achieved an AUC of 0.882 and an accuracy rate of 0.892, outshining conventional methods in several key metrics, including specificity, accuracy, precision, recall, and the F-measure (0.915).

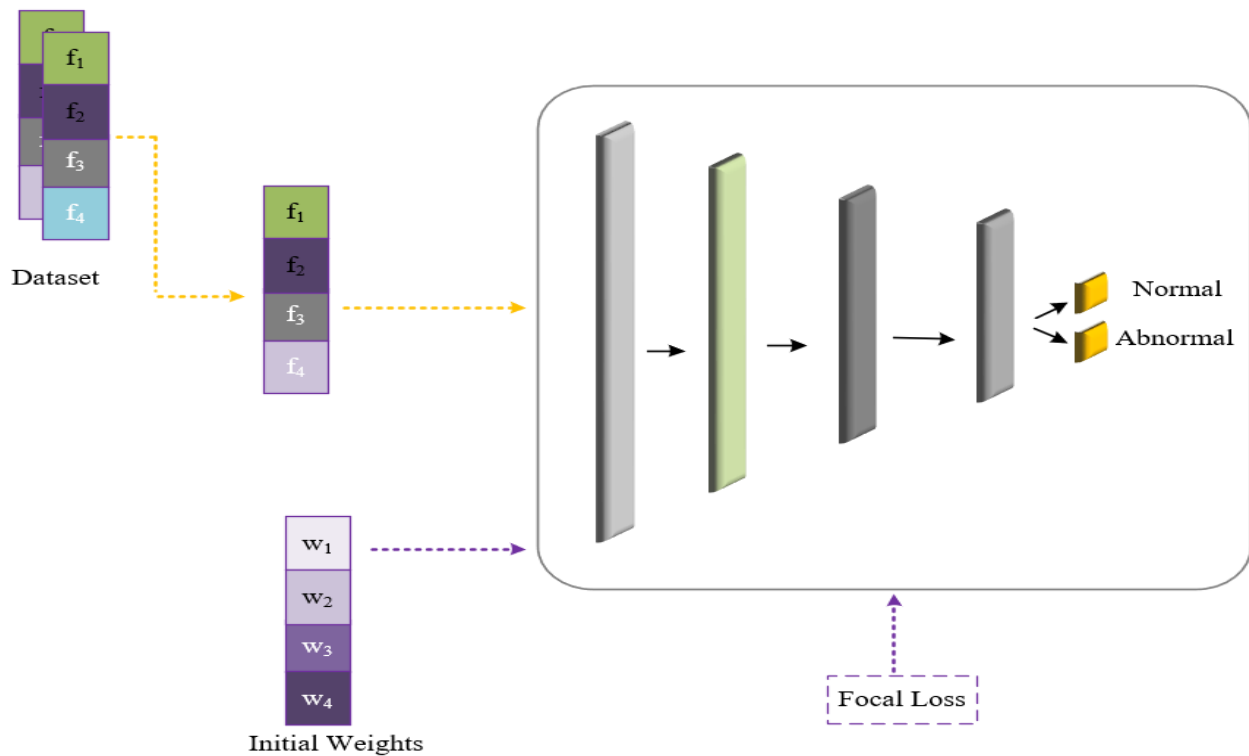


Fig. 1. Overview of the MLP-FL-CRD model.

### III. PROPOSED METHOD

The structure of the MLP-FL-CRD model is shown in Fig. 1. MLP-FL-CRD is specifically engineered to augment the diagnosis of cardiovascular diseases, addressing key challenges such as imbalanced class distribution and the essential need for accurate initial weight determination. The integration of ML-ABC and RL within the framework effectively tackles these pivotal issues where conventional models often falter.

Traditional algorithms typically lack a structured approach for selecting initial weights, potentially impeding the learning process. This can lead to slower rates of convergence and the possibility of converging on less than optimal minima. Furthermore, prevalent models struggle with the issue of class imbalance, a common obstacle in cardiovascular diagnosis where significant events are infrequent. Such models usually exhibit bias towards the majority class, leading to the critical underrepresentation of minority classes. These minority classes are particularly vital to identify accurately in cardiovascular diagnostics. The proposed method, utilizing ML-ABC, introduces a thoughtfully curated and varied assortment of initial weights. This diversity aids the model in avoiding local minima and facilitates more effective convergence towards a comprehensive global solution.

Additionally, the RL aspect of the model is meticulously designed to provide greater rewards for accurately classifying the minority class, thus recalibrating the model's focus towards these crucial predictions. This marks a significant enhancement over traditional supervised learning techniques, which might lack adequate representative data for effective training across diverse classes. The flexible learning policy of RL fosters a more equitable exploration of the decision-making space. This

leads to the formulation of strategies that prioritize the precise classification of lesser-represented classes. The adaptive capacity of RL within the model distinctly differentiates it from existing methodologies, rendering it capable of surmounting the intrinsic challenges typically encountered in conventional classification models, particularly in the realm of cardiovascular diagnostics.

#### A. Artificial Bee Colony Method

ABC [10] is a sophisticated optimization technique that simulates the foraging behavior of honey bees. Central to the ABC algorithm are four key components: employed bees, onlookers, scouts, and food sources. Employed bees are responsible for exploring nearby areas around an initial food source. Once they gather information, they return to the hive and share this knowledge with onlooker bees [41]. These onlooker bees, in turn, evaluate the potential of the food sources based on the information conveyed, including the likelihood of finding nectar. The decision-making process of onlooker bees involves assessing the probability of a food source's availability and its richness in nectar. Should a particular food source become depleted or no longer viable, the employed bee associated with that source undergoes a transformation. This bee becomes a scout, embarking on a random search for new and potentially more lucrative food sources [30]. This aspect of the algorithm exemplifies a dynamic optimization process, mirroring the adaptive and efficient foraging strategies of real-world honey bees. This bee-inspired algorithm is particularly effective in solving complex optimization problems due to its ability to explore and exploit resources. The ABC algorithm's balanced approach to exploration (via scouts) and exploitation (through employed

and onlooker bees) ensures a comprehensive search of the solution space, avoiding premature convergence on suboptimal solutions. This makes it an ideal choice for applications in various fields requiring robust optimization solutions, including data mining, engineering, and, as demonstrated in this study, healthcare analytics.

Eq. (1) presents a methodology for recalculating the position of an employed bee within the context of the ABC algorithm. This reformation of position is contingent upon the nectar quality associated with the new potential location. If the new position offers a higher nectar quality compared to the previous location, the bee is programmed to remember this new position and discard the previous one. This decision-making process is based on the principle of seeking the most rewarding nectar source, which in algorithmic terms translates to finding a more optimal solution. Conversely, if the nectar quality at the new position does not surpass that of the original location, the bee retains its prior position. This aspect of the algorithm ensures that beneficial positions are not forsaken for lesser or equivalent ones, thereby optimizing the search process. The emphasis on comparing and selecting positions based on nectar quality (or solution fitness) mirrors natural foraging behaviors and is a key factor in the ABC algorithm's ability to effectively navigate and exploit the solution space. This rule of position updating based on nectar quality highlights the algorithm's capacity for adaptive learning. It allows the algorithm to dynamically adjust its search strategy based on the evolving understanding of the solution landscape, enhancing its efficiency in locating and converging on optimal or near-optimal solutions.

$$v_i^j = x_i^j + \varphi_i^j(x_i^j - x_k^j) \quad (1)$$

where,  $j$  is the  $j$ -th place, and each answer  $x_i$  contains a size of  $D$ .  $D$  represents the principles to be enhanced, while  $k$  represents an arbitrary answer ( $k \neq i$ ).  $\varphi_i^j$  shows a number randomly selected on the scale of  $[0,1]$ . Modifying a single element of  $x_i$ , the potentially novel solution  $v_i$  can be realized.

In the framework of a  $D$ -dimensional optimization process, a key strategy involves selectively altering the value of a randomly chosen dimension. Following each iteration, the selection of an improved solution is based on its 'athleticism' worth, a term used to signify its fitness or suitability within the context of the problem being solved. According to Formula 1, the newly generated solution  $v_i^j$  is dependent primarily on  $x_i^j$  and  $x_k^j$ , which ensures that the new food source  $v_i^j$  remains unpredictable and dynamic. This element of variability is crucial in preventing the algorithm from stagnating at local optima and aids in exploring a broader range of potential solutions. The study's approach in considering the concept of 'athleticism' is drawn from a collective understanding derived from both current and nearby nutrition resources. By utilizing shared knowledge and insights gathered from these resources, the algorithm continuously seeks to develop a food source with

a higher value. This method is in alignment with the principles of the ABC algorithm, where the objective is to discover and exploit resources that possess superior 'athleticism' worth, or in algorithmic terms, higher fitness values. Such an approach not only enhances the diversity of the solutions explored by the algorithm but also ensures that the process of optimization is dynamic and adaptive. By continuously updating the solution based on shared knowledge and the comparative worth of nearby resources, the algorithm can effectively navigate the solution space, moving towards more promising areas while avoiding less fruitful ones. This dynamic nature of solution generation and selection is integral to the success of the ABC algorithm in solving complex, multi-dimensional optimization problems.

$$v_i^j = \begin{cases} x_i^j + \varphi_i^j(x_k^j - x_i^j), & Fit_i < Fit_k \\ x_k^j + \varphi_i^j(x_i^j - x_k^j), & Fit_i \geq Fit_k \end{cases} \quad (2)$$

where,  $Fit_i$  and  $Fit_k$  represent the fitness values or the athleticism worth of the nearby and current food resources, respectively. The parameter  $\varphi_i^j$  is defined as a uniformly distributed random number within the range  $[0, F]$ , where  $F$ , a positive value, is known as the mutual learning factor. This factor plays a crucial role in the algorithm by guiding the fitness values of newly generated solutions towards superior food sources. It does this by evaluating and contrasting the current food resources with those in proximity. The candidate solution undergoes modification depending on the quality of the current food sources. If the existing food sources are found to be sufficiently rewarding, the solution will be further refined based on these sources. Conversely, if the current sources are deemed inadequate, the solution will shift towards the nearby, potentially more promising food source. This dynamic allows for a balanced approach between exploration of new possibilities and exploitation of known good solutions. The mutual learning factor  $F$  is pivotal in regulating the extent of perturbation between the positions of different food sources. A non-negative value of  $F$  is crucial to ensure that the resultant changes lead to an improved solution. As  $F$  increases from zero, the impact of perturbation on the corresponding food source diminishes, implying that the fitness value of the alternative resource is almost equal to that of the current, superior resource. Conversely, a higher value of  $F$  can reduce the algorithm's ability to effectively explore and exploit, as it lessens the impact of contrasting different food resources. Hence, the choice of a suitable value for  $F$  is vital in preserving a balanced interplay between exploration and exploitation. This balance is key to the ABC algorithm's ability to effectively navigate and identify optimal or near-optimal solutions in intricate optimization scenarios. The proper calibration of  $F$  ensures that the algorithm is neither overly explorative, risking inefficiency, nor excessively exploitative, which might lead to premature convergence. This careful tuning is essential for the algorithm's success in addressing complex optimization challenges efficiently.

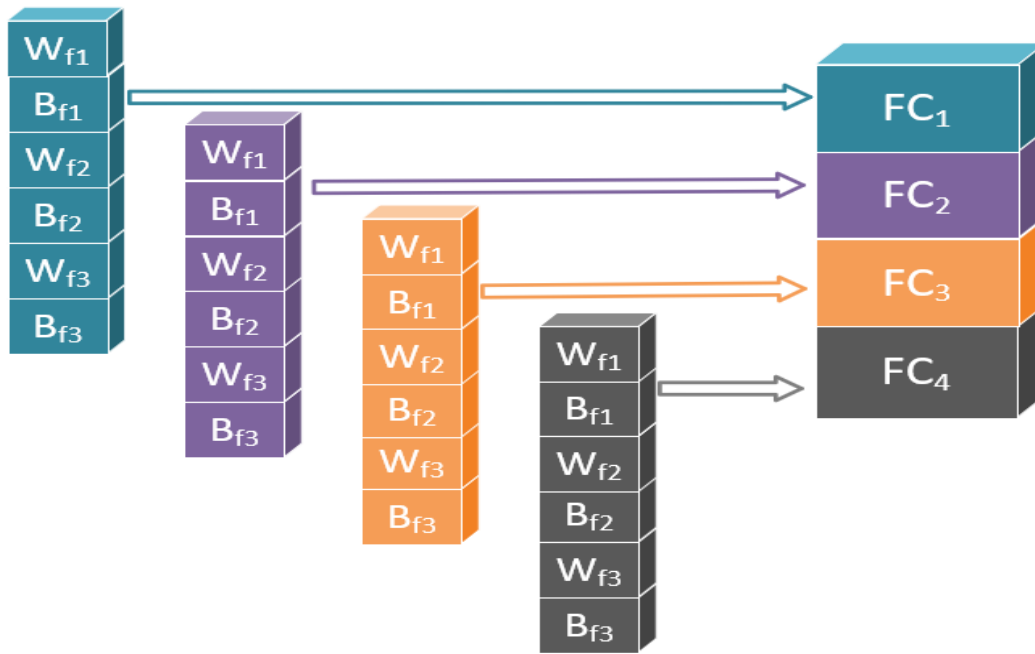


Fig. 2. Encoding approach in the offered method.

### B. Model Architecture

1) *Pre-Training*: In this particular scenario, the management of the MLP's weights is executed through an ABC algorithm, which is enhanced by the principles of mutual learning. The encoding procedure is a critical step that involves arranging these weights into a vector, a process that symbolically represents the positioning of bees within the ABC paradigm [42, 43]. This strategic alignment of weights into a vector format is not just a mere arrangement; it is conceptualized as mirroring the bees' placements in the ABC algorithm, thus establishing a direct correlation between the weights' configuration and the algorithm's operational mechanics. Achieving the optimal configuration for this encoding is a complex task that demands meticulous attention to detail. Despite the inherent challenges, a series of methodical experiments were conducted to ascertain the most effective encoding strategy, ensuring that the weights are optimally aligned for the ABC algorithm's functioning. This experimental approach was crucial in refining the encoding process, thereby enhancing the overall efficiency of the algorithm. As illustrated in Fig. 2, the comprehensive approach to this encoding process is evident. Here, not just the weights but also the bias terms are meticulously arranged into a vector. This arrangement is more than a mere collection of numerical values; it forms the foundation of a candidate solution within the ambit of the proposed ABC algorithm. This candidate solution, representative of the bees' locations in the ABC model, is instrumental in the algorithm's problem-solving process, highlighting the synergy between the MLP's weight management and the ABC algorithm's operational framework.

For assessing the caliber of a candidate solution, the fitness function is delineated as

$$Fitness = \frac{1}{\sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (3)$$

where,  $N$  is the training examples number, with  $\hat{y}_i$  and  $y_i$  indicating the  $i$ -th sample-estimated and goal outcome, respectively.

2) *Focal loss*: In this research, the detection problem is defined within the framework of binary classification, where instances are divided into positive and negative classes. A significant challenge faced is the imbalance in the dataset, particularly manifested in the limited number of samples representing the negative class. This imbalance can adversely affect the model's ability to learn from the underrepresented class. To address this issue, the study employs focal loss (FL) [44], a modified version of the binary cross-entropy (CE) loss function. FL is specifically engineered to refocus the training process on the more complex, less represented samples, which are often crucial for accurate classification but are overshadowed in imbalanced datasets [45]. FL achieves this by adding a modulating factor to the traditional cross-entropy loss, which adjusts the contribution of each sample to the loss based on the ease or difficulty of classifying it. This ensures that the model does not become biased towards the majority class and pays adequate attention to the minority class samples, which are pivotal for a balanced and comprehensive learning process. Such an approach is particularly beneficial in scenarios where the minority class is not just numerically fewer but also qualitatively more challenging to predict. By effectively targeting these challenging samples, FL aids in enhancing the overall robustness and accuracy of the model in

dealing with binary classification tasks under imbalanced conditions.

CE is conceptualized as:

$$CE = \begin{cases} -\log(p), & y = 1 \\ -\log(1-p), & \text{otherwise} \end{cases} \quad (4)$$

where,  $y \in \{-1,1\}$  represents the actual class label, while  $p \in [0,1]$  denotes the model's predicted probability for the class that corresponds to the label  $y = 1$ . The predicted probability is

$$p_t = \begin{cases} -p, & y = 1 \\ 1-p, & \text{otherwise} \end{cases} \quad (5)$$

Hence,

$$CE(p, y) = CE(p_t) = -\log(p_t) \quad (6)$$

FL introduces a modulating factor to the standard cross-entropy loss, resulting in the following modification:

$$FL(p_t) = -\alpha_t(1-p_t)^\gamma \log(p_t) \quad (7)$$

where,  $\gamma > 0$  is a positive parameter (notably, when  $\gamma=1$ , FL resembles the CE loss), and  $\alpha$  lies within the range of 0 to 1, representing the inverse of the class frequency.

#### IV. EXPERIMENTAL RESULTS

In 2012, healthcare practitioners at the Polyclinic for Occupational Health and Sports in Zagreb assembled a comprehensive dataset from 26,002 medical examinations. These examinations were conducted on athletes seeking medical clearance for participation in competitive sports. The collected data encompassed a wide range of vital health parameters for each athlete, including their sex, age, height, weight, resting pulse rate, as well as both diastolic and systolic blood pressure measurements. Additionally, resting electrocardiogram (ECG) data were meticulously recorded for each individual. The dataset revealed a significant imbalance in the classification of results, with a dominant majority, 91.2%, being categorized as N (negative), indicating no apparent risk or health concern that would preclude sports participation.

Conversely, a smaller fraction, 8.8%, fell into the P (positive) category, suggesting potential health risks or conditions that required further medical evaluation and could potentially restrict the athlete's ability to engage in competitive sports. This disproportionate distribution between the N and P classifications underscores the challenges faced in medical diagnostics, particularly in accurately identifying and diagnosing conditions in smaller, potentially high-risk groups. The dataset, therefore, provides an invaluable resource for developing and testing medical diagnostic models that can effectively manage such imbalances, ensuring that high-risk cases are accurately identified and addressed, despite being numerically fewer in the dataset. This approach is crucial in enhancing the safety and health management of athletes, aligning with the overarching goal of ensuring their well-being and fitness for competitive sports participation.

The model put forth operates on a 64-bit Windows OS, supported by a robust 64 GB of RAM, and is further enhanced by a graphics processing unit (GPU) with a capacity of 64 GB. Table I presents the hyperparameters applied to the MLP-FL-CRD model.

The MLP-FL-CRD model is first trained and tested on the introduced dataset in parallel with 6 computer learning samples (SVM [46], Naïve Bayes [47], KNN [48], Random forests [49], Logistic Regression [50], and Decision tree [51]) and two smaller parts of the proposed model, i.e., Proposed+ random weights (which possesses a base architecture similar to the model but uses random weights instead for initialization), and Proposed+ random weights+ FL (which uses FL for classification). Table II shows the parameters applied to machine learning models.

TABLE I. PARAMETER SETTING FOR THE MLP-FL-CRD MODEL

Parameter	Value
Epoch	256
Batch size	128
Learning rate	0.03
Discount factor	0.4

TABLE II. PARAMETER SETTING FOR MACHINE LEARNING MODELS

Algorithm	Parameter	Value
Naïve Bayes	$\alpha$ (Lidstone smoothing parameter)	0.5
KNN	$k$ (Number of neighbors)	5
	Distance Metric	Euclidean ( $p=2$ ), Manhattan ( $p=1$ )
SVM	Kernel	polynomial
	$\gamma$ (Kernel coefficient)	0.5
Random Forests	Number of trees	20
	Max depth of tree	10
Logistic Regression	C (Inverse regularization strength)	0.3
	Solver	liblinear
Decision Tree	Criterion	entropy
	Max depth	10

The results are compared using standard performance metrics (see Table III), of which F-measure and geometric mean are the preferred metrics for imbalanced data [52]. The MLP-FL-CRD model outperforms other models, including the closest competitor, the Decision Tree, across all evaluation metrics. Specifically, this model achieves a reduction in error exceeding 59% and 34% for two primary metrics, namely the F-measure and G-means, respectively. A comparative analysis of the MLP-FL-CRD model against variants such as Proposed+random weights and Proposed+random weights+FL reveals a notable decrease in error rates, approximately 70%. This significant reduction underscores the effectiveness of the enhanced ABC and FL methodologies.

### A. Impact of other Metaheuristics

The next experiment involved comparing the improved ABC algorithm with various metaheuristic optimization algorithms. In this experiment, the initial model parameters were obtained using different metaheuristics while retaining the other model components. Six algorithms were tested, including standard ABC [53], FA [54], BA [55], COA [56], DE [28], and GWO [57]. The default configurations are detailed in Table IV. The results obtained are presented in Table V. The findings indicated that the suggested ABC algorithm reduced the error by approximately 48% compared to the standard ABC algorithm. This result showed that the proposed model outperformed the standard one. Furthermore, the ABC algorithm delivered better results than others, including DE, GWO, and BA.

TABLE III. OUTCOMES OF DIVERSE CATEGORIZATION METHODS

	accuracy	recall	precision	F-measure	G-means
SVM	0.687± 0.051	0.594± 0.068	0.540± 0.125	0.566± 0.068	0.661± 0.005
Naïve Bayes	0.820± 0.159	0.743± 0.004	0.772± 0.105	0.757± 0.109	0.806± 0.045
KNN	0.790± 0.104	0.667± 0.005	0.704± 0.048	0.685± 0.215	0.754± 0.006
Random forests	0.695± 0.105	0.569± 0.226	0.554± 0.103	0.561± 0.100	0.658± 0.256
Logistic Regression	0.818± 0.145	0.781± 0.048	0.715± 0.106	0.747± 0.148	0.809± 0.125
Decision tree	0.844± 0.146	0.825± 0.105	0.765± 0.275	0.804± 0.205	0.830± 0.014
Proposed + random weights	0.800± 0.052	0.810± 0.100	0.791± 0.000	0.781± 0.152	0.820± 0.252
Proposed + random weights+FL	0.851± 0.014	0.860± 0.115	0.840± 0.259	0.840± 0.041	0.850± 0.082
MLP-FL-CRD	0.878± 0.028	0.892± 0.103	0.870± 0.021	0.871± 0.019	0.892± 0.032

TABLE IV. PARAMETER SETTING FOR METAHEURISTIC MODELS

Algorithm	Parameter	Value
DE	scaling ratio	0.6
	probability of crossover	0.8
ABC	limit	$n_e \times \text{dimensionality}$
	$n_o$	50% of the colony
	$n_e$	50% of the colony
	$n_s$	1
FA	coefficient of light absorption	1
	initial attractiveness at $t = 0$	0.3
	proportional factor	0.4
BA	update constant for loudness	0.45
	update constant for rate of emission	0.45
	initial rate of pulse emission	0.002
COA	rate of alien solution discovery	0.30
GWO	no parameters	-

TABLE V. RESULTS OF VARIOUS METAHEURISTIC METHODS

	accuracy	recall	precision	F-measure	G-means
ABC	0.852± 0.148	0.840± 0.123	0.859± 0.051	0.841± 0.009	0.821± 0.412
DE	0.841± 0.158	0.831± 0.103	0.842± 0.251	0.830± 0.015	0.800± 0.025
FA	0.827± 0.015	0.815± 0.123	0.810± 0.261	0.819± 0.071	0.780± 0.014
BA	0.810± 0.032	0.800± 0.014	0.801± 0.071	0.805± 0.132	0.762± 0.125
COA	0.792± 0.123	0.772± 0.016	0.786± 0.274	0.772± 0.171	0.741± 0.156
GWO	0.740± 0.152	0.724± 0.015	0.740± 0.012	0.731± 0.223	0.690± 0.126



### B. Impact of parameter $F$ on the model

The performance of the suggested approach is heavily influenced by the mutual learning factor  $F$ , as expressed in Eq. (2). When  $F$  is set too low, the algorithm may not be able to take full advantage of the mutual learning process, and thus, the performance of the algorithm may suffer. However, if  $F$  is excessively increased, the algorithm may become too ambitious in its learning, resulting in overfitting and poor generalization. Therefore, it is essential to find the optimal value of  $F$  those balances between the benefits of mutual learning and the risks of overfitting. As shown in Figure 3, the performance of the algorithm improves significantly as  $F$  increases from 0.5 to 2.5. This is because a higher value of  $F$  allows the models to exchange more information, leading to better generalization and higher accuracy. However, when  $F$  is increased beyond 2.5, the performance of the algorithm starts to deteriorate. This is because the models become too aggressive in their learning and start to overfit the data. Overall, these results demonstrate that the mutual learning factor  $F$  is a critical parameter in the proposed approach, and its value should be carefully chosen to achieve optimal performance. A moderate value of  $F$  between 1.5 and 2.5 may be a good starting point, and the optimal value can be determined through experimentation and cross-validation.

### C. Exploring the Number of MLP Layers

The article highlights that as the number of layers in a multi-layer perceptron (MLP) increase, the model's complexity increases, leading to a higher risk of overfitting. Conversely, having too few layers may limit the model's ability to represent essential features in the training data. In the proposed approach, six different values (1, 2, 4, 8, 10, 12) are tested as the number of layers in MLP to study its effect on the model's

performance. Table VI displays the achieved results, which show a descending trend for the number of layers from 1 to 4 and an ascending trend for values from 4 to 12. This finding suggests that having four layers in MLP is the optimal value for achieving the best results.

### D. Impact of Loss Function

There are several methods to address data imbalances in machine learning models, including arranging info-augmenting strategies and the choice of the dropping operation or LF. Along with the methods, the choice of the dropping operation or LF is particularly critical since it can help the model learn from the minority class. To test the effectiveness of different loss functions, five functions were chosen, including weighted cross-entropy (WCE) [58], balanced cross-entropy (BCE) [59], Dice loss (DL) [60], Tversky loss (TL) [61], and Combo Loss (CL) [62]. The BCE and WCE loss functions are commonly used to treat both positive and negative examples equally. However, these loss functions may not be suitable for imbalanced datasets where the minority class needs to be emphasized. The DL and TL loss functions are more suitable for imbalanced datasets, as they perform better on the minority class. The CL function is a promising loss function that can benefit applications using unbalanced data. The CL method can decrease the importance of straightforward examples and emphasize learning intricate instances by modifying the weights of the loss function. To evaluate the effectiveness of these loss functions, experiments were conducted and the results are reported in Table VII. According to the findings, the performance of the CL function was better than that of the TL function. It reduced the error rate by 39% and 25% for the F-measure and accuracy metrics in order. However, the CL function operates 60% worse than the FL, which is a specialized loss function for binary classification tasks.

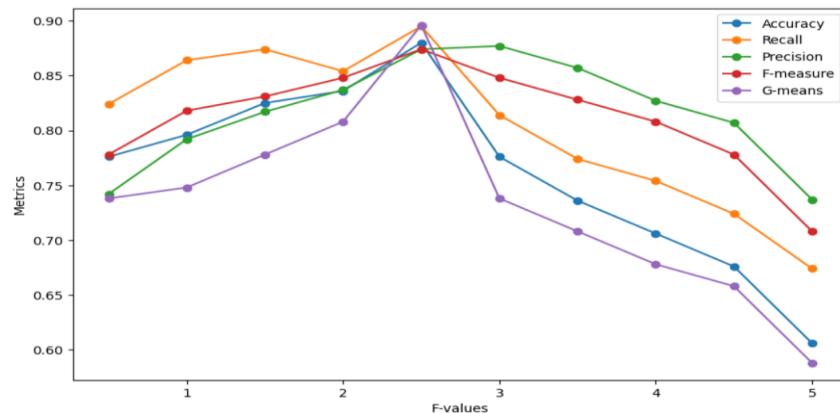


Fig. 3. The performance metrics of the MLP-FL-CRD model in relation to the  $F$  value of the proposed model.

TABLE VI. THE DIVERSE NUMBER OF MLP LAYERS VS. THE EFFICIENCY MEASUREMENTS PLOTTED

Number of layers	accuracy	recall	precision	F-measure	G-means
1	0.750± 0.019	0.772± 0.203	0.731± 0.025	0.732± 0.034	0.763± 0.102
2	0.862± 0.008	0.870± 0.153	0.852± 0.005	0.857± 0.029	0.871± 0.152
4	0.880± 0.028	0.895± 0.103	0.874± 0.021	0.874± 0.019	0.896± 0.032
8	0.842± 0.018	0.851± 0.101	0.840± 0.221	0.820± 0.010	0.852± 0.022
10	0.704± 0.125	0.720± 0.111	0.682± 0.201	0.700± 0.015	0.650± 0.032
12	0.480± 0.055	0.592± 0.021	0.490± 0.221	0.526± 0.002	0.420± 0.152

TABLE VII. OUTCOMES OF DIFFERENT LFS

	accuracy	recall	precision	F-measure	G-means
WCE	0.750± 0.026	0.742± 0.010	0.731± 0.119	0.736± 0.002	0.762± 0.032
BCE	0.801± 0.021	0.793± 0.045	0.772± 0.156	0.770± 0.012	0.811± 0.001
DL	0.812± 0.033	0.801± 0.025	0.780± 0.026	0.798± 0.006	0.820± 0.000
TL	0.821± 0.123	0.824± 0.012	0.800± 0.016	0.813± 0.036	0.840± 0.062
CL	0.861± 0.003	0.852± 0.215	0.840± 0.006	0.840± 0.042	0.863± 0.152

### E. Discussion

This article proposed a novel approach to identifying athletes at risk of developing CVDs by utilizing a multi-layer perceptron model enhanced with focal loss and a ML-ABC optimization technique. This model specifically addresses the challenge of class imbalance in datasets through focal loss, which effectively de-emphasizes simpler cases to focus on more complex ones, thus improving diagnostic accuracy in scenarios with significant class disparities. Additionally, the article introduces the ML-ABC method as a solution to overcome the limitations of traditional gradient-based learning methods, such as sensitivity to initial weight settings. By adjusting candidate food sources based on mutual learning factors and considering initial weights, the model shows improved performance. The effectiveness of this approach is validated through extensive experiments and ablation studies, demonstrating its superiority over other models in accurately assessing the risk of CVDs among athletes.

The susceptibility of the MLP-FL-CRD model to class imbalance, even with the integration of FL, presents a notable constraint in accurately diagnosing cardiovascular risk among athletes. In the referenced 2012 dataset, which included data from 26,002 athletes, there was a pronounced imbalance, predominantly skewed towards negative outcomes. While the implementation of focal loss aims to accentuate the learning from minority class examples, the preponderance of negative results can still hinder the model's ability to generalize across varied scenarios effectively. This imbalance in the dataset could induce a predictive bias within the model, potentially leading to an underestimation of risk in cases that are actually positive but less frequent. Such a bias is particularly concerning in a medical context, where failing to identify at-risk individuals can have serious, if not fatal, consequences. The model's overexposure to negative outcomes might condition it to lean towards these predictions, potentially missing out on identifying athletes who are genuinely at risk of cardiovascular issues. To address this limitation, additional strategies may need to be considered. One approach could involve incorporating more sophisticated balancing techniques that go beyond focal loss, such as synthetic data generation methods like SMOTE (Synthetic Minority Over-sampling Technique) [63] or adaptive resampling [64]. These methods can help in creating a more balanced training environment for the model, thus enhancing its capacity to learn from both majority and minority classes more effectively [65]. Another potential solution lies in expanding and diversifying the dataset. Gathering more comprehensive data that includes a wider range of cardiovascular conditions and outcomes could help in creating a more representative dataset. This expanded dataset would not only provide a broader spectrum of cases for the model to learn from but also reduce the likelihood of

predictive bias, thereby improving the model's accuracy and reliability in real-world applications.

The dependency of the model's performance on the initial weight settings is a critical aspect, particularly prevalent in gradient-based learning methods such as backpropagation. While the model employs a ML-ABC methodology to mitigate this issue by dynamically adjusting the primary weights according to fitness, the inherent reliance on initial weights remains a significant vulnerability. Inadequate calibration of these initial weights could lead to the model converging towards suboptimal solutions, thereby affecting its overall effectiveness and accuracy. The initial weight setting acts as the starting point for the learning process and heavily influences the trajectory of the model's convergence. If these weights are not set in a manner that reflects the complexity and nuances of the data, the model may find itself trapped in local minima, or on a prolonged path towards the global optimum. This challenge is further amplified in the context of complex and high-dimensional data typically encountered in cardiovascular risk assessment. To enhance the resilience of the model against potential pitfalls associated with initial weight selection, exploring alternative strategies for initialization is imperative. One such approach could involve the use of advanced heuristics or algorithms that analyze the data distribution to determine more effective starting weights. Techniques like Xavier or He initialization [66], which consider the size of the network layers in setting the initial weights, could offer more reliable starting points for the model's training. Additionally, incorporating more robust weight optimization methods could further strengthen the model. Techniques like stochastic gradient descent with momentum or adaptive learning rate algorithms like Adam could provide more nuanced adjustments during the training process. These methods help in navigating the weight space more effectively, increasing the likelihood of the model finding a more optimal solution.

The success of the model in the specific context of the study dataset focusing on athletes' cardiovascular risk assessment does not automatically translate to its efficacy in other scenarios or datasets. This limitation in generalizability and adaptability poses a significant challenge, especially considering the diverse nature of CVDs and the varying characteristics of different patient populations. The model's parameters, which have been fine-tuned for this particular dataset, may not be directly applicable or optimal for other datasets that differ in demographics, prevalence of CVD types, or other clinical factors. When transitioning the model to different populations or conditions, it may encounter data that significantly deviates from the characteristics of the original dataset. This deviation can result in reduced accuracy and reliability, as the model's learned patterns and parameters may

not align well with the new data. Consequently, substantial re-tuning of the model's parameters becomes necessary, along with rigorous validation processes to ensure its efficacy in the new context. This re-tuning process can be resource-intensive and time-consuming, requiring a thorough understanding of the new dataset's attributes and underlying distributions. To address these challenges, developing more flexible learning algorithms that can easily adapt to varying data characteristics is essential. Such algorithms should be capable of identifying and adjusting to the nuances of different datasets without extensive manual intervention. This flexibility can be achieved through approaches like meta-learning, where the model learns to quickly adapt to new tasks using only a small amount of data, or through the development of models that are inherently more robust to changes in data distribution. Moreover, incorporating transfer learning techniques can significantly enhance the adaptability of the model. Transfer learning allows a model trained on one task to apply its learned knowledge to a different but related task. By leveraging pre-trained models or transferring knowledge from the original dataset to new ones, the model can achieve better performance with less need for re-tuning. This approach can be particularly beneficial in medical applications like CVD risk assessment, where similarities exist across different datasets, even though they may vary in specific characteristics.

## V. CONCLUSION

CVDs remain a significant global health issue, and identifying individuals at risk of developing CVDs is crucial for early intervention and effective treatment. To this end, the article presents a model based on a multi-layer perceptron that employs focal loss to identify athletes who may be at risk of developing CVDs. Focal loss is a useful technique that reduces the significance of straightforward instances, enabling the model to concentrate on more difficult instances. This technique is particularly useful when dealing with a large disparity between classes. Usually, the training process of the model relies on gradient-based learning methods like backpropagation. These techniques have several limitations, including initialization sensitivity. The article suggests a solution for this problem, which involves utilizing ML-ABC. This approach involves modifying the candidate food source with higher fitness between two individuals using a mutual learning factor. This method takes into account the initial weights of the model and can improve its performance. The offered sample performs better than the rest of the samples, achieving excellent results. Experiments conducted on the study dataset help to determine the ideal values of the critical parameters in the model. Ablation studies further confirm the positive impact of the proposed components on model performance.

Despite these promising results, the article acknowledges the need for further research to test the efficacy of this model in non-athletic and older populations. Additionally, it is necessary to determine an appropriate classification performance measure that balances medical risk and welfare. This research is crucial to ensure that the proposed data mining methods can be effectively applied to improve healthcare policies and reduce unnecessary examinations. The potential impact of this research is significant. By identifying individuals at risk of

developing CVD early, healthcare professionals can intervene with targeted prevention strategies and treatments. This could ultimately lead to improved health outcomes for individuals and reduced healthcare costs for society.

Furthermore, the methodologies integrated into this model hold potential for application beyond the realm of sports medicine, offering opportunities to enhance disease detection and management across various healthcare sectors. The adaptability of these techniques could play a pivotal role in refining diagnostic processes and treatment strategies for a range of medical conditions, thereby contributing to the overall advancement of healthcare practices. To summarize, the research detailed in this article introduces a robust model adept at identifying athletes who are at an increased risk of developing CVDs. Its effectiveness, as demonstrated in the specific context of sports healthcare, offers valuable insights into CVD risk assessment. However, it is imperative to extend this research to encompass a more diverse range of populations. Such expansion is essential to fully ascertain the model's efficacy across different demographic and health profiles. Developing and fine-tuning classification performance metrics tailored to varied populations will also be crucial in this regard. While the model shows considerable promise, its broader applicability and effectiveness in general healthcare settings remain areas for future investigation. Pursuing this line of research is vital for realizing the model's full potential in enhancing early detection and intervention strategies for CVDs. Ultimately, the advancements in early identification and treatment strategies for CVDs, as suggested by this research, could lead to improved patient health outcomes and a reduction in healthcare costs. This underscores the significance of this research as a meaningful contribution to the field of medical diagnostics and patient care.

## REFERENCES

- [1] D. Barbieri et al., "Predicting cardiovascular risk in Athletes: Resampling improves classification performance," *International Journal of Environmental Research and Public Health*, vol. 17, no. 21, p. 7923, 2020.
- [2] I. Wronka, E. Suliga, and R. Pawlińska-Chmara, "Evaluation of lifestyle of underweight, normal weight and overweight young women," *Collegium antropologicum*, vol. 37, no. 2, pp. 359-365, 2013.
- [3] 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.
- [4] C. Campbell and K. Bennett, "A linear programming approach to novelty detection," *Advances in neural information processing systems*, vol. 13, 2000.
- [5] 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.
- [6] 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.
- [7] L. Sacchi, A. Tucker, S. Counsell, D. Garway-Heath, and S. Swift, "Improving predictive models of glaucoma severity by incorporating quality indicators," *Artificial intelligence in medicine*, vol. 60, no. 2, pp. 103-112, 2014.
- [8] F. Masood, J. Masood, H. Zahir, K. Driss, N. Mehmood, and H. Farooq, "Novel approach to evaluate classification algorithms and feature selection filter algorithms using medical data," *Journal of Computational and Cognitive Engineering*, vol. 2, no. 1, pp. 57-67, 2023.

- [9] E. H. Houssein, M. E. Hosney, W. M. Mohamed, A. A. Ali, and E. M. Younis, "Fuzzy-based hunger games search algorithm for global optimization and feature selection using medical data," *Neural Computing and Applications*, vol. 35, no. 7, pp. 5251-5275, 2023.
- [10] S. V. Moravvej et al., "RLMD-PA: A reinforcement learning-based myocarditis diagnosis combined with a population-based algorithm for pretraining weights," *Contrast Media & Molecular Imaging*, vol. 2022, 2022.
- [11] W. Wang et al., "Medical image classification using deep learning," *Deep learning in healthcare: paradigms and applications*, pp. 33-51, 2020.
- [12] 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)*, 2022: IEEE, pp. 1-7.
- [13] S. Zhang, C. Tjortjris, X. Zeng, H. Qiao, I. Buchan, and J. Keane, "Comparing Data Mining Methods with Logistic Regression."
- [14] 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.
- [15] M. Bahadori, M. Soltani, M. Soleimani, and M. Bahadori, "Statistical Modeling in Healthcare: Shaping the Future of Medical Research and Healthcare Delivery," in *AI and IoT-Based Technologies for Precision Medicine: IGI Global*, 2023, pp. 431-446.
- [16] L. Hong et al., "GAN - LSTM - 3D: An efficient method for lung tumour 3D reconstruction enhanced by attention - based LSTM," *CAAI Transactions on Intelligence Technology*, 2023.
- [17] M. Soleimani, Z. Forouzanfar, M. Soltani, and M. J. Harandi, "Imbalanced Multiclass Medical Data Classification based on Learning Automata and Neural Network," *EAI Endorsed Transactions on AI and Robotics*, vol. 2, 2023.
- [18] S. V. Moravvej, A. Mirzaei, and M. Safayani, "Biomedical text summarization using conditional generative adversarial network (CGAN)," *arXiv preprint arXiv:2110.11870*, 2021.
- [19] Z. Ahmed and S. Das, "A Comparative Analysis on Recent Methods for Addressing Imbalance Classification," *SN Computer Science*, vol. 5, no. 1, pp. 1-18, 2024.
- [20] 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)*, 2022: IEEE, pp. 000265-000270.
- [21] C. Huang, Y. Li, C. C. Loy, and X. Tang, "Learning deep representation for imbalanced classification," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 5375-5384.
- [22] Y. Yan, M. Chen, M.-L. Shyu, and S.-C. Chen, "Deep learning for imbalanced multimedia data classification," in *2015 IEEE international symposium on multimedia (ISM)*, 2015: IEEE, pp. 483-488.
- [23] S. Moravvej, M. 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.
- [24] M. Marani, M. Soltani, M. Bahadori, M. Soleimani, and A. Moshayedi, "The Role of Biometric in Banking: A Review," *EAI Endorsed Transactions on AI and Robotics*, vol. 2, no. 1, 2023.
- [25] 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)*, 2021: IEEE, pp. 1-5.
- [26] 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.
- [27] M. Jaderberg et al., "Population based training of neural networks," *arXiv preprint arXiv:1711.09846*, 2017.
- [28] S. J. Mousavirad, D. Oliva, S. Hinojosa, and G. Schaefer, "Differential evolution-based neural network training incorporating a centroid-based strategy and dynamic opposition-based learning," in *2021 IEEE Congress on Evolutionary Computation (CEC)*, 2021: IEEE, pp. 1233-1240.
- [29] D. Karaboga, B. Akay, and C. Ozturk, "Artificial bee colony (ABC) optimization algorithm for training feed-forward neural networks," in *Modeling Decisions for Artificial Intelligence: 4th International Conference, MDAI 2007, Kitakyushu, Japan, August 16-18, 2007. Proceedings 4*, 2007: Springer, pp. 318-329.
- [30] 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," *Computational Intelligence and Neuroscience*, vol. 2022, 2022.
- [31] 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)*, 2021: IEEE, pp. 238-242.
- [32] D. Waigi, D. S. Choudhary, D. P. Fulzele, and D. Mishra, "Predicting the risk of heart disease using advanced machine learning approach," *Eur. J. Mol. Clin. Med*, vol. 7, no. 7, pp. 1638-1645, 2020.
- [33] K. Vanisree and J. Singaraju, "Decision support system for congenital heart disease diagnosis based on signs and symptoms using neural networks," *International Journal of computer applications*, vol. 19, no. 6, pp. 6-12, 2011.
- [34] A. Narin, Y. İşler, and M. Özer, "Early prediction of Paroxysmal Atrial Fibrillation using frequency domain measures of heart rate variability," in *2016 Medical Technologies National Congress (TIPTEKNO)*, 2016: IEEE, pp. 1-4.
- [35] D. Shah, S. Patel, and S. K. Bharti, "Heart disease prediction using machine learning techniques," *SN Computer Science*, vol. 1, pp. 1-6, 2020.
- [36] K. Drożdż et al., "Risk factors for cardiovascular disease in patients with metabolic-associated fatty liver disease: a machine learning approach," *Cardiovascular Diabetology*, vol. 21, no. 1, p. 240, 2022.
- [37] F. S. Alotaibi, "Implementation of machine learning model to predict heart failure disease," *International Journal of Advanced Computer Science and Applications*, vol. 10, no. 6, 2019.
- [38] N. Hasan and Y. Bao, "Comparing different feature selection algorithms for cardiovascular disease prediction," *Health and Technology*, vol. 11, pp. 49-62, 2021.
- [39] S. Mohan, C. Thirumalai, and G. Srivastava, "Effective heart disease prediction using hybrid machine learning techniques," *IEEE access*, vol. 7, pp. 81542-81554, 2019.
- [40] 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.
- [41] S. Vakilian, S. V. Moravvej, and A. Faniyan, "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)*, 2021: IEEE, pp. 509-513.
- [42] S. Vakilian, S. V. Moravvej, and A. Faniyan, "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)*, 2021: IEEE, pp. 1-5.
- [43] 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)*, 2021: IEEE, pp. 460-464.
- [44] D. Sarkar, A. Narang, and S. Rai, "Fed-focal loss for imbalanced data classification in federated learning," *arXiv preprint arXiv:2011.06283*, 2020.
- [45] S. K. Prabhakar, H. Rajaguru, and D.-O. Won, "Performance Analysis of Hybrid Deep Learning Models with Attention Mechanism Positioning and Focal Loss for Text Classification," *Scientific Programming*, vol. 2021, pp. 1-12, 2021.

- [46] M. A. de Almeida, "DATA MINING: DETERMINAC AO DE AGRUPAMENTOS EM GRANDES BASES DE DADOS," Universidade Federal do Rio de Janeiro, 2013.
- [47] G. I. Webb, E. Keogh, and R. Miikkulainen, "Naïve Bayes," *Encyclopedia of machine learning*, vol. 15, pp. 713-714, 2010.
- [48] G. Guo, H. Wang, D. Bell, Y. Bi, and K. Greer, "KNN model-based approach in classification," in *On The Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE: OTM Confederated International Conferences, CoopIS, DOA, and ODBASE 2003*, Catania, Sicily, Italy, November 3-7, 2003. *Proceedings, 2003*: Springer, pp. 986-996.
- [49] L. Breiman, "Random forests," *Machine learning*, vol. 45, pp. 5-32, 2001.
- [50] M. P. LaValley, "Logistic regression," *Circulation*, vol. 117, no. 18, pp. 2395-2399, 2008.
- [51] A. J. Myles, R. N. Feudale, Y. Liu, N. A. Woody, and S. D. Brown, "An introduction to decision tree modeling," *Journal of Chemometrics: A Journal of the Chemometrics Society*, vol. 18, no. 6, pp. 275-285, 2004.
- [52] M. Bekkar, H. K. Djemaa, and T. A. Alitouche, "Evaluation measures for models assessment over imbalanced data sets," *J Inf Eng Appl*, vol. 3, no. 10, 2013.
- [53] D. Karaboga and B. Basturk, "On the performance of artificial bee colony (ABC) algorithm," *Applied soft computing*, vol. 8, no. 1, pp. 687-697, 2008.
- [54] 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.
- [55] X.-S. Yang, "A new metaheuristic bat-inspired algorithm," *Nature inspired cooperative strategies for optimization (NICSO 2010)*, pp. 65-74, 2010.
- [56] X.-S. Yang and S. Deb, "Cuckoo search via Lévy flights," in *2009 World congress on nature & biologically inspired computing (NaBIC), 2009*: Ieee, pp. 210-214.
- [57] S. Mirjalili, S. M. Mirjalili, and A. Lewis, "Grey wolf optimizer," *Advances in engineering software*, vol. 69, pp. 46-61, 2014.
- [58] Ö. Özdemir and E. B. Sönmez, "Weighted cross-entropy for unbalanced data with application on covid x-ray images," in *2020 Innovations in Intelligent Systems and Applications Conference (ASYU)*, 2020: IEEE, pp. 1-6.
- [59] F. Huang, J. Li, and X. Zhu, "Balanced Symmetric Cross Entropy for Large Scale Imbalanced and Noisy Data," *arXiv preprint arXiv:2007.01618*, 2020.
- [60] X. Li, X. Sun, Y. Meng, J. Liang, F. Wu, and J. Li, "Dice loss for data-imbalanced NLP tasks," *arXiv preprint arXiv:1911.02855*, 2019.
- [61] S. S. M. Salehi, D. Erdogmus, and A. Gholipour, "Tversky loss function for image segmentation using 3D fully convolutional deep networks," in *Machine Learning in Medical Imaging: 8th International Workshop, MLMI 2017, Held in Conjunction with MICCAI 2017, Quebec City, QC, Canada, September 10, 2017, Proceedings 8*, 2017: Springer, pp. 379-387.
- [62] S. A. Taghanaki et al., "Combo loss: Handling input and output imbalance in multi-organ segmentation," *Computerized Medical Imaging and Graphics*, vol. 75, pp. 24-33, 2019.
- [63] 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.
- [64] C. Kuptamete, Z.-H. Michalopoulou, and N. Aunsri, "Adaptive genetic algorithm-based particle herding scheme for mitigating particle impoverishment," *Measurement*, vol. 214, p. 112785, 2023.
- [65] 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*, 2021: Springer, pp. 690-701.
- [66] J. Ye, Z. Zhu, F. Liu, R. Shokri, and V. Cevher, "Initialization Matters: Privacy-Utility Analysis of Overparameterized Neural Networks," *arXiv preprint arXiv:2310.20579*, 2023.