2020 17th International Conference on Electrical Engineering, Computing Science and Automatic Control (CCE). Mexico City, Mexico. November 11-13, 2020

Artificial Neural Network for Classification of Possible Cardiovascular Risk Using Indexes of Heart Rate Variability

Gabriel Vega Martínez
Sección Bioelectrónica
Centro de Investigación y Estudios Avanzados del Instituto Politécnico Nacional
CDMX, México
ORCID 0000-0002-3984-7124

Diego Mirabent Amor
CENIAMED
Instituto Nacional de Rehabilitación "Luis Guillermo Ibarra Ibarra"
CDMX, México
diegomirabent@outlook.com

J. Gilberto Franco Sánchez
CENIAMED
Instituto Nacional de Rehabilitación "Luis Guillermo Ibarra Ibarra"
CDMX, México
jfranco@inr.gob.mx

Francisco José Ramos Becerrril
Rehabilitación Cardiaca
Instituto Nacional de Rehabilitación "Luis Guillermo Ibarra Ibarra"
CDMX, México
ORCID 0000-0003-2947-5678

Carlos Alvarado Serrano
Sección Bioelectrónica
Centro de Investigación y Estudios Avanzados del Instituto Politécnico Nacional
CDMX, México
ORCID 0000-0003-4835-6906

Arturo Vera Hernández
Sección Bioelectrónica
Centro de Investigación y Estudios Avanzados del Instituto Politécnico Nacional
CDMX, México
ORCID 0000-0001-6258-154X

Lorenzo Leija Salas
Sección Bioelectrónica
Centro de Investigación y Estudios Avanzados del Instituto Politécnico Nacional
CDMX, México
ORCID 0000-0001-8437-6520

***Abstract*— Prevention is a key factor to avoid chronic diseases and premature death. The assessment of risk of cardiovascular disease through the Framingham score could help in taking action on time. This paper proposes the use of an Artificial Neural Network as a classifier of risk based on the indexes of Heart Rate Variability. 60 electrocardiographic records from the database of the PhysioBC® project are used to calculate time domain, frequency domain and nonlinear indexes. These parameters, in addition to age and body mass index, will be used to classify 4 levels of risk. These levels are established using the Framingham score. The proposed architecture has a training efficiency of 91.7 %, 100 % using test vectors and 95 % with validation vectors.**

Keywords—Heart Rate Variability, Framingham Risk Score, Artificial Neural Network, Wavelet Transform

I. INTRODUCTION

Current sedentary lifestyles are having a major impact on people health and are precursors to chronic diseases and premature death [1]. The world is currently undergoing a global health crisis where isolation and confinement is necessary, causing psychological and physical changes in people with implications for their health states [2]. Prevention plays an important role and must be part of government policies. Childhood obesity is already an epidemic, being a population with high risk of developing cardiovascular diseases [3].

In the specialty of cardiology, it has been established the need to implement a subspecialty more focused on prevention than treatment [4]. There are different methods to know the risks that a person may have in the appearance and progression of a disease such as: genetic markers, epigenetic markers, the environment, and cultural issues. Knowing them is important but quantifying them is an even bigger task. The consumption of tobacco, alcohol, obesity, physical inactivity and weight loss without reason are reported as risk of cardiovascular mortality, the number one cause in the world [5].

A great impact research is the study of the heart of Framingham [6], which was the basis for the development of an algorithm that allows stratifying, expressed as a percentage, the 30-year risk probability of presenting cardiovascular disease. There are two models; the first uses sex, age, blood pressure, hypertension, smoker, diabetic, HDL and total cholesterol. The second model replaces the lipid parameters with the body mass index (BMI). New proposals seek a more personalized approach, use more parameters of the clinical and laboratory aspects, and also consider physical activity and how healthy the diet is [7].

An electrocardiogram (ECG) is an instrument of cardiac electrophysiology for the diagnosis of cardiovascular diseases. The diagnostic information is obtained by studying the morphological changes and the time that waves, segments and intervals that make up the ECG record may present. Processing alternatives such as the analysis of Heart Rate Variability (HRV) evolve to study the control of the autonomic nervous system (ANS) in its sympathetic and parasympathetic branch and how they manifest as changes in the duration of the heartbeat. The HRV indexes have been used for the prevention of fatigue in runners [8], with the quality of sleep and its impact on sports performance [9], with more clinical issues such as its relationship with inflammatory processes, the primary mechanism of defense of the organism against infections [10] and with greater impact in its use as a predictor of sudden cardiac death [11].

The funding for the development of the work was provided by CYTED-DITECROD-218RT0545 and Proyecto IV-8 call Amexcid-Auci 2018-2020.

978-1-7281-8987-1/20/$31.00 ©2020 IEEE

It is difficult to be able to name an indicator or variable as a diagnostic factor, but it is possible to analyze the probability of being at risk using a set of variables, which is the paradigm used by risk calculators. These classification processes are normally faced with nonlinear models and for this reason the use of tools such as artificial neural networks (ANN) is necessary. Algorithms capable of solving complex problems such as the classification of EEG signals in control tasks [12], in tasks of patient supervision where, after training and learning, it is able to identify changes in ECG signals that may compromise health status of a patient [13] and finally as a risk classifier in pathologies where early intervention can improve the treatment and outcome of the disease. [14].

In this work, it is proposed to use an ANN as a classifier in order to stratify the possibility of having low, medium, moderate or high risk of developing cardiovascular disease using HRV indexes obtained from 5-minute ECG recordings at rest from an heterogeneous population of the manufacturing industry in Mexicali, Mexico from the PhysioBC® project database [15].

II. METHODOLOGY

The database of the PhysioBC® project has 114 ECG records of a heterogeneous and representative population of the City of Mexicali, Baja California, Mexico. It includes 57 female and 57 male subjects, with an age range of 18 to 60 years old. The records were taken in rest (sitting) with a 12-bit resolution, 500 samples/second and a duration of 5 minutes, necessary parameters to carry out the analysis of HRV. ECG records were processed using MATLAB (The Mathworks, Inc. ©) version 2020a software and the Deep Learning and Wavelet toolbox.

A. Heart Rate Variability

Defined as the study of the changes in the duration of the RR intervals analyzed in each beat, the HRV has become a useful indicator in the prevention, diagnosis and treatment of various health conditions. The tachogram is obtained by measuring the duration between each R wave in an ECG record. Any given HRV index is obtained from the tachogram analysis and is a value that is calculated in various domains: temporal, spectral, and nonlinear.

For records taken at rest, the influence of respiration should be considered, but for outpatient records, movement artifacts, changes in the electrode-skin interface, the possible recording of electromyographic signals, and others should be considered. In this work, the Continuous Wavelet Transform (CWT) is used to identify the R wave in the ECG records of the PhysioBC® database. Being a convolution operation, the selected mother wavelet is the Daubechies 4 due to the morphological similarity that exists with ECG records in leads where R waves of considerable amplitude can be observed such as DII, DIII, aVF, V5 and V6. The proposed algorithm identifies regions based on the scalogram, where it is found the scale in which the highest energy of the QRS complex is located. This value is considered a selection threshold.

Once you have a scale value as a threshold, wavelet coefficients are calculated only at that level and stored in a vector. The next operation is the identification of peaks and the proposal of a width that ensures that the region identified in the wavelet coefficient vector contains the R wave. Misidentification of R waves generate artifacts within the tachogram, known as ectopic heartbeats.

In a convolution operation between the ECG record and the region vector calculated with the wavelet coefficients, the R wave is located, and its location is stored in time. In the new created time vector, which has the information of when an R wave occurred, a subtraction operation (1) is performed to calculate the duration between each beat (NN), thus finally generating the tachogram.

$$NN_{i=1,2,3\ldots}^{n} = \{a_{i+1} - a_i, \ldots, a_{n+1} - a_n\} \quad (1)$$

It is in the tachogram where operations are performed to calculate indexes of HRV. The concept, mathematical definition and units of the HRV indexes are described by the Task Force HRV Guidelines [16]. Temporal indexes are obtained by applying mathematical operations based mainly on measures of central tendency and dispersion. The spectral indexes use auto-regressive methods to know the components in 3 different frequency ranges, whose graphic representation is an aid to understand the balance and contribution of the sympathetic and parasympathetic branches of the ANS. Nonlinear methods apply more complex mathematical operations to understand the physiological dynamics of cardiac control and have been useful in characterizing study groups. The indexes of HRV that were calculated from the PhysioBC® database records are shown in Table I.

TABLE I. HRV INDEXES

Domain		
Time Domain	***Frequency Domain***	***Nonlinear***
• SDNN (ms) • SDANN (ms) • NNx (count) • pNNx (%) • RMSSD (ms) • SDNNi (ms)	Method: • AR Burg (autoregressive) Frequency range: • VLF (0-0.04 Hz) • LF (0.04-0.15 Hz) • HF (0.15-0.4 Hz) • LF/HF	Poincaré: • SD1 (ms) • SD2 (ms)

B. Cardivascular Risk

Risk is a concept that analyzes the interaction of various variables and predicts an outcome. Its robustness lies in selecting the variables that have the most effect on the expected outcome. In the Medical branch it is widely used to infer whether a person may or may not have a health problem. For example, the BMI is a classifier of the weight status of a person, it allows to assign states such as normal weight, overweight and obesity at different stages and establish a level of risk.

There are several risk tables. The PhysioBC® project also includes, in addition to ECG records, Framingham scale evaluations that establish a person's risk of developing

cardiovascular disease. To establish this scale, the 30-year risk predictors are used, the percentage of risk arises from the interaction of the following variables: sex, age, blood pressure, if there is treatment for hypertension, if you are a smoker, if you have a diagnosis of diabetes and the BMI.

C. *Artificial Neural Network*

The most robust risk factors in the health area are those that contain clinical variables and results of other evaluations. This data is captured in questionnaires and, after analysis, the risk percentage is finally obtained, as in the Framingham test [6].

A technological application that can work in a similar way is the ANN. They are applied to supervised learning problems, they train in a set of input-output pairs, where they require a train stage to describe what should occur in response to an input.

The proposed architecture for the classifier is a multilayer perceptron that uses the Levenberg Marquardt algorithm. It has 2 intermediate layers of 4 neurons each, and 1 neuron at the output layer. The transfer function is *tansig* for the hidden layers, and *purelin* for the output layer. The architecture is implemented in MATLAB (The Mathworks, Inc. ©) using the Deep Learning toolbox.

The input vector for the classifier is composed by 11 HRV indexes (SDNN, RMSSD, NNx, pNNx, SDNNi, VLF, LF, HF, LF/HF, SD1 and SD2) of each of the records in the database. The output vector represents 4 levels of risk: low, medium, moderate and high. After a training stage, the ANN classifies the possible risk that a person has of developing cardiovascular disease considering the results of the analysis of various HRV indexes.

III. Results

Inclusion criteria during the analysis of the PhysioBC® project database include the duration and quality of the registry. With the first criterion, 75 records with a duration of at least 5 minutes are selected, as required by the norm for the analysis of the HRV; when applying the second criterion, 15 records are eliminated, since they present more than 3 ectopic heartbeats in the tachogram.

A. *Heart Rate Variability*

Two types of records are found in the database, 12-lead ECGs and in other cases only tachograms are available. 38 ECG records were processed to identify the R wave and generate the tachogram, leads DII, V5 and V6 were selected according to the lesser presence of artifacts or noise. For the identification of the R wave, the algorithm by regions was used. The mother wavelet used was Daubechies 4 and 64 scales were calculated, Fig. 1.

From the 38 records, since they are taken at rest, there are not many changes at the scales where the greatest contribution of energy manifests, the scales ranged from 22 to 26. With the identified scale it is ensured that the CWT coefficients oscillate with greater amplitude around the QRS complex, Fig. 2.

In the last operation of convolution between the vector of regions, found at the DII ECG record, it is now possible to identify the location of the R wave as a maximum, Fig. 3. With this we now have in a vector the information of when the heartbeat occurred, the remaining operation is the application of (1) to generate the tachogram and be able to calculate the HRV indexes.

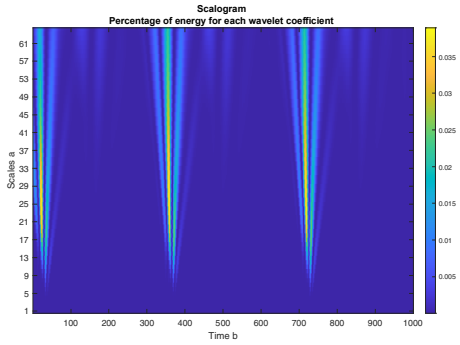


Fig. 1. Scalogram for the detection of an optimal escale considering the maximun change of porcentage of energy in wavelete coeficients.

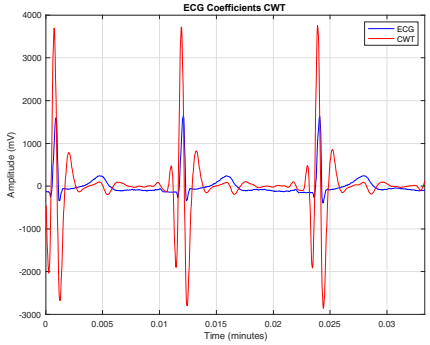


Fig. 2. The selected scale encourages the maximum oscillation between wavelet coefficient and the R wave of the ECG.

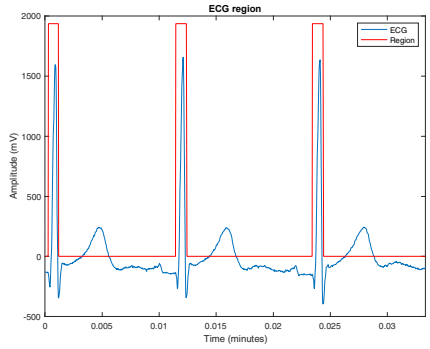


Fig. 3. R wave identification consists in the selection of a maximum value within the region using DII lead.

The tachogram is already a first indicator of HRV, in general terms, a high variability is related to a good response of

the autonomic nervous system and its interaction and control with other organs and systems of the human body and could be related to a state of health with no apparent complications, Fig. 4a.

On the other hand, a decreased variability could indicate alterations in the actions of control and interaction of the ANS, in addition to the influence of comorbidities, Fig 4b.

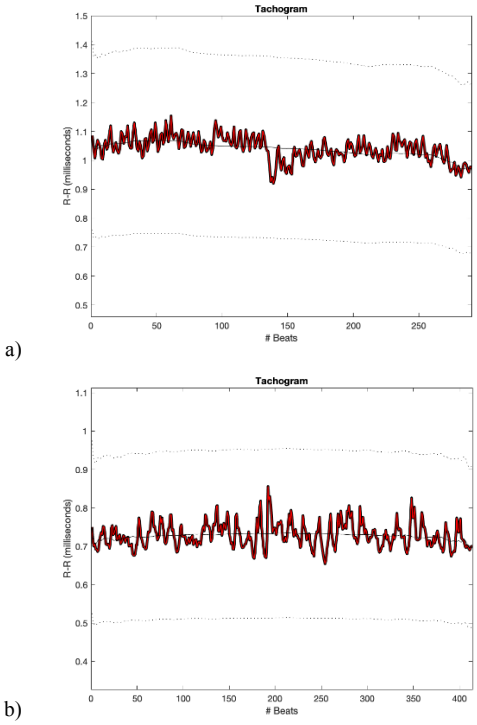


Fig. 4. Tachograms processed by the regions algorithm. a) Tachogram of patient 4, variability is considerable, current Framingham risk factor at 30 years is 5 %. b) Tachogram of patient 11even when variations are 100 ms in average, there is a lower variability, current Framingham risk factor at 30 years is 41%.

37 records were obtained directly from the tachograms published in the database, eliminating those that contain more than 3 ectopic heartbeats. Finally, the indexes of the HRV in temporal domain, frequency domain and nonlinear domain are calculated to 60 records. The indexes in the frequency domain study the interaction of the sympathetic and parasympathetic branches, the literature refers that the values in the range called high frequency (HF) are related to the parasympathetic branch, while the contribution observed in the band Low frequency (LF) is an interaction of both branches, but with a greater contribution from the sympathetic branch. This interaction can also be an indicator and be related to possible health states.

In a resting test, there should be a greater contribution in the LH band, Fig. 5a, if there is a greater contribution in the HF band, it is possible that something is altering the interaction and control with the parasympathetic branch, Fig. 5b.

In nonlinear indexes, it is also possible to analyze these changes. The Poincare index is an indicator that is related to a scatter plot of the duration of the NN intervals [16], Fig. 6.

The frequency and nonlinear indexes showed the greatest change and a possible relationship with risk stratification. The 11 HRV indexes used in this proposal are: SDNN (ms), RMSSD (ms), NNx (1min-Count), pNNx (50ms), SDNNi (1 min), VLF (Power %), LF (Power %), HF (Power %), LF/HF, SD1 (ms) and SD2 (ms).

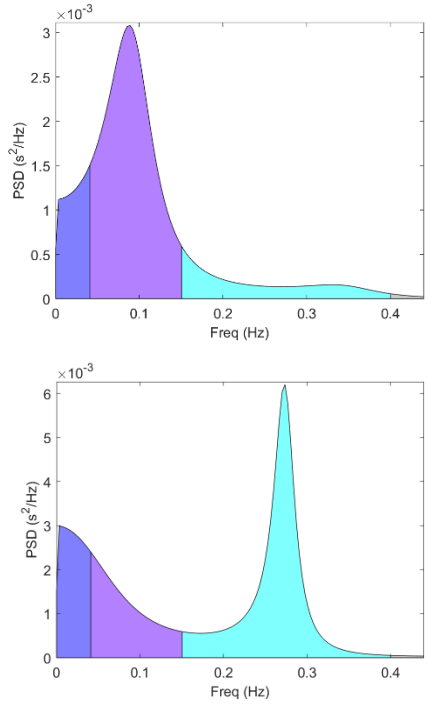


Fig. 5. HRV Indexes in the frequency domain. a) Analysis of patient 3, in a resting test, there is a larger contribution from LF band, current Framingham risk factor at 30 years is 5 %. b) Analysis of patient 6, in a resting test, there is a larger contribution from HF band, which possibly should not manifest in a resting test, in a resting test, current Framingham risk factor at 30 years is 62%.

B. *Cardivascular Risk*

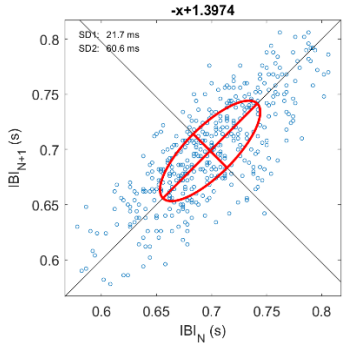
To establishing the use of a variable or set of variables as a risk factor is such a complex task and is not the purpose of this work.

In order to establish if these parameters have any relationship with risk factors, an association study should be performed among the variables, and it is neither an objective in this work. The HRV indexes are affected by age, so for any analysis that must be carried out it is important to consider grouping the data again, the proposal is to define 4 groups, those of age up to 30 years (Group 1), a second group with up to 40 years (Group 2), a third group with up to 50 years (Group 3) and a fourth and last group with those over 50 (Group 4). Once this new distribution of data was performed, the means and standard deviations of each of the HRV indexes were analyzed, the results are shown in Table II.

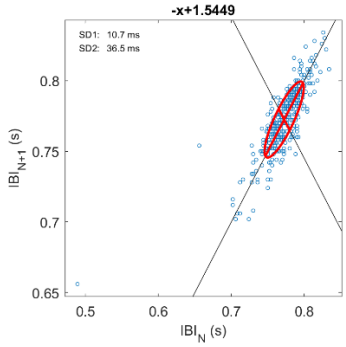
TABLE II. ANALYSIS OF THE HRV INDEXES FOR 4 GROUPS OF AGE

		SD1(ms)	SD2(ms)	VLF (%)	LF (%)	HF (%)	LF/HF (ratio)	SDNN (ms)	RMSSD (ms)	NNx (1min-Count)	pNNx (50ms)	SDNNi (1min)	Risk factor CV (%)
G1	**Mean**	20.4	64.0	22.0	54.7	23.3	3.2	47.5	28.8	38.8	9.6	308.9	5.8
	STD	8.8	22.9	12.1	13.9	11.9	1.8	17.1	12.4	41.0	10.5	28.7	4.9
G2	**Mean**	19.1	58.0	18.3	59.4	22.3	3.4	43.3	27.0	30.5	7.7	279.0	16.7
	STD	5.9	12.2	7.6	13.4	8.4	2.3	9.1	8.3	26.5	6.7	97.6	6.5
G3	**Mean**	16.4	53.7	26.9	52.6	20.5	4.5	39.8	23.1	20.4	5.2	313.7	40.3
	STD	7.4	17.8	10.3	17.5	15.8	3.6	13.3	10.4	30.6	8.0	38.9	17.5
G4	**Mean**	15.1	52.5	35.0	51.1	13.8	7.0	38.7	21.3	8.1	2.6	338.1	59.3
	STD	7.7	16.5	10.4	18.0	9.3	6.6	12.5	10.9	10.2	3.8	63.9	11.8

The classification by age groups allows the analysis of the HRV indexes to be carried out and to be able to consider whether the changes can be related to the possibility of risk. From Table II it can be seen how the mean value of indexes such as SD1, SD2, HF, RMSSD, NNx, and pNNx decrease in each age group, while the risk factor for developing cardiovascular disease increases with each defined group.



a)



b)

Fig. 6. HRV Indexes for nonlineal domain, Poincaré a) Analysis of patient 7, dispersion in axes SD1 and SD2, current Framingham risk percentage at 30 years is 3 %. b) Analysis of patient 46, dispersion in axes SD1 and SD2, current Framingham risk percentage at 30 years is 53 %.

When analyzing the standard deviations, none of the variables can classify any level of risk since the limit values of the groups overlap. In the case of the risk factor and its relationship with the established age groups, even with considerable standard deviations, 4 risk groups could be established: low, medium, moderate, and high.

C. Artificial Neural Network

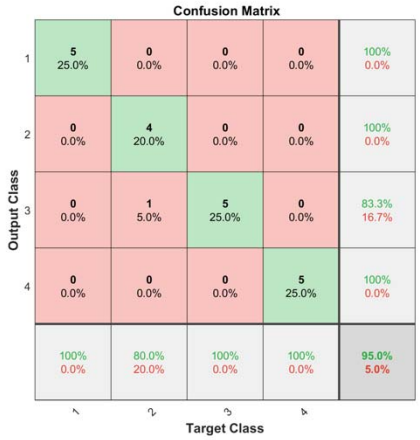
Using only the age to establish a level of risk is not enough. In addition, the limits of the groups data overlap complicating the classification, there are subjects in Group 1 with a high risk and/or people in Group 4 with a low risk level.

The ANN characteristic vector is composed using 11 indexes of HRV, from Table II, age, and BMI to classify the possibility of being in one of the 4 defined risk groups.

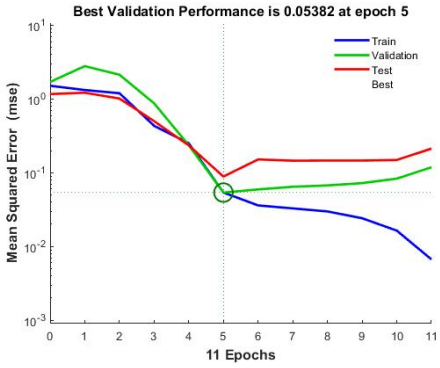
The final sample of 60 characteristic vectors of 11 elements each, that correspond to 60 patients, will be divided into 2 sets, the first for training and testing made up of 40 records, 10 records for each age group. 60 % of this group is used for the training set and 40 % for the testing set.

The second group is made up of 20 records, 5 for each age group and is used to validate the ANN.

To validate the classification a confusion matrix is used, which allows the analysis of the sensitivity and specificity. An average classification of 79 % ± 10 % is achieved in 70 % of the executions of the architecture, the most efficient architecture reported has a classification percentage of 91.7 % in training, 100 % using the test vectors and finally 95 % in the validation group, Fig. 7.



a)



b)

Fig. 7. ANN classifier results. a) Confusion matrix of the 13 charcateristic vectors built using HRV indexes, age and BMI. The classifier efficiency in the validation group is 95 %. b) Performance plot of the classifier for training, validation and test.

IV. Discussion and Conclusion

The HRV indexes present changes useful for studying possible health stages only if they are grouped by age. As shown in Figures 4, 5 and 6, in some cases the possibility of classification with the indexes shown is very evident, however this does not happen with all records.

The database of the PhysioBC® project, due to its technical characteristics, is useful for the analysis of the HRV. In future work it will be interesting to look for the possible relationship of the indexes with other parameters besides the risk factor such as the performance of physical activity, variable that was also collected by the research group.

The ANN is a viable option for the development of a classifier, even with a basic architecture such as the multilayer perceptron it was possible to meet the objective of the work, new proposals should be used to optimize the classification. Vectors proposed as input to ANN with HRV indexes may be a good option for characterizing populations.

It is necessary to have a greater number of repositories of biological signals (ECG, electromyography and electroencephalography) that allow various research groups to develop and validate auxiliary technological tools that can be used in the prevention, diagnosis and treatment of diseases.

References

[1] J. I. Arocha Rodulfo, "Sedentarismo, la enfermedad del siglo xxi," *Clínica e Investigación en Arteriosclerosis*, vol. 31, no. 5, pp. 233–240, Sep. 2019.

[2] M. Narici *et al.*, "Impact of sedentarism due to the COVID-19 home confinement on neuromuscular, cardiovascular and metabolic health: Physiological and pathophysiological implications and recommendations for physical and nutritional countermeasures," *European Journal of Sport Science*, pp. 1–22, May 2020.

[3] M. Di Cesare *et al.*, "The epidemiological burden of obesity in childhood: a worldwide epidemic requiring urgent action," *BMC Medicine*, vol. 17, no. 1, Nov. 2019.

[4] M. D. Shapiro *et al.*, "Preventive Cardiology as a Subspecialty of Cardiovascular Medicine," *Journal of the American College of Cardiology*, vol. 74, no. 15, pp. 1926–1942, Oct. 2019.

[5] I. Lee, S. Kim, and H. Kang, "Lifestyle Risk Factors and All-Cause and Cardiovascular Disease Mortality: Data from the Korean Longitudinal Study of Aging," *International Journal of Environmental Research and Public Health*, vol. 16, no. 17, p. 3040, Aug. 2019.

[6] "Cardiovascular Disease (10-year risk) | Framingham Heart Study," *Framinghamheartstudy.org*, 2020. [Online]. Available: https://framinghamheartstudy.org/fhs-risk-functions/cardiovascular-disease-10-year-risk/.

[7] D. Zdrenghea *et al.*, "CV RISK – A new relative cardiovascular risk score," *Medical Hypotheses*, vol. 132, p. 109362, Nov. 2019.

[8] T. Leti and V. A. Bricout, "Interest of analyses of heart rate variability in the prevention of fatigue states in senior runners," *Autonomic Neuroscience*, vol. 173, no. 1–2, pp. 14–21, Jan. 2013.

[9] Y. Sekiguchi, W. M. Adams, C. L. Benjamin, R. M. Curtis, G. E. W. Giersch, and D. J. Casa, "Relationships between resting heart rate, heart rate variability and sleep characteristics among female collegiate cross country athletes," *Journal of Sleep Research*, vol. 28, no. 6, Mar. 2019.

[10] D. P. Williams *et al.*, "Heart rate variability and inflammation: A meta-analysis of human studies," *Brain, Behavior, and Immunity*, vol. 80, pp. 219–226, Aug. 2019.

[11] F. Sessa *et al.*, "Heart rate variability as predictive factor for sudden cardiac death," *Aging*, vol. 10, no. 2, pp. 166–177, Feb. 2018.

[12] V. Asanza, A. Constantine, S. Valarezo, and E. Pelaez, "Implementation of a Classification System of EEG Signals Based on FPGA," *2020 Seventh International Conference on eDemocracy & eGovernment (ICEDEG)*, Apr. 2020.

[13] Y. Cao, T. Wei, N. Lin, D. Zhang, and J. J. P. C. Rodrigues, "Multi-Channel Lightweight Convolutional Neural Network for Remote Myocardial Infarction Monitoring," *2020 IEEE Wireless Communications and Networking Conference Workshops (WCNCW)*, Apr. 2020.

[14] J. He, L. Shen, X. Ai, and X. Li, "Diabetic Retinopathy Grade and Macular Edema Risk Classification Using Convolutional Neural Networks," *2019 IEEE International Conference on Power, Intelligent Computing and Systems (ICPICS)*, Jul. 2019.

[15] N. Flores, R. L. Avitia, M. A. Reyna, and C. García, "Readily available ECG databases," Journal of Electrocardiology., vol. 51, no. 6, pp. 1095–1097, 2018. DOI: 10.1016/j.jelectrocard.2018.09.012.

[16] T.F. of the E. S. Electrophysiology, "Heart Rate Variability," *Circulation*, vol. 93, no. 5, pp. 1043-1065, Mar. 1996.