Research article

Analyzing ECG signals in professional football players using machine learning techniques

A.A. Munoz-Macho [a,b,*], M.J. Dominguez-Morales [a], J.L. Sevillano-Ramos [a]

[a] Computer Architecture and Technology Department, University of Seville, Spain
[b] Performance and Medical Department, RCD Mallorca SAD, Palma de Mallorca, Spain

ABSTRACT

Background: Football player's health is important, and preventing sudden cardiac arrest may be a critical issue. Professional football players have different ECG signals than the average population, yet there are considerable gaps in study whereas the general population has been extensively studied.
Objectives: (a) Generate a reference and innovative resting 12-lead ECG database from 54 UEFA PRO level male football players from La Liga. This is a novel approach to cope the ECG and possible arrythmias in athletes. (b) Manage each XML athlete ECG data and develop a free-use program to visualize, denoise and filter the signal with the capacity to automate the labelling of the waves and save the reports. (c) Study the ECG wave shape and generate models through ML to analyse its utility to automate basic diagnosis.
Methods: The dataset collection is based on a prospective observational cohort and includes 10 s, 12-lead ECGs and rhythm and condition labels for each athlete. Physiological sport arrhythmias, T-Wave shape and other findings were studied and labelled. ECG Visualizer was developed and used for 3 machine learning (ML) methods to automate sinus bradycardia arrhythmia diagnosis.
Results: A dataset with 163 ECGs in XML format was collected comprising the Pro Football 12-lead Resting Electrocardiogram Database (PF12RED). "ECG Visualizer" software was developed, and ML was shown to be useful in detecting sinus bradycardia.
Conclusions: The study demonstrates that AI and machine learning can detect simple arrhythmias with accuracy, also it provides a valuable dataset and a free software application.

1. Introduction

An electrocardiogram (ECG) is a graphical representation of the electrical changes occurring inside the heart's components and their voltage over time. Every pulse causes the muscles to depolarize and repolarize, as the heart muscle can adjust its voltage to induce contractions over time. The normal ECG of a normal heartbeat is comprised of P, QRS, and T waves.

The first QRS detection algorithm method was created in 1985 [1], and recently thanks to machine learning (ML) it will be possible to recognise arrhythmias, even matching or surpassing the classifications made by experts [2].

In professional athletes such as professional football players, the ECG undergoes physiological changes as a result of adaptations to the continual and highly demanding training and competitions.

Changes in the heart's electrical conduction could potentially occur because of these alterations, favouring the development of physiological arrhythmias such as sinus sport bradycardia (SB) or incomplete right bundle branch block (iRBBB). These physiological alterations are well-described in recurrent meetings of sports cardiologists and sports physician specialists, the most recent of which established the International Criteria for ECG Screening [3]. An abnormal ECG finding in athletes is unrelated to regular training or expected physiological adaptation to exercise and may suggest the presence of pathological cardiovascular disease or borderline situation and require further diagnostic investigation such as echocardiogram, cardiac magnetic resonances, and stress electrocardiogram tests (Figure 1).

Concerned about the problem in football, FIFA and other international statements have tried to protect players, promote pre-competitive medical screenings using electrocardiograms [4], and compile statistical data on sudden deaths [5].

Finally, the actual databases for arrhythmia studies are based on common populations (See Table 1), eventually sedentary, with health conditions, or populations with possible Berkson's bias because they are studied at hospitals [6,7].

2. Objectives

(a) Generate an initial database from 12-lead ECG with 10 s 5000 blocks of data at 500 Mhz.
(b) Manage XML from ECG and patient data and construct a program to view, denoise, and filter the signal while automating wave labelling and CSV report exporting.
(c) Examine the ECG waveform and develop models using machine learning to assess its applicability for automating basic diagnostics.

3. Methods

3.1. Participants

The findings are based on 163 resting 12-lead ECGs from 54 male UEFA Europa League-level professional football players. Detailed information of the participants can be found in Table 2. With the anthropometric and cardiac values. Each football player could have multiple ECGs, but only one ECG from each participant was chosen for the study; the selected ECGs are highlighted in yellow in Table 3 XML&PDF in the Github repository [8].

3.2. Ethics, trial registration

The study is part of a project focused on exploring the systems models and the application of AI and ML in professional football. The author AAMM inscribed the project as "Development and Implementation of Model-Based Systems for Professional Football Teams, Aimed at Optimizing Health and Performance" and it is registered in ClinicalTrials.gov (No. NCT05872945).

The study was approved by the Autonomous Community of Andalusia Ethics Committee (Spain) (Protocol Number: 1573-N-19, December 2019), which granted the form to obtain informed consent, and allowed the data to be shared publicly after anonymization.

3.3. Diversity, equity, and Inclusion Statement

Concerning our consistent dedication to Diversity, Equity, and Inclusion, we hereby disclose the demographic makeup of the 54 professional male footballers who participated in the research. Forty members of European Caucasian descent comprise the cohort, while seven members are of Latin American origin and race, six are of African descent, and one is of Asian descent. The percentages in question are as follows: 74.07 percent, 12.96 percent, 11.11 percent, and 1.85 percent, respectively. It is imperative to acknowledge that the absence of professional teams in the study area prevented the inclusion of any female participants in this research. Our organisation remains committed to advancing diversity and inclusivity in every facet of our research endeavours, and we shall persistently pursue fair and balanced representation in forthcoming studies.

3.4. CHecklist for statistical assessment of medical papers (CHAMP)

This study has completed and sent the CHAMP file [9] and the STROBE Statement [10] as a checklist file during the authorship process.

3.5. Data acquisition and signal processing

The data was recompiled on a prospective observational cohort, gathered in five phases and saved in XML format. Most of XML have their reference in PDF format. First, in the 2018–2019 postseason with 24 XML, in the 2019–20 preseason with 41 XML; in the 2019–20 postseason with 34 XML; in the 2020–21 preseason with 37 XML; and fifth, with 27 XML Registries in the 2021–22 preseason. It can be found up to six players with the complete series.

Each athlete underwent a 10-s resting 12-lead ECG as part of a medical evaluation. The data was stored in the General Electrics (GE) ECG Software CardioSoft V6.73 12 S L V21. The principal investigator AAMM, a registered sport and exercise physician and expert in sports cardiology, then labelled the characteristics and results of each ECG. The final diagnoses were stored in the GE ECG Software

CardioSoft V6.73 12 S L V21 system and a resume CSV in the GitHub repository [8].

ECG data and diagnostic information were exported from the local server to XML files that were encoded with a specific naming conversion defined by boolean methods using phyton via Google Collab. Finally, we developed a tool named ECG Visualizer [11] for translating ECG data and diagnostic information from XML files to CSV format.

3.6. Data filtering methods and peaks detection

In this investigation, although the electrocardiograms are of excellent quality, the sources of ECG noise contamination were power line interference, electrode contact noise, motion artefacts, muscle contraction, and baseline noise. Therefore, we devised and implemented a sequential noise reduction method for ECG raw data processing.

3.6.1. Data filtering

Various algorithms were used for data processing, the signal filtering process was based on these 4 types on all ECG leads.

1. **Fixed window average filter** (mean/smoothing filter). The average value of the signal within a given window width is calculated. This filter does not include overlapping and therefore the signal is divided into blocks of size the width of the window to finally obtain a signal with a lower number of samples (number of original samples divided by the width of the window). Therefore, it can be seen in the application that applying this filter causes a reduction in the sampling hertz of the signal rate to 100 Hz. This filter removes the digital noise from the sampling [12].
3. **Sliding window average filter** (moving average filter Sliding-window average filter (or moving average filter): as with the previous filter, the average value of the signal within a given window width is calculated. However, on this occasion, signal overlapping is performed so that, after each calculation of the mean, the next calculation is performed with the same window but shifted one sample forward. This causes each mean calculation to include all the samples of the previous calculation except the oldest one (adding a new one in its place). This causes the number of samples to only be reduced at its final margin by N-1 samples (N being the size of the window); thus, practically the original number of samples is maintained. This filter removes the high-frequency and low-frequency ripples [13].
4. **Sliding window median filter** (moving median filter). It works the same way as the previous filter with the difference that, this time, the median operation is applied over the sample window, instead of the average. It flattens deeply the signal peaks and reduces the noise [14].
6. **Band rejection filter** (band-stop filter) periodic samples are removed with a repetition of a frequency determined by the user. Since the frequency is usually not exact, this band rejection is applied to the base frequency ±5 Hz. This filter is especially useful if we know certain frequencies that are introducing noise to the original signal, such as the frequency of the alternating current produced by the mains connection (50 or 60 Hz depending on the continent) [15].

After this filtering, the peak detection process is performed by the same tool. To obtain a precise output, the tool allows the user to select the target user between a common person, an athlete or a custom class determined by the user. This selection is important, as the class determines PR, QRS and QT intervals according to some authors [16,17]. The application usage diagram is presented in Figure 2.

3.6.2. Peaks detection

To visualize the information from the ECGs, we used "ECG Visualizer" which can represent the 12-lead waveform, filtering the signal to eliminate noise and automatically detecting all the P, Q, R, S and T peaks of each lead as described below and performed in this order: [11]. The algorithms used were.

- R peaks: The first value was searched using a local maximum search from the beginning. After that, the other R peaks were searched looking for peaks of similar amplitudes.
- Q peaks: Starting from the R peaks location, the search window was established by the QRS maximum and minimum intervals configured according to the type of user.
- S peaks: They were detected following the same procedure as that used for the Q peaks but in reverse.
- P peaks: Starting from the Q peaks location, the search window was established by the PQ maximum and minimum intervals configured according to the type of user.
- T peaks: A similar procedure was performed as for P peaks but using the QT interval configured for the search window. The only aspect to take into account was that it cannot be considered that the polarity of the T peak is the same as that of the corresponding P and R peaks. This aspect must be detected too.

Finally, the basic ECG information is stored locally in a CSV file, as well as the mean distance of the most significant segments and detected T-wave inversions.

3.7. Data records

The dataset comprises 163 raw ECG data in XML format, as well as a descriptive results file, up to 121 ECG in PDF format, with most of them containing its corresponding XML file, and a glossary. 51 players with a complete profile with XML and PDF format ECG were

selected for the ML study [8].

For each subject, the raw ECG data from each stage were saved as a single XML file that was named by unique IDs. These IDs were also located in the diagnostics file that contains all the information for each subject.

3.8. ML development

The third objective of the paper was to study the ECG wave shape and generate models through ML to analyse them and give advice according to the international criteria for ECG interpretation [3].

After generating the automatic report via ECG Visualizer, we proceeded to implement an automatic classifier that makes use of the data obtained from the tool's reports and the labels provided by AAMM.

According to the report, three classifiers are used to assess sinus bradycardia in the test instance below. Based on the mean RR segment distance (or beats per minute), sinus bradycardia is usually easy for a doctor to diagnose. This test tests whether an autonomic system can recognise the relationship between the parameters. Figure 3 shows the process.

The software tool generates age, race, height, weight, PQ, QR, RS, ST, RR, QRS, QT and corrected QT intervals, T-wave inversions in each derivation, race, and other classifiers, totalling 27 input variables.

Next, the Hold-Out technique is used, dividing the dataset into a training subset and a test subset, using in our case a 70-30 division. The first subset is used to train three different types of classifiers: an Artificial Neural Network, a Support Vector Machine, and a Random Forest.

The rationale behind selecting these specific ML algorithms was to evaluate the mechanisms that have been used in similar works that analyse physiological signals [18].

Although the use of ANNs usually obtains better results, in occurrences that are not very complex to detect, other simpler algorithms such as SVM and decision trees, have been shown to obtain good results. In addition, the latter are computationally much less expensive and therefore do not require high-computational cost equipment.

3.8.1. Training and test subsets

From the information generated by the GE electrocardiogram, only the RAW data of the 12 leads were used. None of the parameters pre-calculated by the electrocardiogram were used. With the RAW data of the 12 leads, the filtering and peak detection process was performed with the software tool developed. After that, the features are extracted with the same tool and the report is generated.

On the other hand, physician interpretation was considered the gold standard to train the AI algorithm. AAMM registered sports & exercise physician with more than 20 years of experience and expert in sports cardiology labelled the information on the rhythm and pathologies of each player and annotated one by one. In this way, the information extracted with the software tool is used as training input data and the information annotated by the health professional as labels of those data.

After the splitting process, training and test subsets were checked obtaining the Loss Curves to observe insights into the model's performance and the learning process (See Figure 4).

3.8.2. Random forest (RF)

This model consists of an ensemble of decision trees [19], where each tree obtains a random vector sampled independently, having uncorrelated tree models [20]. For this test, 10 estimators were used and an unlimited maximum number of features.

3.8.3. Support vector machine (SVM)

This classifier searches for a hyperplane in a space of dimension based on the number of input variables to separate the data into different classes [21]During the process a SVM optimisation process was performed with a Bayesian hyperparameter optimisation using functions integrated in the TensorFlow library. Through this optimisation, we seek to find the optimal values of 'C' and 'gamma' from a range of more than 20 values for each. The result yields a value of 10.0 for 'C' and 0.01 for 'gamma'. Other parameters needed to be adjusted, like the maximal margin hyperplane (maximizing the distance between the support vectors and the hyperplane) [22].

3.8.4. Artificial Neural Network (ANN)

Finally, an ANN classifier is used, more precisely a multilayer perceptron network (MLP) [23]. In this work, we have not performed a grid search to find the best parameter combination; so, the neural network used has 4 layers: an input with 27 inputs, one hidden layer with 16 neurons, a second hidden layer with 8 neurons, and the output layer with 2 neurons (one for each class). (See Figure 5).

The hyperparameter selection was developed with a Grid Search process that has been carried out using different values for each hyperparameter: 4 different values for the learning rate, 4 different values for the batch size and 4 different customised architectures.

For the architectures, what has been taken into account is the number of hidden layers (between 0 and 3); in addition, assuming that the input and output layers have a fixed number of neurons, the intermediate layers have been created with a decreasing number of neurons concerning the input layer, forming a linear reduction from the input neurons to the output neurons (for example, if the input layer is 24 neurons and the output layer is 2, 11 neurons have been used with a single intermediate layer).

Tests have also been carried out with and without dropout. All this resulted in a total of 128 combinations. For each of them, the dataset was randomly divided into training and test subsets and, with the results of the test, the option with the best result was the one with 2 hidden layers, batch size of 4, learning rate of 0.001 and with dropout.

In addition, although it is not indicated in the article so that it is not so dense, the two with the best results after the Grid Search process underwent a Cross-Validation process to measure the robustness of the system to changes in the training and test sets. And,

after these results is when the neural network was finally decided. Finally, the values were 5e-4 for learning rate, 0.2 for dropout (used between each pair of layers) and 4 for batch size.

4. Results

In these players, the results can be seen in Table 3. During the screening main results were detected in 64.81% for physiological sport-related Sinus Bradycardia (SB), 66.67% for incomplete Right Bundle Branch Block (iRBBB) 66.87% for Right T Wave Inversions (RTWI) and 1.85% Mobitz Type I - 2° degree AV block (MTI) based on the International Criteria for ECG Interpretation in athletes. Finally, we found one player (1,85%) with T-Waves inverted in left leads V4, V5 and V6 that needed complete cardiologist testing to approve his medical aptitude.

4.1. ML results

After the training process, the classification results are evaluated using the metrics explained below.

- True Positive (TP), False Positive (FP), True Negative (TN), False Negative (FN).
- Accuracy: proportion of TP and TN in all evaluated cases (see Equation (1)).
- Sensitivity (or Recall): proportion of TP in all the cases that belong to this class (see Equation (2)).
- Specificity: proportion of TN in all cases that don't belong to this class (see Equation (3)).
- Precision: proportion of TP in all cases that have been classified as it (see Equation (4)).
- F1-score: a measure of a test's accuracy. It considers both the precision and the sensitivity (recall) of the test to compute the score. It is the harmonic mean of both parameters see Equation (5).

$$accuracy = \sum_c \frac{TP_c + TN_c}{TP_c + TN_c + FP_c + FN_c}, c \; \epsilon \; classes \tag{1}$$

$$sensitivity = \sum_c \frac{TP_c}{TP_c + FN_c}, c \; \epsilon \; classes \tag{2}$$

$$specificity = \sum_c \frac{TN_c}{TN_c + FP_c}, c \; \epsilon \; classes \tag{3}$$

$$precision = \sum_c \frac{TP_c}{TP_c + FP_c}, c \; \epsilon \; classes \tag{4}$$

$$F1_{score} = 2 * \frac{precision * sensitivity}{precision + sensntivity} \tag{5}$$

The confusion matrices obtained for each classifier are shown in Figure 6.
Next, the results obtained for each classifier: RF, SVM and ANN are detailed in Table 4.

5. Discussion

Artificial intelligence (AI) and Machine Learning (ML) have shown significant development in recent years. Indeed, it has helped to predict and give advice in complex calculus in several fields: architecture, aerospace engineering, etc [24]. In the sports industry, data organisation is still being implemented and it would be important that the AI achieves the challenges to be respectful of human-centred activities [25]. Also is important to create ordered data to generate new knowledge to aid practitioners in their day-to-day planning and decision-making during training and competitions, as well as the individual condition of each team member, to promote better performance and player health. Indeed, technology can help understand complex systems and biological signals [26] and some authors highlight how machine learning (ML) may be utilised to enhance performance, avoid injuries, and monitor player health [27–30].

Football has many complex situations that can impair the health of the players, is a sport where the players must develop great efforts [31], and the availability of players and its effects are very important [32]. The fixtures are congested [33] with concatenated matches [34] trip plannings, and sleep management [35] and it may be possible to study this by analysing blood and cardiac bio-markers [36].

ML could be important in aiding in ECG pattern recognition according to the consensus and experts [3] and our research reveals how three of them have good results in detecting sinus bradycardia. SB is classified with an accuracy that exceeds 96% for all cases. It is interesting to discuss how more complex classifiers like SVN and ANN obtain worse results than the less complex ones, RF.

This is because neural networks and support vector machines try to use all the input features to classify the samples, even if some of them do not intervene in the results. This means that if one variable distorts the information provided by the others, it may be counterproductive to include it in the training. However, for a Tree-based classifier such as an RF, the variables to be used in each decision step of the tree are determined during the training. Therefore, the latter classifier may only need to focus on a subset of

variables without paying attention to the others.

In summary, for problems that are easily linearizable, it is preferable to use simpler classifiers such as decision trees or linear regressions. In these cases, more complex classifiers, such as SVMs or neural networks, seem excessive for the problem to be classified. However, for classifiers that must look for combined relationships between variables that are not easily discernible (or that do not follow a clear pattern), the latter classifiers are more useful.

This conclusion can be observed in the graphical representation of some of the trees provided after the random forest classification. These representations are shown in Figure 7.

In Figure 7, we can observe that the classifiers use similar decisions to determine sinus bradycardia as the ones used by healthcare professionals. The other classifiers do not provide the parameters used for their decisions, as they act as black boxes, so we cannot obtain a similar graph to the one represented in Figure 7 for random forest.

Our dataset is the first of its kind in terms of composition, featuring XMLs containing raw data from records and select records in PDF format for visual clinical assessment. Regarding tools akin to the ECG Visualizer, it is pertinent to note the following:

Table 5 shows analogous software tools capable of analysing ECG signals. We are able to evaluate the type of input, the number of leads, the filtering process, the peaks detection process, and the feature extraction in order to juxtapose them with our own.

As can be discerned, merely a handful of these tools permit the application of manual filters. However, the principal issue lies in the fact that none facilitates the configuration of peak detection and feature extraction; consequently, these other tools cannot be tailored to specific types of individuals, such as athletes.

Hence, it can be asserted that our tool is the sole entity currently capable of detecting anomalies in ECG signals from athletes.' Concerning the application of AI in ECGs, the work of certain authors has been noted [18].

In relation to the work of other authors, ours bears resemblance to that of certain individuals who have created Arduino-based ECG systems to automate the detection of fundamental arrhythmias such as sinus bradycardia, tachycardia, and bradycardia, as well as other potential findings like cardiac hypertrophy, but not in a professional context [37]. Although some authors have achieved noteworthy advancements in the use of ANN to identify athletes based on their gait and ECG patterns, applying these findings to elite sports is still a long way from clinical use [38]. Others, however, suggest classificators for investigating the morphology of electro-cardiograms (ECGs) in physically demanding settings and utilising ANN; nevertheless, these individuals are not professional sportsmen [39].

5.1. Clinical implications

This work attempted to continue the path set by other authors [40] by generating a specific ECG database for athletes and pro football players describing and expanding the present applications of machine learning in elite team sports.

In addition, this information can serve as guidance for professional leagues, which should use this data collection and harmonise it to create ever-larger data sets that can be utilised to generate more accurate predictions or conduct a more in-depth analysis of these complex ecosystems.

Some authors have previously described how pro teams can benefit from these automated or artificial intelligence technologies [41]. Looking for inclusion and diversity findings to give more specific guidelines is important to classify and do studies and datasets in several populations. The female population has its characteristics that we must consider [42].

Sports cardiology and pre-competition health screenings improve day by day, the dataset published will serve as a reference as the first pro-athlete ECG dataset. It could be used for research or practitioners' assistance and perhaps will change athlete protection policies. This study proposes repeated ECGs during the season to monitor player health and detect possible cardiac changes or systems to promote continuous ECG registration during exercise [43]. It also provides a free "ECG Visualizer" tool for practitioners and researchers and a preview of how machine learning and AI could automate diagnosis in sports with the first example of detecting sinus bradycardia.

5.2. Limitations

The research was conducted during typical seasons, and it was challenging to organise all the data. The missing data and the loss of follow-up were addressed accommodating the amount of total data and analysis.

The ECG Visualizer compatibility was limited to only XML format which was a significant drawback. The authors are expanding its compatibility to other formats like CSV to increase its utility and accessibility. Initially, this tool was developed to open and analyse the files generated by a specific commercial electrocardiogram and the authors are developing a new integration of another commercial ECG into this tool and are currently developing a utility to load ECG files recorded in RAW (CSV and XML).

Finally, as indicated in the Diversity, Equity, and Inclusion Statement, the study comprised only male athletes with limited racial and ethnic characteristics. More diversity was sought, but the availability of individuals of different genders and ages was limited, this constitutes a limitation to extrapolating results to the whole population. The authors of this research are concerned and are developing new projects to improve these aspects.

6. Conclusion

The study demonstrates that AI and machine learning can detect simple arrhythmias such as sinus bradycardia with accuracy, also it provides a valuable dataset and a free software application.

To reduce health problems and sports-related deaths, additional research and larger data sets are required, as well as multicentric data resources and rigorous clinical trials.

Future studies will concentrate on the automatic identification and data gathering of complex physiological arrhythmias such as early repolarization, ST segment elevations, and T Wave inversions in athletes and their correlation with normal or abnormal clinical findings.

Data and code availability

The dataset and other documentation can be found at https://github.com/dradolfomunoz/PF12REDhttps://github.com/mjdominguez/ECGVisualizer/tree/main/ECGVisualizer/src [8], the source code of the ECG Visualizer can be found at https://github.com/mjdominguez/ECGVisualizer [11], which contains also binary executable files, source code and user manual and other documentation in several folders.

CRediT authorship contribution statement

A.A. Munoz-Macho: Conceptualization, Formal analysis, Investigation, Methodology, Resources, Validation, Writing – original draft. **M.J. Dominguez-Morales:** Data curation, Formal analysis, Methodology, Resources, Software, Validation, Writing – review & editing. **J.L. Sevillano-Ramos:** Formal analysis, Methodology, Supervision, Validation, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Appendix C. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e26789.

Appendix A. Tables

Table 1
ECG Databases Comparison

Name	Subjects	Records	Age	Sampling Rate	Leads (n)
MIT-BIH	47	48 (30min)	23–89	360 Hz	2
AHA	N/A	154 (180 min)	N/A	250 Hz	2
EDB	79	90 (120min)	30–84	250 Hz	2
CU	35	35 (8 min)	N/A	250 Hz	2
NSD	2	12 (30 min)	51–69	360 Hz	2
St. Petersburg DB	32	75 (30min)	18–80	257 Hz	2
Shaoxing Hospital Zhejiang DB	10,646	10,646 (10 s)	4–98	500 Hz	12
PF12RED	54	163 (10 s)	18–37	500 Hz	12

Table 2

Population characteristics and ECG basic parameters. SysBP, systolic blood pressure; DiaBP, diastolic blood pressure VR, ventricular rate; PQInt, PQ interval; QRSDur, QRS duration; QTInt, QT interval; QTCInt, QTC interval; RRInt, RR interval; PPInt, PP Interval. PAxis, RAxis, TAxis: angle of P, R and T in �.

	Age (y)		Height (cm)		Weight (kg)
Average	25.74		180.61		74.90
±SD	5.16		6.92		6.78

	SysBP (mmHg)	DiaBP (mmHg)	VR (bpm)	PQInt (ms)	QRSDur (ms)	QTInt (ms)	QTCInt (ms)	RRInt (ms)	PPInt (ms)
Average	118.45	77.18	54.33	169.98	101.32	434.01	408.85	1136.43	1129.32
±SD	9.86	5.02	9.60	38.71	8.01	28.08	23.00	192.04	191.01

	PAxis		RAxis		TAxis
Average (°)	52.09		67.63		40.96
±SD (°)	25.21		25.08		28.63

Table 3

Rhythm information and baseline characteristics of participants

Acronym Name	Full Name	Frequency, n (%)	Age, mean ± SD
SR	Sinus Rhythm	19 (35.19%)	24.58 ± 4.57
SB	Sinus Bradycardia	35 (64.81%)	26.37 ± 5.41
iRBBB	Incomplete Right Bundle Branch Block	11 (20.37%)	25.65 ± 5.76
RTWI	Right T Wave Inversion	36 (66.67%)	25.19 ± 5.22
LTWI	Left T Wave Inversion	1 (1.85%)	32
MTI (2°AVB)	2° grade Atrioventricular Block (Mobitz I)	1 (1.85%)	30
All	All	54 (100%)	25.74 ± 5.16

Table 4

Performance report for the 3 different ML techniques.

	Accuracy	Sensitivity	Specificity	Precision	F1-score
Random Forest	100%	100%	100%	100%	100%
SVM	96.078%	96.666%	95.238%	96.666%	96.666%
Neural Network	98.039%	100%	95.454%	96.666%	98.305%

Table 5

Characteristics of actual ECG management software.

| Software	Input	Leads	Filtering		Peaks		Features		Free
			Auto	Manual	Auto	Manual	Auto	Manual	
Edelman et al. (2019)	.mat	1	Yes	No	Yes	No	Yes	No	Yes
Encord ECG (2023)	DICOM	12	No	No	Yes	No	No	No	No
OHIF ECG Viewer (2023)	DICOM	12	No	No	Yes	No	No	No	Yes
Waveform ECG (2008)	.xml	12	No	No	Yes	No	No	No	Yes
ECG Viewer (2022)	.dat, .txt, .csv	12	Yes	Yes	Yes	No	No	No	Yes
ECG Visualizer (2023)	.xml	12	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Appendix B. Figures
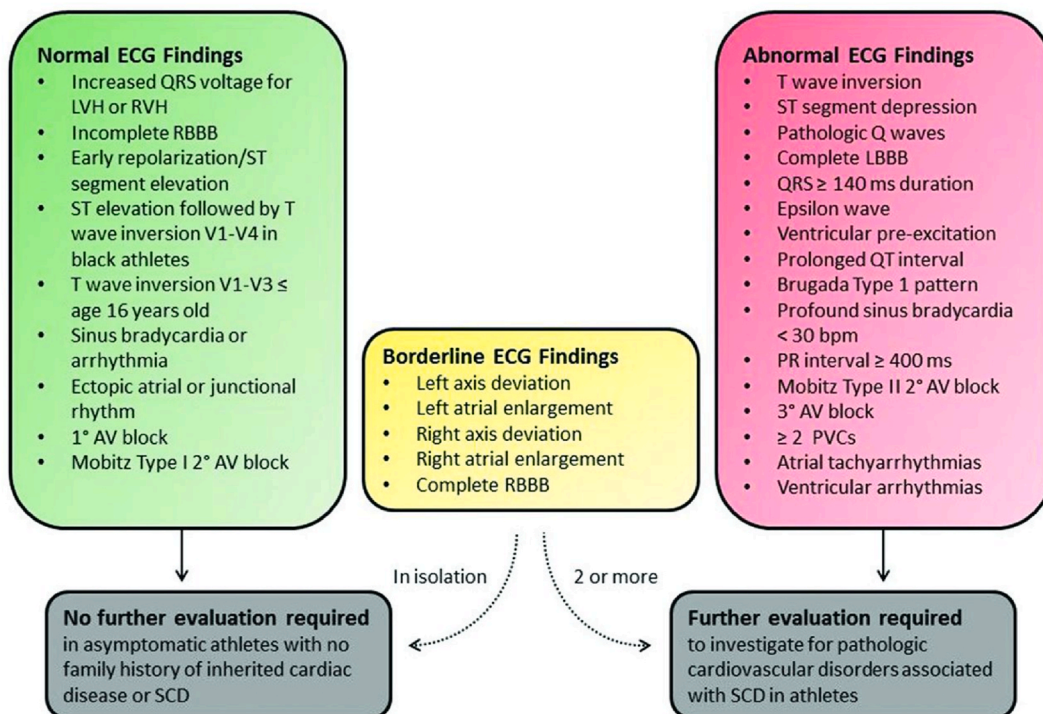
Fig. 1. International consensus standards for ECG interpretation in athletes [3]. AV, atrioventricular; LBBB, left bundle branch block; LVH, left ventricular hypertrophy; PVC, premature ventricular contraction; RBBB, right bundle branch block; RVH, right ventricular hypertrophy; SCD, sudden cardiac death. Drezner et al., 2017 https://doi.org/10.1136/bjsports-2016-097331 License number CCC: 5,630,411,149,549. Authorized by BMJ per re-use Figure.
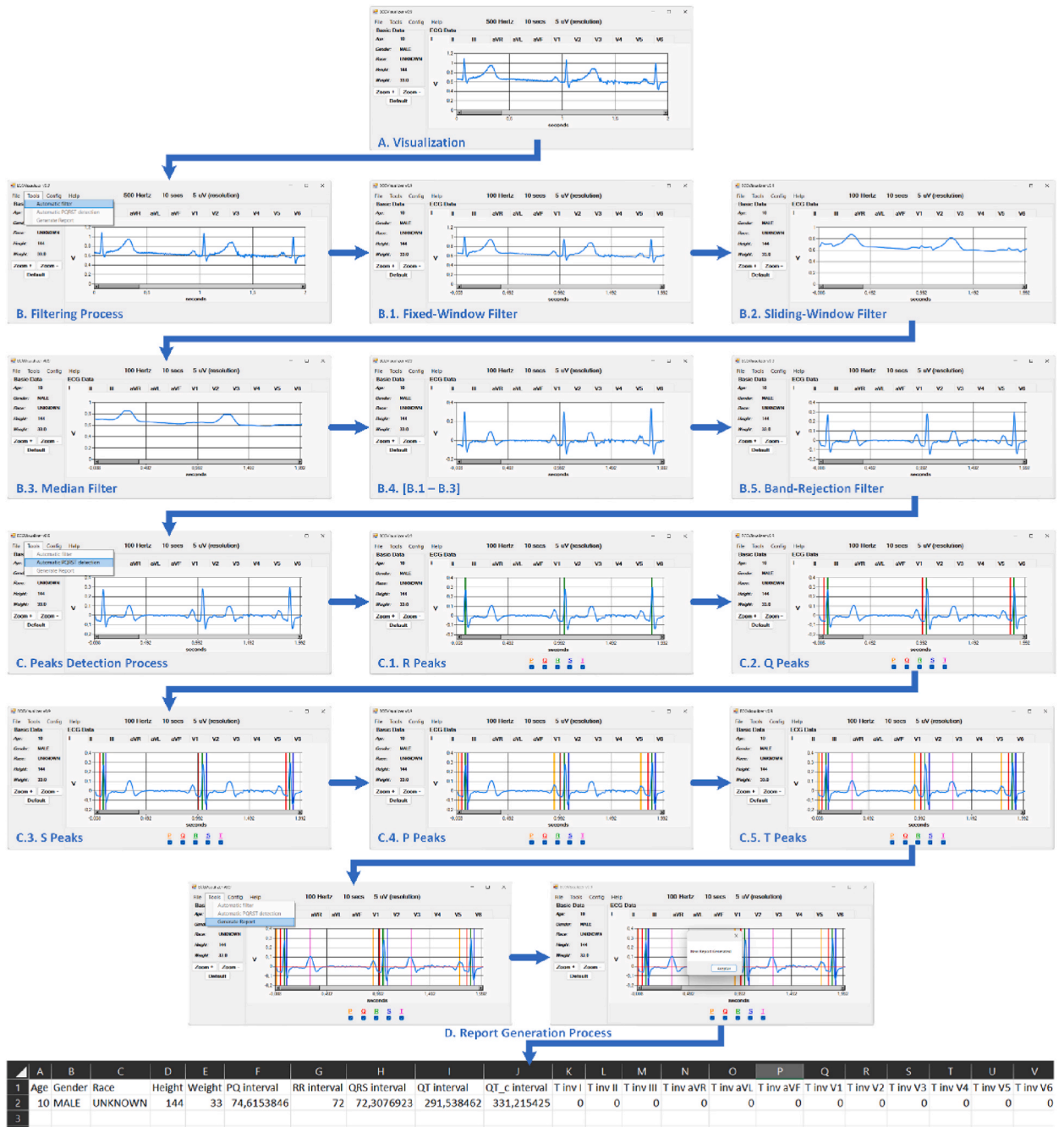
Fig. 2. Software tool usage diagram.
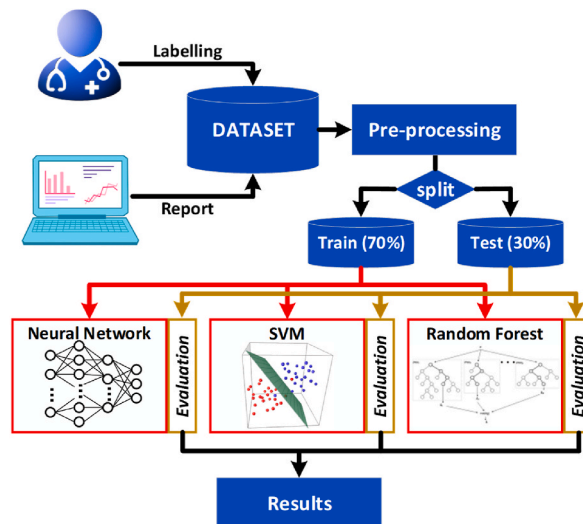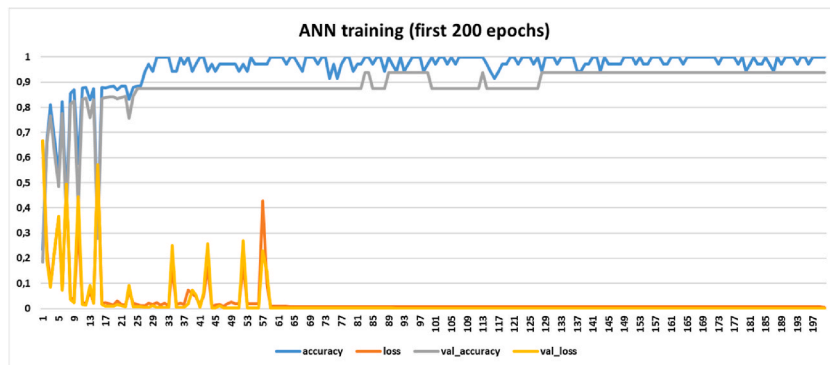
Fig. 3. Classifiers evaluation process.

Fig. 4. Loss Curves Inclusion, showing training and testing from the model's performance and learning process over time.
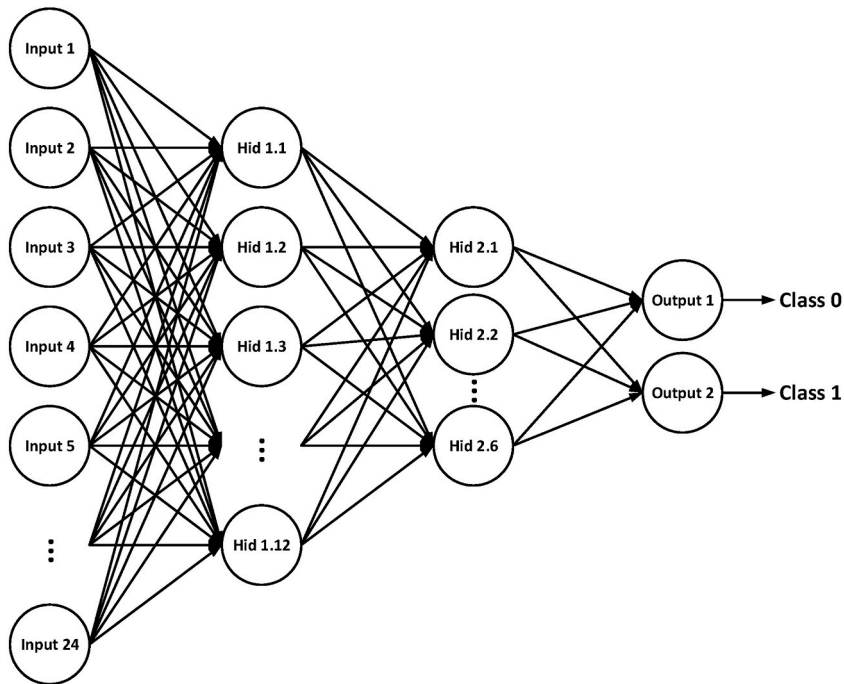
Fig. 5. Visual Representation of the ANN: A diagrammatic representation of the ANN would enhance understanding, particularly for readers who are more visually inclined.
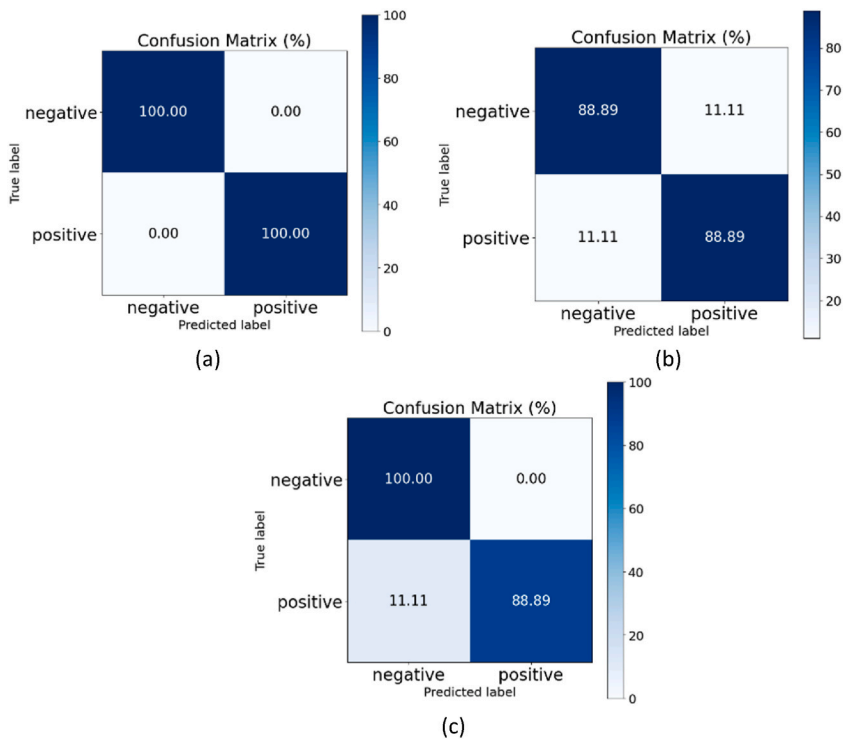

Fig. 6. Confusion matrices for a) Random Forest; b) Support Vector Machine; and c) Neural Network
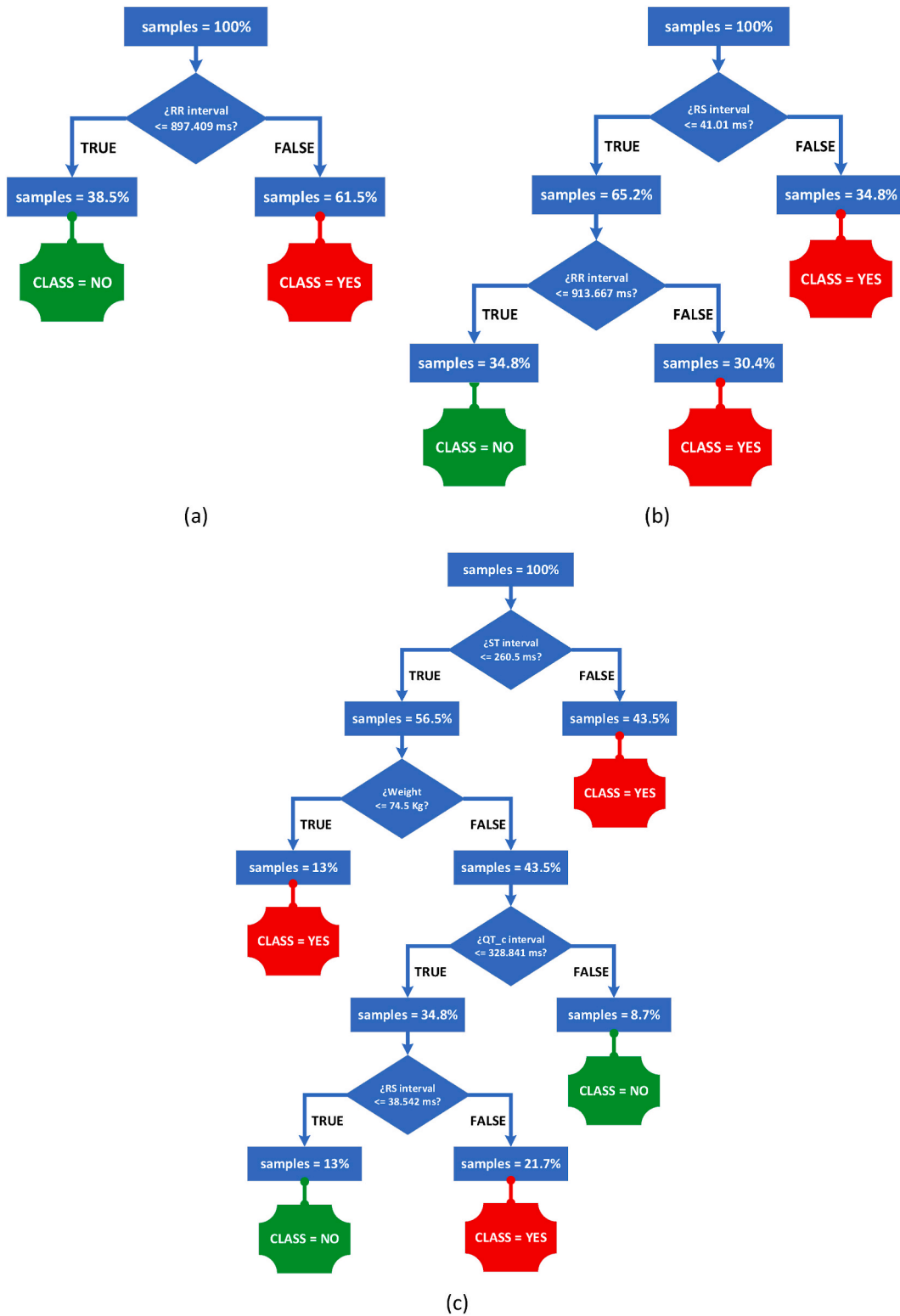
Fig. 7. Parameters used by each tree: 'a' only uses the RR interval parameter; 'b' uses RS and RR intervals; 'c' uses RS, RR and corrected QT intervals with user Weight.

References

[1] J. Pan, W.J. Tompkins, A real-time QRS detection algorithm, IEEE Trans Biomed Eng BME- 32 (1985) 230–236, https://doi.org/10.1109/TBME.1985.325532.

[2] P. Rajpurkar, A.Y. Hannun, M. Haghpanahi, C. Bourn, A.Y. Ng, Cardiologist-Level Arrhythmia Detection with Convolutional Neural Networks, Htttps://Arxiv. Org/Pdf/1707.01836.Pdf, 2017 (https://stanfordmlgroup).

[3] J.A. Drezner, S. Sharma, A. Baggish, M. Papadakis, M.G. Wilson, J.M. Prutkin, A. La Gerche, M.J. Ackerman, M. Borjesson, J.C. Salerno, I.M. Asif, D.S. Owens, E. H. Chung, M.S. Emery, V.F. Froelicher, H. Heidbuchel, C. Adamuz, C.A. Asplund, G. Cohen, K.G. Harmon, J.C. Marek, S. Molossi, J. Niebauer, H.F. Pelto, M. V. Perez, N.R. Riding, T. Saarel, C.M. Schmied, D.M. Shipon, R. Stein, V.L. Vetter, A. Pelliccia, D. Corrado, International criteria for electrocardiographic interpretation in athletes: consensus statement, Br. J. Sports Med. 51 (2017) 704–731, https://doi.org/10.1136/bjsports-2016-097331.

[4] II - Medical examination of players • UEFA Medical Regulations • Lector • Documents UEFA, (n.d.). https://documents.uefa.com/r/e_a_ 0zs~8Ut55Hay0CW8yQ/ir0ZJfBWq_wZK2aVPEea7A.

[5] F. Egger, J. Scharhag, A. Kästner, J. Dvořak, P. Bohm, T. Meyer, FIFA Sudden Death Registry (FIFA-SDR): a prospective, observational study of sudden death in worldwide football from 2014 to 2018, Br. J. Sports Med. 56 (2022) 80–87, https://doi.org/10.1136/bjsports-2020-102368.

[6] G.B. Moody, R.G. Mark, The impact of the MIT-BIH arrhythmia database, IEEE Eng. Med. Biol. Mag. 20 (2010) 45–50. http://www.ncbi.nlm.nih.gov/pubmed/ 11446209.

[7] A.L. Goldberger, L.A.N. Amaral, G. Leon, J.M. Hausdorff, P.C. Ivanov, R.G. Mark, J.E. Mietus, G.B. Moody, C.-K. Peng, H.E. Stanley, PhysioBank, PhysioToolkit, and PhysioNet components of a new research resource for complex physiologic signals, Circulation 101 (2000) e215–e220, https://doi.org/10.1161/01. CIR.101.23.e215.

[8] A.A. Munoz-Macho, M.J. Dominguez-Morales, J.L. Sevillano-Ramos, Pro Football 12-lead Resting Electrocardiogram Database (PF12RED), 2023. Https:// Github.Com/Dradolfomunoz/PF12RED.

[9] M.A. Mansournia, G.S. Collins, R.O. Nielsen, M. Nazemipour, N.P. Jewell, D.G. Altman, M.J. Campbell, A CHecklist for statistical assessment of medical papers (the CHAMP statement): explanation and elaboration, Br. J. Sports Med. 55 (2021) 1009–1017, https://doi.org/10.1136/bjsports-2020-103652.

[10] S.J. Pocock, J.P. Vandenbroucke, Strengthening the reporting of observational studies in epidemiology (StroBE) statement: guidelines for reporting observational studies, BMJ 335 (2007) 806–808, https://doi.org/10.1136/bmj.39335.541782.AD.

[11] M. Dominguez-Morales, A.A. Munoz-Macho, J. Sevillano-Ramos, ECG Visualizer Software Tool, 2023. Https://Github.Com/Mjdominguez/ECGVisualizer.

[12] B.A. Shenoi, Introduction to Digital Signal Processing and Filter Design, John Wiley and Sons, 2005, https://doi.org/10.1002/0471656372.

[13] H.C. Chen, S.W. Chen, A moving average based filtering system with its application to real-time QRS detection, Comput. Cardiol. 30 (2003) 585–588, https:// doi.org/10.1109/CIC.2003.1291223.

[14] B.I. Justusson, Two-Dimensional Digital Signal Processing II, Springer-Verlag, 2006, https://doi.org/10.1007/BFB0057592.

[15] H. Zumbahlen, Basic Linear Design, Basic linear design., 2007.

[16] S. Sharma, J.A. Drezner, A. Baggish, M. Papadakis, M.G. Wilson, J.M. Prutkin, A. La Gerche, M.J. Ackerman, M. Borjesson, J.C. Salerno, I.M. Asif, D.S. Owens, E. H. Chung, M.S. Emery, V.F. Froelicher, H. Heidbuchel, C. Adamuz, C.A. Asplund, G. Cohen, K.G. Harmon, J.C. Marek, S. Molossi, J. Niebauer, H.F. Pelto, M. V. Perez, N.R. Riding, T. Saarel, C.M. Schmied, D.M. Shipon, R. Stein, V.L. Vetter, A. Pelliccia, D. Corrado, International recommendations for electrocardiographic interpretation in athletes, J. Am. Coll. Cardiol. 69 (2017) 1057–1075, https://doi.org/10.1016/j.jacc.2017.01.015.

[17] School of Health Sciences (The University of Nottingham), A Beginners Guide to Normal Heart Function, Sinus Rhythm & Common Cardiac Arrhythmias, 2023. Https://Www.Nottingham.Ac.Uk/Nursing/Practice/Resources/Cardiology/Function/Normal_duration.Php.

[18] R.A.A. Bellfield, S. Ortega-Martorell, G.Y.H. Lip, D. Oxborough, I. Olier, The athlete's heart and machine learning: a review of current implementations and gaps for future research, J Cardiovasc Dev Dis 9 (2022), https://doi.org/10.3390/jcdd9110382.

[19] T. Kam Ho, Random decision forests, in: Proceedings of 3rd International Conference on Document Analysis and Recognition, vol. 1, 1995, pp. 278–282.

[20] L. Breiman, Random forests, Mach. Learn. 45 (2001) 5–32, https://doi.org/10.1023/A:1010933404324.

[21] B.E. Boser, I.M. Guyon, V.N. Vapnik, Training algorithm for optimal margin classifiers, Proceedings of the Fifth Annual ACM Workshop on Computational Learning Theory (1992) 144–152, https://doi.org/10.1145/130385.130401.

[22] W.S. Noble, What is a support vector machine? Nat. Biotechnol. 24 (2006) 1565–1567, https://doi.org/10.1038/nbt1206-1565, 12 24 (2006).

[23] J. Schmidhuber, Deep learning in neural networks: an overview, Neural Network. 61 (2014) 85–117, https://doi.org/10.1016/j.neunet.2014.09.003.

[24] S. Russell, P. Norivg, Artificial intelligence: a modern approach, in: Artificial Intelligence: A Modern Approach, Global Edition, 2021.

[25] O. Ozmen Garibay, B. Winslow, S. Andolina, M. Antona, A. Bodenschatz, C. Coursaris, G. Falco, S.M. Fiore, I. Garibay, K. Grieman, J.C. Havens, M. Jirotka, H. Kacorri, W. Karwowski, J. Kider, J. Konstan, S. Koon, M. Lopez-Gonzalez, I. Maifeld-Carucci, S. McGregor, G. Salvendy, B. Shneiderman, C. Stephanidis, C. Strobel, C. Ten Holter, W. Xu, Six human-centered artificial intelligence grand challenges, Int. J. Hum. Comput. Interact. 39 (2023) 391–437, https://doi.org/ 10.1080/10447318.2022.2153320.

[26] M.E. Cohen, D.L. Hudson, Neural network models for biosignal analysis, in: 2006 International Conference of the IEEE Engineering in Medicine and Biology Society, IEEE, 2006, pp. 3537–3540, https://doi.org/10.1109/IEMBS.2006.260393.

[27] G. Nassis, E. Verhagen, J. Brito, P. Figueiredo, P. Krustrup, A review of machine learning applications in soccer with an emphasis on injury risk, Biol. Sport (2023), https://doi.org/10.5114/biolsport.2023.114283.

[28] A. Rossi, L. Pappalardo, P. Cintia, A narrative review for a machine learning application in sports: an example based on injury forecasting in soccer, Sports 10 (2022), https://doi.org/10.3390/sports10010005.

[29] N.F.N. Bittencourt, W.H. Meeuwisse, L.D. Mendonça, A. Nettel-Aguirre, J.M. Ocarino, S.T. Fonseca, Complex systems approach for sports injuries: moving from risk factor identification to injury pattern recognition—narrative review and new concept, Br. J. Sports Med. 50 (2016) 1309–1314, https://doi.org/10.1136/ bjsports-2015-095850.

[30] A. López-Valenciano, F. Ayala, J.M. Puerta, M.B.A. De Ste Croix, F.J. Vera-Garcia, S. Hernández-Sánchez, I. Ruiz-Pérez, G.D. Myer, A preventive model for muscle injuries: a novel approach based on learning algorithms, Med. Sci. Sports Exerc. 50 (2018) 915–927, https://doi.org/10.1249/MSS.0000000000001535.

[31] C. Carling, F. le Gall, G. Dupont, Analysis of repeated high-intensity running performance in professional soccer, J. Sports Sci. 30 (2012) 325–336, https://doi. org/10.1080/02640414.2011.652655.

[32] J. Ekstrand, Keeping your top players on the pitch: the key to football medicine at a professional level, Br. J. Sports Med. 47 (2013), https://doi.org/10.1136/ bjsports-2013-092771.

[33] A. Dellal, C. Lago-Peñas, E. Rey, K. Chamari, E. Orhant, The effects of a congested fixture period on physical performance, technical activity and injury rate during matches in a professional soccer team, Br. J. Sports Med. 49 (2015) 390–394, https://doi.org/10.1136/bjsports-2012-091290.

[34] J. Ekstrand, Playing too many matches is negative for both performance and player availability - results from the on-going UEFA injury study, German Journal Of Sports Medicine 64 (2013) 5–9, https://doi.org/10.5960/dzsm.2012.038.

[35] J. Calleja-Gonzalez, D. Marques-Jimenez, M. Jones, T. Huyghe, F. Navarro, A. Delextrat, I. Jukic, S. Ostojic, J. Sampaio, X. Schelling, P. Alcaraz, F. Sanchez-Bañuelos, X. Leibar, J. Mielgo-Ayuso, N. Terrados, What are we doing wrong when athletes report higher levels of fatigue from traveling than from training or competition? Front. Psychol. 11 (2020) https://doi.org/10.3389/FPSYG.2020.00194.

[36] A. Rossi, L. Pappalardo, P. Cintia, F.M. Iaia, J. Fernàndez, D. Medina, Effective injury forecasting in soccer with GPS training data and machine learning, PLoS One (2018) 1–15, https://doi.org/10.1371/journal.pone.0201264.

[37] E. Adetiba, V.C. Iweanya, S.I. Popoola, J.N. Adetiba, C. Menon, Automated detection of heart defects in athletes based on electrocardiography and artificial neural network, Cogent Eng 4 (2017), https://doi.org/10.1080/23311916.2017.1411220.

[38] P. Christ, U. Rückert, Identification of athletes during walking and jogging based on gait and electrocardiographic patterns, Communications in Computer and Information Science 452 (2014) 240–257, https://doi.org/10.1007/978-3-662-44485-6_17.

[39] M. Laurino, A. Piarulli, R. Bedini, A. Gemignani, A. Pingitore, A. L'Abbate, A. Landi, P. Piaggi, D. Menicucci, Comparative study of morphological ECG features classificators: an application on athletes undergone to acute physical stress, in: International Conference on Intelligent Systems Design and Applications, ISDA, 2011, pp. 242–246, https://doi.org/10.1109/ISDA.2011.6121662.

[40] J. Zheng, J. Zhang, S. Danioko, H. Yao, H. Guo, C. Rakovski, A 12-lead electrocardiogram database for arrhythmia research covering more than 10,000 patients, Sci. Data 7 (2020) 1–8, https://doi.org/10.1038/s41597-020-0386-x, 1 7 (2020).

[41] J.G. Claudino, D. de O. Capanema, T.V. de Souza, J.C. Serrão, A.C. Machado Pereira, G.P. Nassis, Current approaches to the use of artificial intelligence for injury risk assessment and performance prediction in team sports: a systematic review, Sports Med Open 5 (2019), https://doi.org/10.1186/s40798-019-0202-3.

[42] C.S. Harris, V.F. Froelicher, D. Hadley, M.T. Wheeler, Guide to the female student athlete ECG: a comprehensive study of 3466 young, racially diverse athletes, Am. J. Med. 135 (2022) 1478–1487.e4, https://doi.org/10.1016/j.amjmed.2022.07.013.

[43] O. Fabregat-Andrés, A. Muñoz-Macho, G. Adell-Beltrán, L. Fácila, Feasibility of using a new generation wireless device for electrocardiographic monitoring of professional soccer players during an exercise test in field, J. Sports Med. Phys. Fit. 55 (2015) 1593–1595, https://pubmed.ncbi.nlm.nih.gov/25069964/.

Munoz-Macho, Adolfo Antonio (https://orcid.org/0000-0002-9133-4860A sport and exercise medical doctor, he has accumulated substantial experience working as a performance medical director for several years with professional football teams, including UEFA Europa League level teams. . His scientific concerns are high performance, neuromusculoskeletal system, exercise biology and physiology and sudden death prevention in sports. Presently the medical and performance director at RCD Mallorca SAD. Doctoral candidate at Seville University.

Dominguez-Morales, Manuel Jesus (https://orcid.org/0000-0001-5669-9111) Computer Engineer, M.Sc. in Software Engineering and Technology, M.Sc. in Computer and Networks Engineering, Ph.D. in Industrial Informatics. Associate Professor at Computer Architecture and Technology Department (University of Seville, Spain). Research focused on intelligent embedded devices, e-health, physiological signal processing and AI diagnostic-aid systems.

Sevillano-Ramos, José Luis https://orcid.org/0000-0002-1392-1832 received in 1993 a Ph.D. from the University of Seville, Spain, where he is currently a Full Professor of computer technology and architecture. In the period 2014-2022 he has been Director of the School of Computer Engineering at the University of Seville, Spain. He is also Associate Editor of the "International Journal of Communication Systems" (Wiley) and of "Simulation: Transactions of The Society for Modeling and Simulation International" (Sage). In the period 2009-2011 he served as Vice-President of Membership and member of the Executive Committee of The Society for Modelling and Simulation International (SCS). He has also served in different roles on several international conferences. He has published more than 100 articles in international journals and conferences. His research interests include real-time communications and architectures, Mobile Robots and eHealth and Rehabilitation Systems.