ECG Pre-processing and Feature Extraction Tool for Intelligent Simulation Systems

Manuel Domínguez-Morales[1,2(✉)], Adolfo Muñoz-Macho[3], and José L. Sevillano[1,2]

[1] Architecture and Computer Technology Department, E.T.S. Ingeniería Informática, Universidad de Sevilla, Sevilla, Spain
mjdominguez@us.es
[2] Computer Science Research Institute, Universidad de Sevilla, Sevilla, Spain
[3] Performance and Medical Department at RCD Mallorca SAD, Palma de Mallorca, Spain

Abstract. Sudden cardiac death events and fatal cardiac problems are a field of vital importance for physicians working with elite athletes. For this reason, it is common to periodically perform cardiac monitoring with professional ECG devices to detect certain risk markers. As these doctors often work with many athletes (as is the case with professional football teams), an artificial intelligence-based system would help mass screening and allow these exams to be carried out more regularly. Because physicians often evaluate the printed reports generated by ECG devices, few manufacturers provide powerful and configurable software tools. Moreover, for teaching purposes, a simulation tool that would allow working with previously collected ECG files would be very useful. In this paper, we present a software tool to be used with General Electric CardioSoft 12SL electrocardiograph. This tool allows importing the XML files generated by this device, perform a manual or automatic signal filtering process and PQRST peak detection, and finally generate a customisable report as a CSV file containing the features obtained after signal analysis. This pre-processed information can be used as input of ECG simulators and in artificial intelligence systems to develop diagnostic support systems.

Keywords: ECG · Signal processing · Feature extraction · Report generation · Artificial intelligence

1 Introduction

Concerned about cardiac problems in athletes and the sudden deaths caused by them, international bodies such as the *Fédération Internationale de Football Association* (FIFA) in the case of football have made an enormous effort to carry

J.-L. Guisado-Lizar et al. (Eds.): SIMUtools 2023, LNICST 519, pp. 209–216, 2024.
https://doi.org/10.1007/978-3-031-57523-5_16

out mass screening of athletes with the use of electrocardiograms (ECG) [11]. According to those tests, sport physicians and cardiologists review ECG Data and determine the players' aptitude [5]. This assessment is performed with professional 12-lead electrocardiographs, which provide the information in printed form (which is the most commonly used), but also in an XML file.

However, due to the high number of professional athletes and the complexity of analysing all the peaks from the 12 leads of an ECG signal, these tests are not carried out many times a year (only one time a year in professional football teams). This fact contrasts with the devastating results of a sudden death, which could require a great precision, and maybe it could be better diagnosed with more frequent screenings.

On the other hand, the use of artificial intelligence (AI) in the field of cardiac disease detection for elite professionals may help to reduce the time needed to detect harmful markers, as has been demonstrated with other physiological signals [7,9]. Therefore, prior to the use of this type of AI systems, the processes of pre-processing and extraction of useful features are essential. Moreover, it is also necessary to provide a simulation tool to teach medical professionals how to detect dangerous markers (in this regard, there are other simulations tools like [6], centred on ECG signals). In addition, such a software tool could help in preprocessing and visualization of ECG signals in a more user-friendly and usable way; or in research, for example, to generate synthetic signals that can be used to enrich databases [8].

For this reason, a new software tool called *ECGVisualizer* is presented. This tool allows loading XML files obtained from General Electric CardioSoft 12SL electrocardiogram, visualising the 12-lead information, filtering the signals, performing PQRST peaks detection, extracting features and generating a summary report. Some manufacturers provide very closed visualization tools without customisation and, in other cases, generic visualization tools are used that do not work properly for all devices.

This tool includes two main features: *1) full customization*, as user can decide what filters to apply, the type of patient for the peaks' detection and what features to be extracted from the signals; and *2) report generation* in a CSV open format, according to the features selected and the baseline information of the patient. This report can eventually be used as input of EGC simulators or in AI-based diagnostic support systems. *ECGVisualizer* can be downloaded from the next URL: https://github.com/mjdominguez/ECGVisualizer.

The objective of this paper is to present the software tool and, using a dataset collected and processed with this tool, to show an example of its use and the results obtained. The effectiveness of this tool has been verified by a cardiac physician.

This paper is organized as follows: firstly, the software tool and the collected dataset are presented; secondly, a use case of this tool is shown with several ECGs obtained from professional soccer players; then, the characteristics of this tool are compared with other similar tools and, finally, the conclusions of this work are shown.

2 Materials

First, the software tool developed (ECGVisualizer) and its functionalities are presented, followed by a summary of the dataset collected for testing with this tool.

2.1 ECGVisualizer Tool

The ECG Visualizer software is a tool that allows to load XML files generated by a professional 12-lead ECG, visualise its content, apply customised filters on the signals and generate a report of the most significative features obtained from the signal. Currently, this tool supports the electrocardiograph CardioSoft 12SL from General Electric, but soon it will support also the Contec 1200G device.

The whole processing chain is fully configurable by the user, allowing to indicate which filters to apply and when, the characteristics of the patient whose ECG we are processing and the characteristics that we want to be included in the final report.

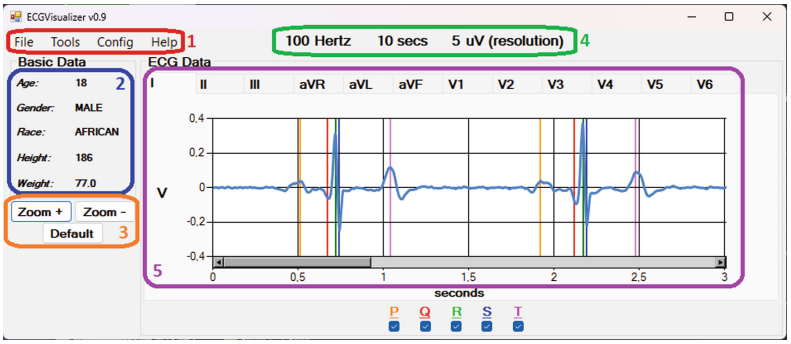

Fig. 1. ECGVisualizer main form.

Software Architecture. This tool has been developed with Visual Studio IDE, using C# language and Windows Forms. In addition to the common operations, all signal filters and peak detection algorithms have been implemented, including some utilities to operate with signals. Finally, regarding the detection of the ECG signal peaks, this task is performed taking into account the maximum and minimum ranges of the usual intervals (QRS, QT and PR) stipulated for the configured patient type. The main form of this tool is presented in Fig. 1. According to it, the tool is divided in:

– Section 1 (red) - Toolbar: options that allow the user loading, filtering, selecting features and configuring the report generated from the ECG file loaded. See the GitHub User Manual[1].
– Section 2 (blue) - Basic patient information: baseline information obtained from the ECG file.
– Section 3 (orange) - Visualization tools: tools that allow to vary zoom.
– Section 4 (green) - Signal information: informative section, where the main characteristics of the loaded ECG signal are shown.
– Section 5 (purple) - ECG signal visualization: the user may select which lead to visualize. The result of the processing is represented here too.

The tool allows an automatic processing (default filter cascade) and manual processing (user can select what filters and when). The usual processing chain is presented in the central part of Fig. 2.

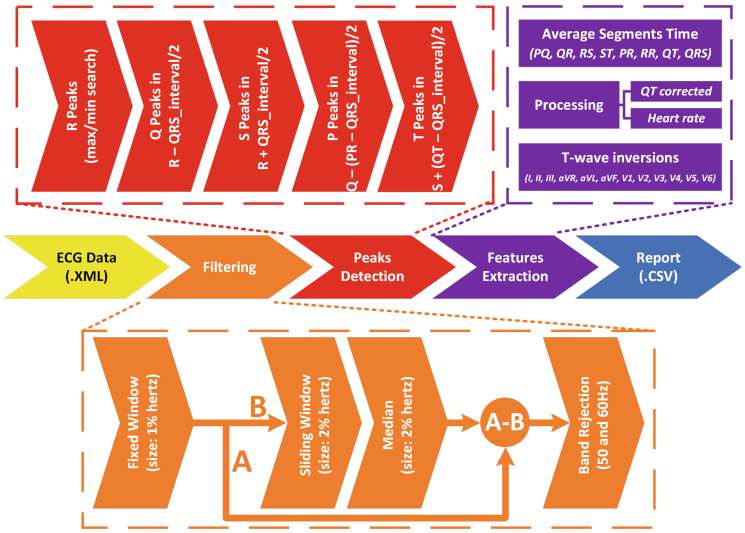

Fig. 2. Summary of the automatic processing chain.

Software Functionalities. The functionalities of this tool are described next:

– Load: XML files from GE CardioSoft 12SL ECG can be loaded. These files include the patient's baseline information and the RAW information of the 12 leads (collected at 500 hz).

[1] https://github.com/mjdominguez/ECGVisualizer/tree/main/ECGVisualizer/documentation.

- Visualization: the user can navigate between the 12 leads within the time period. The user can also zoom in and out on those areas he/she wants to enlarge or reduce.
- Filtering: it allows for automatic filtering (Fig. 2-down) and manual filtering. The configurable filters included are:
  - Fixed-window average filter (mean/smoothing filter). It removes the digital noise from the sampling and, as a side effect, reduces the sampling rate.
  - Sliding-window average filter (moving average filter). It removes the high and low-frequency ripples.
  - Sliding-window median filter (moving median filter). It flattens deeply the signal peaks and reduces the noise.
  - Band rejection filter (band-stop filter). It is used to eliminate the persistent noise caused by the electrocardiograph mains connection.
- Peaks detection: it allows selecting the target patient between a common person, an athlete or a custom class [10,11]. The detailed process is described in the upper-left part of Fig. 2.
- Feature extraction & Report Generation: it allows configuring the features included in the report (see Fig. 2 upper-right for the full features list). The result is a CSV file with each feature in a different column.

2.2 Dataset

The dataset collected is named PF12RED[2], including 163 raw ECG data in XML format from 54 football players. It was collected from professional UEFA football players of a team from La Liga EA SPORTS, taken by the co-author *Dr. Adolfo Muñoz-Macho*. It is part of a project focused on using AI in professional football teams for diagnosis aid, inscribed in ClinicalTrials.gov (No. NCT05872945).

3 Results

This section will show a use case of the tool. After that, a comparison with other ECG signal visualization and processing tools is performed.

3.1 Illustrative Example

The tool was configured with the automatic filtering process, patient type "athlete", and with the default features. The result of the peak detection process and the report generated can be seen in Fig. 3.

The tool shows the baseline data on the left and the ECG signals on the right of Fig. 3 top-left, including the peaks detected. The report generated is shown at the bottom of Fig. 3, including a T-wave inversion in aVL lead, which can be visually verified by looking closely at aVL lead (see Fig. 3 top-right). This process

[2] https://github.com/dradolfomunoz/PF12RED.

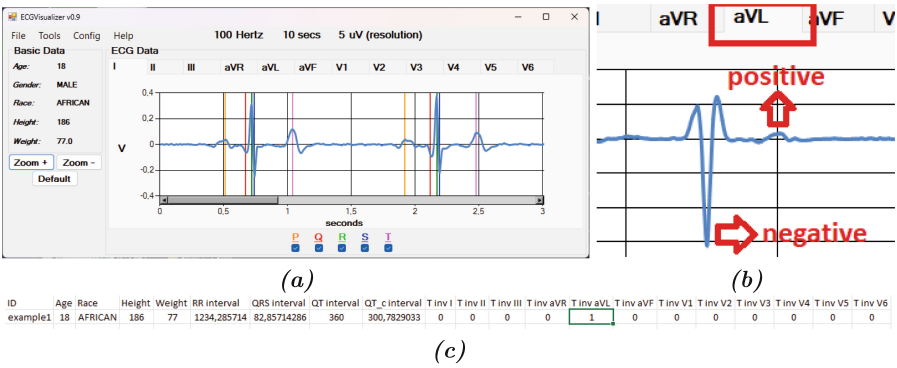

Fig. 3. Visual representation of an ECG processing: (a) peaks detection result, (b) T-wave inversion in aVL, (c) report generated.

was repeated for the whole dataset and verified by *Dr Adolfo Muñoz-Macho*. Results show an accuracy of almost 100% compared with the markers obtained manually: the few discrepancies obtained relies on the confusion between peak T and peak U in some signals (where peak U is an unusual wave, like a T-wave echo).

3.2 Tools Comparison

A search for software tools capable of visualizing ECG information was carried out. A summary of this information can be seen in Table 1.

Table 1. ECG visualization, simulation and processing tools.

Software Tool	Input	Leads	Filtering		Peaks		Features		Free
			Auto	Manual	Auto	Manual	Auto	Manual	
Edelmann et al. (2019) [6]	.mat	1	Yes	No	Yes	No	Yes	No	Yes
Encord ECG (2023) [2]	DICOM	12	No	No	Yes	No	No	No	No
OHIF ECG Viewer (2023) [3]	DICOM	12	No	No	Yes	No	No	No	Yes
Waveform ECG (2008) [4]	.xml	12	No	No	Yes	No	No	No	Yes
ECG-Viewer (2022) [1]	.dat, .txt, .csv	12	Yes	Yes	Yes	No	No	No	Yes
ECGVisualizer (2023) [this work]	.xml	12	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Some of the tools found use DICOM format and, because of this, they cannot perform advanced processing on the information, only visualization and manual

labelling. However, it is common that the tools found do not allow filtering and feature extraction: apart from our tool, only two perform some processing:

- Edelmann et al. (2019) [6]: this is the most similar tool compared to ours. However, it has two drawbacks: first, it only allows visualising one lead at a time (and feature extraction is only performed to it); and, second, the input format is specific from MATLAB, requiring the data to be pre-processed with the MATLAB Toolkit before using the software.
- ECG-Viewer (2022) [1]: this is the most widely used, as it allows several open-source input formats and several customisations. However, it has two major drawbacks: first, it does not allow configuration in the peak detection process (not being suitable for athletes); and, second, it does not extract features from the segments and waves (it only includes a general report).

So, our tool has features that are not found in others. Mainly, it is worth highlighting the high degree of customisation of our tool, which allows it to be adapted to the type of user and to select the characteristics that we want to extract from the signals. On the downside, as we use the format of a specific ECG device, we are limited to the information obtained with this type of device.

4 Conclusions

Medical professionals need simulation tools to help them in their learning process. This applies to all areas, including the detection of cardiac problems.

When working with ECG data, it is necessary to pay attention to the PQRST peaks of each of the twelve leads. Moreover, within the sporting environment, the times between peaks and the characteristics to be observed vary in professional athletes.

Therefore, this work presents a free software tool that, thanks to its filtering and peak detection capability, serves as a simulation tool to evaluate the expertise of the future physician.

ECGVisualizer tool includes XMLs loading from a 12-lead ECG, visualization, filtering, feature extraction and fully customized report generation. The two main contributions of this tool are the total customisation of the filtering process and the feature extraction; and, secondly, the possibility of generating a report in a completely open and editable format (being able to join the reports of several patients in a single file).

The results demonstrate the correct feature detection and the future usefulness of this tool.

Declaration of Competing Interest

The authors declare that they have no competing interests that could have appeared to influence this work.

Acknowledgements. This work is part of the project SANEVEC TED2021-130825B-I00, funded by the Ministerio de Ciencia e Innovación (MCIN), Agencia Estatal de Investigación (AEI) of Spain, MCIN/AEI/10.13039/501100011033, and by the European Union NextGenerationEU/PRTR. We also want to thank the professional UEFA football players and the team from La Liga EA SPORTS involved in the collected dataset for allowing us to work with the information from their electrocardiograms.

References

1. ECG-Viewer. https://github.com/CBLRIT/ECG-Viewer. Accessed 29 Sept 2022
2. Encord ECG. https://encord.com/try-it-free/?&utm_campaign=cta-blog-medical-light. Accessed 29 Sept 2022
3. OHIF ECG Viewer. https://github.com/RadicalImaging/OHIFExtensionsAndModes/. Accessed 29 Sept 2022
4. Waveform ECG. https://www.cvrgrid.org/tools/waveform-ecg.html. Accessed 29 Sept 2022
5. Drezner, J., et al.: International criteria for electrocardiographic interpretation in athletes: consensus statement. Br. J. Sports Med. **51**(9), 704–731 (2017)
6. Edelmann, J.C., Mair, D., Ziesel, D., Burtscher, M., Ussmueller, T.: An ecg simulator with a novel ecg profile for physiological signals. J. Med. Eng. Technol. **42**(7), 501–509 (2018)
7. Escobar-Linero, E., Domínguez-Morales, M., Sevillano, J.L.: Worker's physical fatigue classification using neural networks. Expert Syst. Appl. **198**, 116784 (2022)
8. Gillette, K., et al.: Medalcare-xl: 16,900 healthy and pathological synthetic 12 lead ecgs from electrophysiological simulations. Sci. Data **10**(1), 531 (2023)
9. Muñoz-Saavedra, L., Luna-Perejón, F., Civit-Masot, J., Miró-Amarante, L., Civit, A., Domínguez-Morales, M.: Affective state assistant for helping users with cognition disabilities using neural networks. Electronics **9**(11), 1843 (2020)
10. School of Health Sciences (The University of Nottingham): A beginners guide to normal heart function, sinus rhythm & common cardiac arrhythmias. https://www.nottingham.ac.uk/nursing/practice/resources/cardiology/function/normal_duration.php. Accessed 1 June 2023
11. Sharma, S., et al.: International recommendations for electrocardiographic interpretation in athletes. Eur. heart J. **39**(16), 1466–1480 (2018)