Biomedical time series preprocessing and expert-system based feature extraction in MULTISAB platform

Alan Jovic*, Davor Kukolja*, Kresimir Friganovic*, Kresimir Jozic**, and Sinisa Car***
* University of Zagreb Faculty of Electrical Engineering and Computing, Zagreb, Croatia
** INA - industrija nafte, d.d., Zagreb, Croatia
*** Dom zdravlja Dubrovnik, Ambulanta Babino polje, Babino polje, Croatia
Corresponding author: alan.jovic@fer.hr

Abstract - In this paper, we review the current state of implementation of the MULTISAB platform, a web platform whose main goal is to provide a user with detailed analysis capabilities for heterogeneous biomedical time series. These time series are often encumbered by noise that prohibits accurate calculation of clinically significant features. The goal of preprocessing is either to completely remove the noise or at least to ameliorate the quality of the recorded series. The focus of this paper is on the description of an expert feature recommendation system for electrocardiogram analysis. We demonstrate the process through which one arrives at a point where significant expert features are proposed to a platform user, based on time series at hand, analysis goal, and available length of the time series. We also provide the description of implemented preprocessing techniques and feature extraction procedure within the platform.

I. INTRODUCTION

The need for web-based biomedical software is continuously growing in the healthcare community [1,2]. Although, according to [1], web based solutions could, among many other things, reduce the cost of healthcare expenditure and facilitate access to a better healthcare, specialized software developed for medical professionals is vastly limited to continuous monitoring of biomedical time series (BTS) [3]. Hence, in the currently running Croatian Science Foundation research project HRZZ-MULTISAB[1], we pursue the development of an efficient and upgradeable BTS analysis system for automatic classification of human body disorders in a form of an integrative web platform. The aims of such a system are to help medical specialists in diagnostics and early detection of various diseases and to improve scientific investigation of BTS disorders.

In our previous work, we described the early efforts in designing such a system, which included an architectural overview, an overview of used technologies, detailed description of the MULTISAB project's frameworks [4] and use case based scenarios specification [3].

The architectural overview of the platform is shown in Fig. 1. MULTISAB project is divided into three subprojects: frontend on the client side; backend and processing subproject on the server side. The frontend and backend subprojects support communication between end users and servers, enabling a user (client) to send requests to the server computer through his web browser. Besides processing requests from users, the backend subproject serves for communication with a database and with data analysis hosts, on which processing subproject is running. The processing subproject is designed for heterogeneous BTS data analysis. It contains various frameworks for data handling, signal visualization, signal preprocessing, feature extraction, and data mining. Such web-based architecture provides users with the ability of distant access and enables biomedical data upload to the server, where efficient analysis is performed. Also, a web platform enables easier maintenance and a larger user base.

BTS data analysis process, described in detail in [3], is divided into 8 steps, some of which may be skipped, depending on the user: 1) analysis type selection, 2) scenario selection, 3) input data selection, 4) records inspection, 5) records preprocessing, 6) feature extraction, 7) model construction, and 8) reporting.

The aim of this paper is to review the current state of the MULTISAB platform implementation with the focus on the description of an expert feature recommendation system for ECG analysis and its initial implementation through predefined analysis scenarios. We also present the implemented preprocessing techniques and feature extraction procedure within the platform.

With the reported progress on the MULTISAB platform implementation, we have achieved the functionality of the first six analysis process steps (finishing with feature extraction). However, it should be noted that the current implementation is still in the test phase and that the user interface (UI) is in Croatian language. Nevertheless, we expect that in a few months a fully operational version will be available in Croatian and in English. The functionality will be limited to the first six steps, ending with the possibility of download of calculated features to a local file, thus enabling further independent data mining and reporting.

[1] This work has been fully supported by the Croatian Science Foundation under the project number UIP-2014-09-6889.

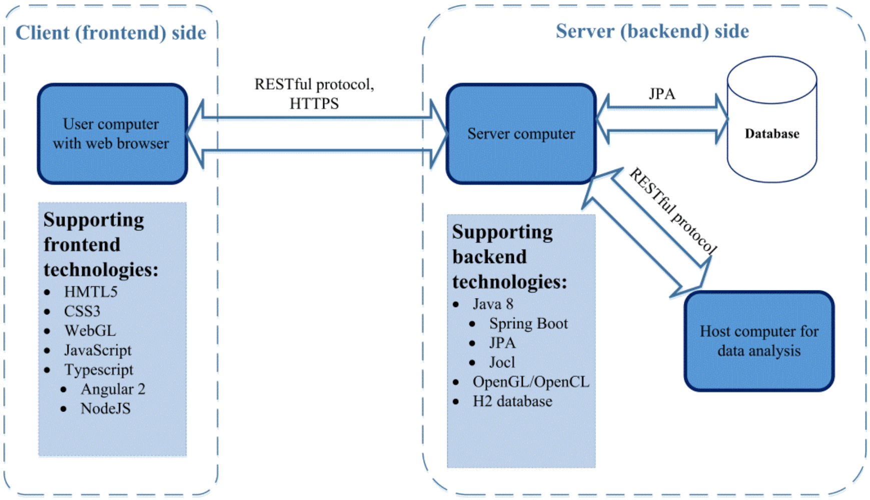

Figure 1. Architectural overview of the MULTISAB project web platform (modified from [4])

II. PREDEFINED ANALYSIS SCENARIOS WITH EXPERT FEATURES

Scenario selection UI, shown in Fig. 2, enables the user to select a predefined scenario ("Odabir spremljenih scenarija") and the construction of a completely new scenario through full customization ("Specifikacija novog scenarija"). A current list of predefined analysis scenarios includes: detection of acute myocardial ischemia ("Detekcija akutne miokardijalne ishemije"), detection of congestive heart failure ("Detekcija kongestivnog zatajenja srca"), and detection of atrial fibrillation ("Detekcija fibrilacije atrija") from ECG signals. The user can confirm a scenario with a button ("Dalje"). In later steps of the analysis, it is possible to change some of scenarios' aspects, such as preprocessing methods, feature extraction methods, or model construction methods. The predefined scenarios contain necessary preprocessing steps and expert features that are to be extracted from signals, which are selected by the medical expert system specifically designed for this purpose.

Implementation of the medical expert system is a continuously evolving process that gradually introduces new predefined scenarios. New scenarios are based on currently valid medical guidelines, standards, consultations with medical specialists, and relevant medical and biomedical engineering literature.

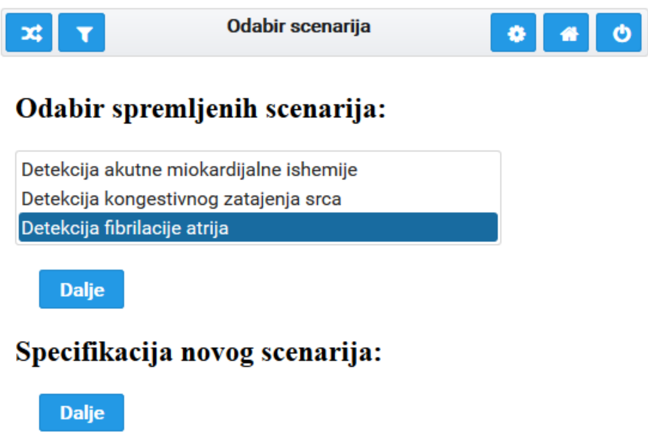

Figure 2. Scenario selection user interface; currently in Croatian language

The current initial version of the MULTISAB expert system, implemented in cooperation with cardiologists, includes the recommended features for detection of acute myocardial ischemia (AMI), atrial fibrillation (AF) and congestive heart failure (CHF) from ECG signals.

In the following subsections, we demonstrate the process of designing the predefined analysis scenarios with expert features for detection of specified disorders. To facilitate readers' understanding of the underlying terminology, we provide an image of characteristic ECG points and waves, with a particular focus on ST elevation, in Fig. 3 (courtesy of Rob Kruger and ECGpedia.org).

A. *Acute Myocardial Ischemia*

According to the medical guidelines [5], criteria for diagnosis of AMI (in absence of left ventricular hypertrophy and left bundle branch block) are:

1. ST elevation: New ST elevation at the J point in two contiguous leads with the cut-points: ≥0.1 mV in all leads other than leads V_2–V_3,

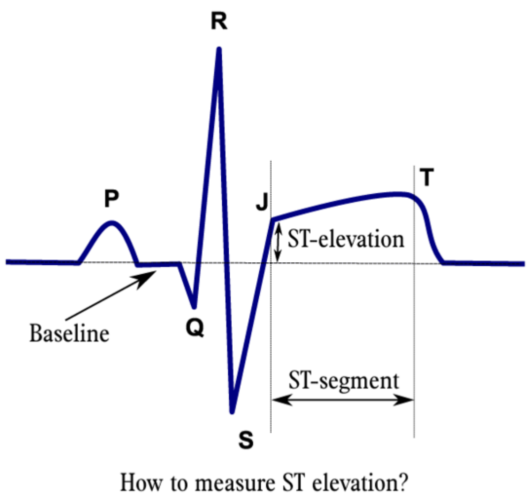

Figure 3. Characteristic ECG points and waves

where the following cut points apply: ≥0.2 mV in men ≥40 years; ≥0.25 mV in men <40 years, or ≥0.15 mV in women.

2. ST depression and T wave changes: New horizontal or down-sloping ST depression ≥0.05 mV in two contiguous leads and/or T inversion ≥0.1 mV in two contiguous leads with prominent R wave or R/S ratio >1.

From criteria for diagnosis of AMI, we can easily define expert morphological ECG features:

Morphological ECG features: ECG signal value (in regard to the baseline) at the J point, R/S ratio, T wave amplitude, and ST segment slope (for all leads).

According to guidelines [5], dynamic changes in the ECG waveforms during AMI episodes often require acquisition of multiple ECGs, particularly when the ECG at the initial presentation is non-diagnostic. Serial recordings in symptomatic patients with an initial non-diagnostic ECG should be performed at 15 - 30 min intervals or, if available, continuous computer-assisted 12-lead ECG recording.

MULTISAB platform currently supports only continuous signals, so dynamic changes in the ECG can only be monitored if a segment of ECG signal of at least 15 min exists. According to [5,6], ECG manifestation of AMI includes dynamical ST segment. For quantification of the ST segment dynamics, we can use some measure of dispersion, such as standard deviation, which was recommended by cardiologists.

Dynamical ECG features: standard deviation of ECG signal value at the J points (for all leads).

In the case of a continuous recording, we can also analyze heart rate variability (HRV) and QT Interval Variability (QTV). These time series could be of great importance, because, according to the guidelines [5], further ECG signs associated with AMI include: cardiac arrhythmias, intraventricular and atrioventricular (AV) conduction delays, and loss of precordial R wave amplitude. According to [7], comparison of a normality zone against the ischaemic episode in the same record has showed increases in power spectral density for low frequency (LF) and high frequency (HF) bands.

HRV features: LF, HF.

There are also indications that QTV can have clinical meaning [8]. In [9], normalized QTV (QTVnorm) and QT variability index (QTVI) were greater during ischemic episodes than during nonischemic episodes.

QTV features: QTVnorm, QTVI.

For AMI detection, metadata which would include age and sex and additional information, like, e.g. body mass index (BMI) would also be useful, but MULTISAB currently does not support import of anamnesis and metadata. We plan to add this feature in the future.

Additional expert features, which demand expanding the current frameworks with new operations, could include calculation of energy and entropy measures in wavelet domain [10] or additional information in the ST segments, such as shape, width, height, curvature, etc. In [11], a five-order polynomial fitting is applied to extract additional features from ST segments.

B. Atrial Fibrillation

According to the medical guidelines [12], and with up to date medical research [13], AF is associated with the following changes in ECG:

1. Lack of discrete P waves.
2. Fibrillatory or f waves are present at a rate that is generally between 350 and 600 beats/minute.
3. Ventricular response follows no repetitive pattern; the variability in the intervals between QRS complexes is often termed "irregularly irregular."
4. The ventricular rate, especially in the absence of AV nodal blocking agents or intrinsic conduction disease, usually ranges between 90 and 170 beats/min.
5. The QRS complexes are narrow, unless AV conduction through the His Purkinje system is abnormal due to functional (rate-related) aberration, pre-existing bundle branch or fascicular block, or ventricular preexcitation with conduction down the accessory pathway.

From ECG manifestation, we can define the following expert morphological ECG features:

- P wave absence - to detect a lack of discrete P waves
- Fibrillatory rate - must be between 350 and 600 beats/minute
- QRS complexes duration - to detect narrow QRS complexes

Morphological ECG features: P wave absence, fibrillatory rate, QRS complexes duration.

In the case of continuous recording, we can also analyze HRV, because one of AF manifestations is that ventricular response follows no repetitive pattern. Many researchers [14-16] have studied HRV features for AF detection and also gave their pathophysiological meaning. For expert features, we have chosen those with best discriminatory capability and those with the most important objective according to [14].

It should also be noted that the choice of HRV features depends on the length of the analyzed segment. The features are usually categorized into short-term HRV features, which can be calculated from segments duration of several minutes (normally 5 minutes), and long-term HRV features, which are calculated from 24 hours of recordings [17].

Short-term HRV features: Heart rate, standard deviation of all NN (normal beat to normal beat R-R) intervals (SDNN), the square root of the mean of the sum of the squares of differences between adjacent NN intervals (RMSSD), percentage of the number of pairs of adjacent NN intervals differing by more than 50 ms (pNN50), LF, HF, LF/HF ratio, approximate entropy (ApEn), sample entropy (SampEn).

Long-term HRV features: total power spectral density (TP), very low frequency band power spectral density (VLF), 1/f power-law exponent α.

Additional expert features, which demand expanding the current frameworks with new operations, could include calculation of wavelet entropy (WE) [18] or relative wavelet energy (RWE) [19], which are based on the wavelet transformation of the TQ interval. Both features can perform successfully even under heart rates with no variability. For example, in [19], AF was detected in less than 7 beats, with an accuracy higher than 90%.

C. *Congestive Heart Failure*

ECG manifestations of CHF are unspecific. In the guidelines for the diagnosis and treatment of acute and chronic heart failure [20], there is only a statement that an abnormal ECG increases the likelihood of the diagnosis of CHF.

The medical diagnosis of CHF is based on presenting signs and symptoms, patient's prior clinical history, physical examination and resting ECG. If all the elements are normal, CHF is highly unlikely. If at least one element is abnormal, plasma natriuretic peptides (NPs) should be measured, to identify those who need echocardiography, because echocardiography is the most useful and widely available test in patients with suspected CHF used to establish the diagnosis.

For CHF detection from ECG signals, expert morphological ECG features were selected based on ECG manifestations of various heart diseases described in [21] and [22].

Morphological ECG features: Q wave amplitude, PR interval duration, P peak amplitude, R peak amplitude, QRS complex duration, ratio of QT segment and heart rate corrected QT segment (QT/QTc), S wave duration in leads I, V_5, V_6, V_1 and V_2, R wave duration in leads V_5 and V_6.

For expert HRV features, we have chosen those with the best predictive values according to [23-26] and from recommendations by cardiologists.

Short-term HRV features: Heart rate (or 1/RRmean), short term detrended fluctuation analysis coefficient α1 (DFA α1), symbolic dynamics one variation pattern (1VP), LF, SDNN.

Long-term HRV features: VLF, 1/f power-law exponent α.

Although some features of heart rate turbulence (HRT), such as turbulence slope (TS), have significant predictive value [26], they are not used in the expert system because they cannot be calculated from all ECG recordings.

III. PREPROCESSING

Records preprocessing assumes various procedures for signal filtering, characteristic waveform detection, and data transformations as we can see in the Preprocessing ("Predobrada signala") UI, Fig. 4, where the user can select various methods for signal filtering, characteristic ECG waveform detection and data transformation.

Biomedical time series are often encumbered by noise that prohibits accurate calculation of clinically significant features. The goal of preprocessing is either to completely remove the noise or at least to ameliorate the quality of the recorded series. Signal filtering includes noise filtering (e.g. PLI notch filters for 50 / 60 Hz), baseline wandering corrections (e.g. for ECG because of breathing, body movements or electrodes impedance changes), and records segmentation into windows of certain width prior to applying preprocessing method [18]. Window width may be chosen explicitly by the user or it may be implicit, determined by the selected method preprocessing method.

The user can also select which characteristic waveforms are needed to be detected in an ECG and the corresponding detection methods, because locating waveforms is sometimes necessary for domain specific feature calculation. For example, if R peak annotations are not uploaded together with the ECG data signals, it is necessary to detect R peaks in ECG signals for HRV domain specific feature calculation (e.g. with Pan-Thompkins or Elgendi methods [27]).

Data transformations are performed after the detection of characteristic waveforms. The user can select various time (e.g. principal component analysis - PCA), frequency (e.g. fast Fourier transform, AR Burg method, Lomb-Scargle periodogram), time-frequency (e.g. wavelet transform, Hilbert Huang transform) and other types of data transformations.

The goal of data transformations is to obtain the data in a form from which it is easier to calculate precise domain or general features for describing specific subjects' states. For example, it is possible to transform ECG signal into wavelet domain and then apply some of the common signal extractors (see the list in [4]) in order to calculate features from the transformed signal.

IV. FEATURE EXTRACTION

Feature extraction is the central step in the analysis of biomedical time-series. In the Feature ("Značajke") tab of Feature extraction ("Izlučivanje značajki") UI, Fig. 5, the user will already be provided with a list of (expert) features selected (during construction of a completely new scenario) or predefined (for predefined analysis scenarios) in the scenario selection phase.

Fig. 5 shows how Feature extraction UI looks like for AF detection analysis scenario. Regardless of the scenario, the user can select additional features ("Dodatne značajke") in the feature extraction phase. User can also specify parameters for calculation of a particular feature. For parameters specification, the user must click on the feature to open a dialog box where it is possible to specify parameters for those features that are parametric. Before starting feature extraction, the user must specify signal segmentation parameters ("Prozor primjene").

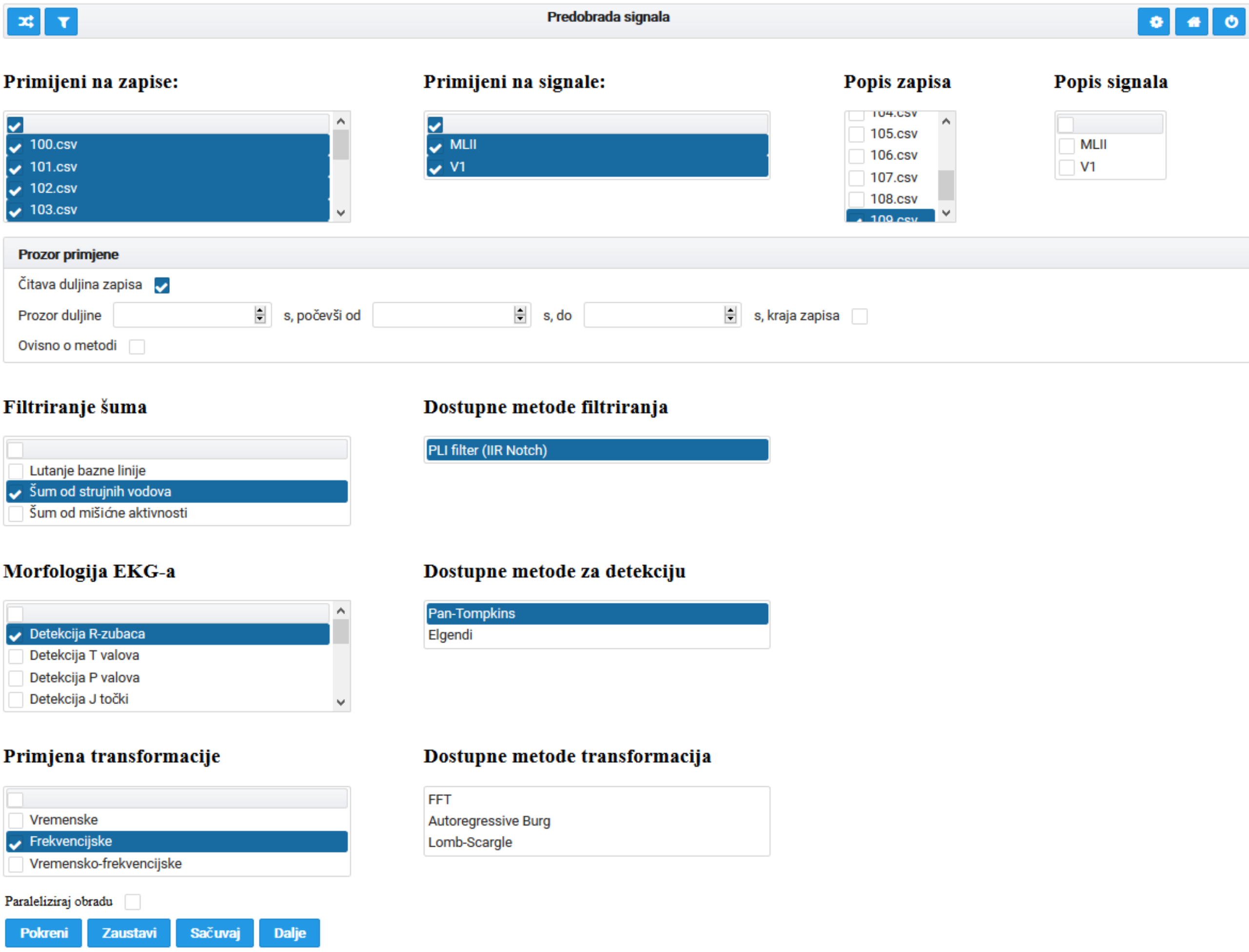

Figure 4. Preprocessing user interface; the user interface is currently implemented in Croatian language

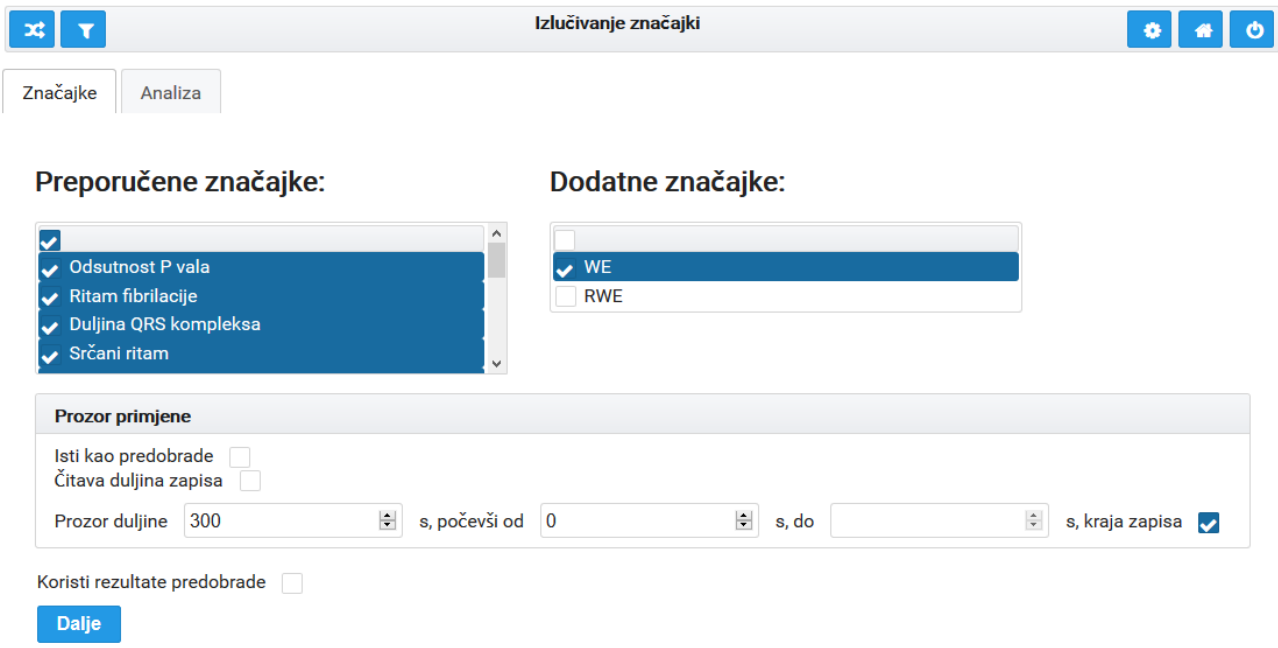

Figure 5. Feature extraction user interface; the user interface is currently implemented in Croatian language

In the Analysis ("Analiza") tab, the user can specify the type of calculation parallelization that will be performed. After feature extraction, one can save the extracted feature vectors in a file for future analysis, if one wants to have the intermediate results recorded.

V. Conclusion

In the paper, we have described an expert feature recommendation system for ECG analysis through predefined analysis scenarios. Initial version of the MULTISAB expert system is implemented in cooperation with cardiologists and includes the detection of AMI, AF, and CHF from ECG signals.

We have also reported the progress on the MULTISAB platform implementation. Currently, we have achieved the functionality of the first six analysis process steps, ending with the possibility of calculated features download. However, it should be noted that the current implementation is still in the test phase, which we will strive to complete in a few months.

Acknowledgments

We would like to thank Goran Krstačić, MD, PhD, FESC (from the Institute for Cardiovascular Diseases and Rehabilitation) for his valuable assistance with the expert system design.

References

[1] J. Oster, J. Behar, R. Colloca, Q. Li, Q. Li, and G. D. Clifford, “Open source Java-based ECG analysis software and Android app for atrial fibrillation screening,” in: Computing in Cardiology Conference (CinC) 2013, IEEE, pp. 731–734, 2013.

[2] A. Szczepanski and K. Saeed, “A mobile device system for early warning of ECG anomalies,” Sensors, vol. 14, no. 6, pp. 11031–11044, 2014.

[3] A. Jovic, D. Kukolja, K. Jozic, and M. Cifrek, “Use case diagram based scenarios design for a biomedical time-series analysis web platform,” in: Proceedings of the MIPRO 2016 Conference, P. Biljanovic, Ed., Rijeka: MIPRO Croatian Society, pp. 326–331, 2016.

[4] K. Friganovic, A. Jovic, K. Jozic, D. Kukolja and M. Cifrek, “MULTISAB project: a web platform based on specialized frameworks for heterogeneous biomedical time series analysis - an architectural overview,” in: Proceedings of the CMBEBiH 2017 Conference, A. Badnjevic (Ed.), IFMBE Proceedings vol. 62, Springer Nature Singapore, pp. 9–15, 2017.

[5] K. Thygesen, J. S. Alpert, A. S. Jaffe, et al., “Third universal definition of myocardial infarction,” European Heart Journal, vol. 33, pp. 2551–2567, 2012.

[6] E. Burns, “Myocardial ischaemia,” Life in the fastlane, http://lifeinthefastlane.com/ecg-library/myocardial-ischaemia/, accessed on: 2017-02-20.

[7] L. G. Gamero, J. Vila, and F. Palacios, “Wavelet transform analysis of heart rate variability during myocardial ischaemia,” Medical & Biological Engineering & Computing, vol. 40, no. 1, pp. 72–78, 2002.

[8] Y. Peng and Z. Sun, “Characterization of QT and RR interval series during acute myocardial ischemia by means of recurrence quantification analysis,” Medical & Biological Engineering & Computing, vol. 49, no. 1, pp. 25–31, 2011.

[9] T. Murabayashi, B. Fetics, D. Kass, et al., “Beat-to-Beat QT Interval Variability Associated With Acute Myocardial Ischemia,” Journal of Electrocardiology, vol. 35, no. 1, pp. 19–25, 2002.

[10] E. S. Jayachandran, K. P. Joseph, and R. U. Acharya, “Analysis of myocardial infarction using discrete wavelet transform,” Journal of Medical Systems, vol. 34, no. 6, pp. 985–992, 2010.

[11] L. Sun, Y. Lu, K. Yang, and S. Li, “ECG analysis using multiple instance learning for myocardial infarction detection,” IEEE Transactions on Biomedical Engineering, vol. 59, no. 12, pp. 3348–3356, 2012.

[12] P. Kirchhof, S. Benussi, D. Kotecha, et al., “2016 ESC Guidelines for the management of atrial fibrillation developed in collaboration with EACTS,” European Heart Journal, vol. 37, pp. 2893–2962, 2016.

[13] B. Olshansky, “The electrocardiogram in atrial fibrillation,” UpToDate, http://ultra-medica.net/Uptodate21.6/contents/mobipreview.htm?25/42/26278, accessed on: 2017-02-20.

[14] A. Bollmann, D. Husser, L. Mainardi, et al., “Analysis of surface electrocardiograms in atrial fibrillation: techniques, research, and clinical applications,” Europace, vol. 8, no. 11, pp. 911–926, 2006.

[15] D. G. Shin, C. S. Yoo, S. H. Yi, et al., “Prediction of paroxysmal atrial fibrillation using nonlinear analysis of the R-R interval dynamics before the spontaneous onset of atrial fibrillation,” Circulation Journal, vol. 70, no. 1, pp. 94–99, 2006.

[16] R. Alcaraz and J. J. Rieta, “A review on sample entropy applications for the non-invasive analysis of atrial fibrillation electrocardiograms,” Biomedical Signal Processing and Control, vol. 5, no. 1, pp 1–14, 2010.

[17] R. Sassi, S. Cerutti, F. Lombardi, et al., “Advances in heart rate variability signal analysis: joint position statement by the e-Cardiology ESC Working Group and the European Heart Rhythm Association co-endorsed by the Asia Pacific Heart Rhythm Society,” Europace, vol. 17, pp. 1341–1353, 2015.

[18] J. Ródenas, M. García, R. Alcaraz, and J. J. Rieta, “Wavelet entropy automatically detects episodes of atrial fibrillation from single-lead electrocardiograms,” Entropy, vol. 17, no. 9, pp. 6179–6199, 2015.

[19] M. García, J. Ródenas, R. Alcaraz, and J. J. Rieta, “Application of the relative wavelet energy to heart rate independent detection of atrial fibrillation,” Computer Methods and Programs in Biomedicine, vol. 131, pp. 157–168, 2016.

[20] P. Ponikowski, A. A. Voors, S. D. Anker, et al., “2016 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure,” European Heart Journal, vol. 37, pp. 2129–2200, 2016.

[21] J. Malmivuo and R. Plonsey, “Bioelectromagnetism: principles and applications of bioelectric and biomagnetic fields,” chapter 19: “The basis of ECG diagnosis,” Oxford University Press, New York, 1995.

[22] DailyCare BioMedical Inc., “Introductory guide to identifying ECG irregularities,” http://www.dcbiomed.com/proimages/materials/EKG_Introduction/portableEKG_weitere_Informationen.pdf, accessed on: 2017-02-20.

[23] S. Guzzetti, M. T. La Rovere, G. D. Pinna, et al., “Different spectral components of 24 h heart rate variability are related to different modes of death in chronic heart failure,” European Heart Journal, vol. 26, pp. 357–362, 2005.

[24] T. H. Makikallio, H. V. Huikuri, U. Hintze, et al., “Fractal analysis and time- and frequency-domain measures of heart rate variability as predictors of mortality in patients with heart failure,” American Journal of Cardiology, vol. 87, issue 2, pp. 178–182, 2001.

[25] R. Maestri, G. D. Pinna, A. Accardo, et al., “Nonlinear indices of heart rate variability in chronic heart failure patients: redundancy and comparative clinical value,” Journal of Cardiovascular Electrophysiology, vol. 18, no. 4, pp. 1–9, 2007.

[26] M. T. La Rovere, G. D. Pinna, R. Maestri, et al., “Autonomic markers and cardiovascular and arrhythmic events in heart failure patients: still a place in prognostication? Data from the GISSI-HF trial,” European Journal of Heart Failure, vol. 14, issue 12, pp. 1410–1419, 2012.

[27] M. Elgendi, “Fast QRS detection with an optimized knowledge-based method: evaluation on 11 standard ECG databases,” PLOS ONE, vol. 8, no. 9: e73557, 2013.