MULTISAB project: a web platform based on specialized frameworks for heterogeneous biomedical time series analysis - an architectural overview

Kresimir Friganovic*, Alan Jovic*[c], Kresimir Jozic**, Davor Kukolja* and Mario Cifrek*

*University of Zagreb Faculty of Electrical Engineering and Computing, Unska 3, 10000 Zagreb, Croatia
**INA – industrija nafte, d.d., Avenija Veceslava Holjevca 10, p.p. 555, HR – 10002 Zagreb, Croatia
kresimir.friganovic@fer.hr, [c](corr. author): alan.jovic@fer.hr, kjozic@gmail.com, davor.kukolja@fer.hr, mario.cifrek@fer.hr

Abstract. The aim of this work is to present an architectural overview of a novel web platform used for heterogeneous biomedical time series analysis. Its architecture is based on three subprojects: frontend, backend, and processing. Frontend uses several contemporary web technologies to present a fast, responsive and pleasing user interface. Backend, written in Java, communicates with a database and with other servers, on which the processing subproject is deployed. The processing subproject contains several frameworks intended for: record input handling, signal preprocessing, data visualization, general time series features extraction, specific time series features extraction (e.g. heart rate variability and electroencephalogram), data mining, and reporting. The platform is in an early phase of implementation, but we demonstrate its features and capabilities, of which feature extraction frameworks and signal visualization currently stand out.

Keywords: framework, web platform, biomedical time series, frontend development, signal visualization

1   Introduction

The need for biomedical software is showing growth in healthcare community and in general population users. Despite a great number of web and mobile applications targeting general users, such as sports and fitness enthusiasts, specialized software developed for medical professionals is limited to continuous monitoring of biomedical time series (BTS). To overcome this problem, we pursue the development of an efficient and upgradeable BTS analysis system in a form of an integrative web platform for automatic classification of human body disorders. We call it "MULTISAB project" (short for: Multiple Time Series Analysis in Biomedicine). The system will

serve as a help for medical specialists in diagnostics and early detection of various diseases [1].

In our previous work, we described the early efforts in designing such a system, which included an overview of used technologies and use case based scenarios specification [1,2]. The platform predicts a total of 8 fundamental use case scenarios: analysis type selection, scenario selection, input data selection, records inspection, records preprocessing, feature extraction, model construction, and reporting [2]. Additional scenarios for handling the user account and platform administration were also provided. The approach that we followed allows adaptive control of user needs, and can be specialized for different kinds of analysis goal (e.g. classification, visualization, prediction). As the system is under development as a web platform, the users will be able to access it from afar and transmit biomedical data to the server, where efficient analysis can be performed. Also, a web platform enables easier maintenance and larger user base. This distinguishes our platform from local or hospital-specific software solutions [2].

The aim of the current paper is to provide an architectural overview and detailed description of the MULTISAB project's frameworks, and to show the currently implemented visualization aspects of the platform.

The architectural overview of the platform is shown in Fig. 1. MULTISAB project is divided into three subprojects: *frontend*, *backend* and *processing*. The *frontend* and *backend* subprojects support communication between end users and servers. Basic functions are currently supported: web browser login into the system, uploading and visualization of the heterogeneous BTS, specification of the steps included in the analysis and defining the outcomes of the analysis. The *processing* subproject contains several frameworks for heterogeneous BTS data analysis and will
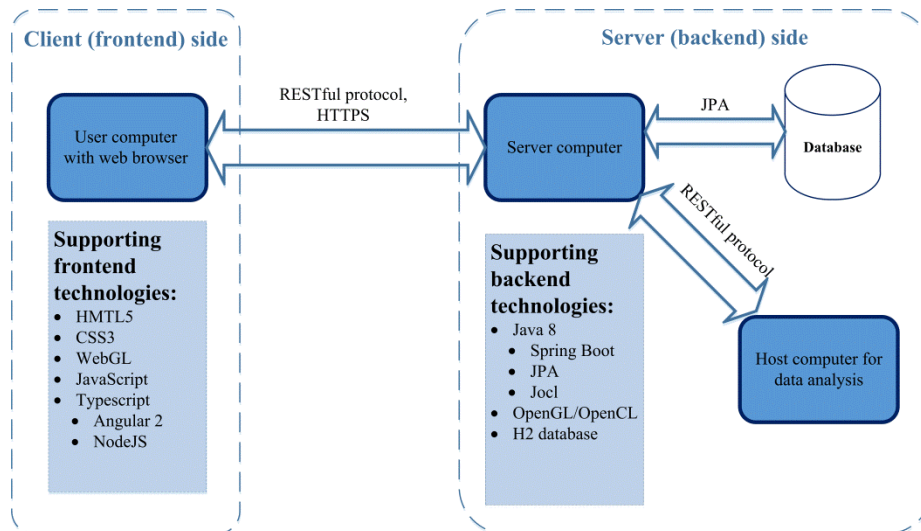
Fig. 1. Architectural overview of the MULTISAB project web platform

be discussed in more details in Section 2. In ideal conditions, every subproject runs on different computer host(s), therefore, they do not share physical computer resources. TypeScript, which is used on *frontend*, is a strict superset of JavaScript that compiles to the selected version of JavaScript. Through his web browser, a user (client) can send requests to the server implemented on *backend*.

The *backend* subproject is written in Java 8 and serves for processing requests from users and for communication with a database and data analysis hosts. Java 8 is used, because it enables efficient parallel processing with data streams. A JPA framework (EclipseLink) is used to establish communication between the Java-written *backend* and a relational database by using object-relational mapping. It also implements a persistent control of data. In order to run all the necessary resources on the backend host, Spring Boot framework is used, which facilitates web application implementation. In the future, 3D visualization of body disorders using WebGL on the client-hand side will be supported. In order to communicate with the graphical processing units in Java, OpenGL/OpenCL are used with jocl Java framework.

Communication between *backend* and *frontend* uses HTTPS protocol to insure secure transfer and communication. For communication between Java 8 on *backend* and JavaScript on *frontend*, RESTful protocol is used.

2   Processing frameworks

The *processing* subproject is written entirely in Java 8 and implements everything regarding data processing and analysis. The subproject is divided into several frameworks in order to improve maintainability and scalability of the integrated system. The frameworks consist of one or more packages specialized for different tasks in signal analysis, which correspond to the mentioned use case scenarios.

Detailed description of some of the implemented frameworks is given in the following subsections. Frameworks for ECG analysis and data mining algorithms currently under construction are discussed in Section 5.

2.1   Common signal features framework

The framework for common signal features extraction includes a great number of features for time series analysis in time (both linear and nonlinear), frequency and time-frequency domain. The EEGframe and HRVframe [3,4] frameworks have been refactored to extrude general BTS features for MULTISAB processing into a common signal features framework. The idea behind such a great number of common features is a more accurate and more all-encompassing analysis than it is usually pursued for BTS (comparable to the work of Fulcher et al. [5]). The common features can be used alone or together with specialized features to train data mining classifiers in order to efficiently classify diseases.

Power spectral density (PSD) of a signal, which is used for frequency domain analysis, is implemented using three different algorithms: fast Fourier transform (FFT), Burg method and Lomb-Scargle method. Three window types are supported

for FFT-based PSD estimate: rectangular, Hann, and Hamming window. Other included transformations are in the time-frequency domain: wavelet (of Haar type) and Hilbert-Huang transformation.

Time domain analysis package includes the following statistical features: mean value of a segment, standard deviation value of a segment, mean of first differences, mean of first differences normalized by standard deviation, mean of second differences, mean of second differences normalized by standard deviation, autocorrelation coefficient, and Fano factor [6,7]. The package also contains Principal Component Analysis (PCA), which is also used for dimensionality reduction in data mining.

A great amount of literature was investigated in order to establish a list of relevant nonlinear features in analysis of BTS in general. Namely, alphabet entropy [8], Renyi entropy [9], sample entropy [10], Hurst exponent [6], synchronization likelihood [11] and others can all be used in BTS analysis [12]. Details on these methods can be found in the abovementioned literature. Under the nonlinear features package, several sub packages are included for easier management: entropy, fractal, multivariate and phase space features' packages. Features included in common signal features framework are listed in Fig. 2.

2.2   Data handling framework

Data handling framework is used to import various types of data for biomedical time series. The most common formats are EDF/EDF+, which are used to store ECG, EEG and other time dependent biomedical recordings. The formats supported so far are:
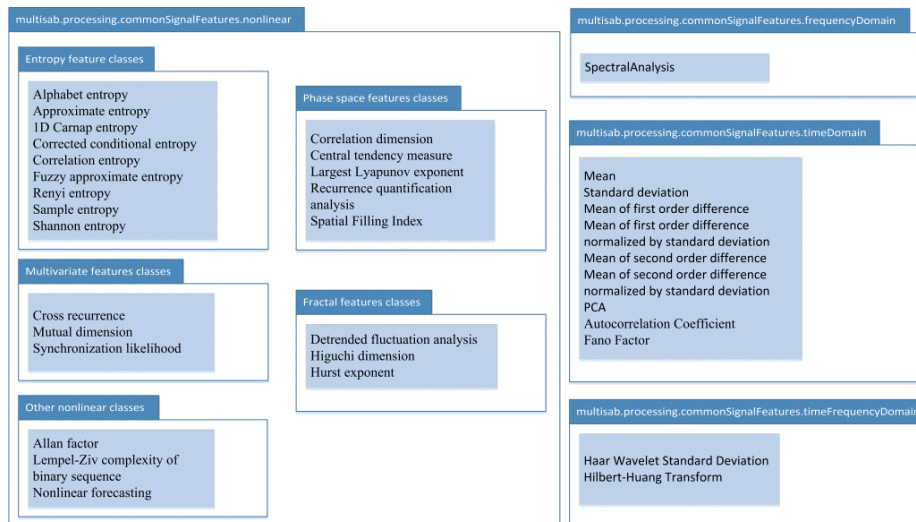- .ann
- .csv

Fig. 2. Common signal features

- .edf/.edf+c
- .txt

The textual (.txt), annotation (.ann), and comma-separated value (.csv) formats are supported based on PhysioNet [13] representation of the recordings. Various auxiliary components of data: annotations (time and sample of the annotation, duration of annotation, type), data record (start time offset, signal data), metadata (version of data format, local patient identification, local record identification, etc.) and data parameters (label, physical minimum and maximum, digital minimum and maximum, prefiltering information) are included in various classes related to record input processing.

2.3 Specialized biomedical time series analysis frameworks

For feature extraction from specific BTS (ECG, EEG, EMG, …), specialized frameworks can be constructed. Currently, EEG analysis and HRV analysis frameworks are implemented. The EEG analysis framework contains specific spectral features' extractors used in the analysis of an EEG signal. PSD for the five most common brain rhythms are estimated: alpha (7.5 – 12.5 Hz), beta (13 – 25 Hz), gamma (25 – 100 Hz), delta (1 – 4 Hz) and theta (4 – 7 Hz) rhythm [14]. Spectral entropy [15] is calculated using any of the previously obtained PSD estimates.

Similarly, for HRV spectral analysis, ultralow (0.00001 – 0.003 Hz), low (0.003 – 0.04 Hz), high (0.15 – 0.4 Hz), and low / high ratio frequency features are calculated from the estimated PSD.

Time domain features included in the HRV analysis are the majority of the standard statistical ones used in literature [12]: the mean NN interval length, SDNN, RMSSD, SDSD, NNX, pNNX, SDANN, SDNNindex, HRV triangular index, and TINN (we omit the features' descriptions due to lack of space).

Also, HRV standard deviation ratios (SD1, SD2, SD1/SD2, CSI, CVI) can be calculated in order to assess cardiac sympathetic and vagal indices [16].

3 Database

As it is important to understand the underlying data organization in order to make better use of the platform, database implementation will be shortly described here. Our project uses H2 database engine, which is a fast, server-based and in-memory database engine. The database's entity-relationship model is shown in Fig. 3. The structure of the project's database is focused on user session data (SessionData). SessionData is connected to general user data (UserData) and data analysis scenario (Scenario). The main idea of the database design is to support storing the analysis steps of a currently ongoing session and to enable continuation of previous session, if a user decides to stop his work and then continue later on. MULTISAB database is completely built in the normalized form (3NF). As the MULTISAB project is in development, not all database connections are currently enabled. Complete database will be available in the later stages of the project.
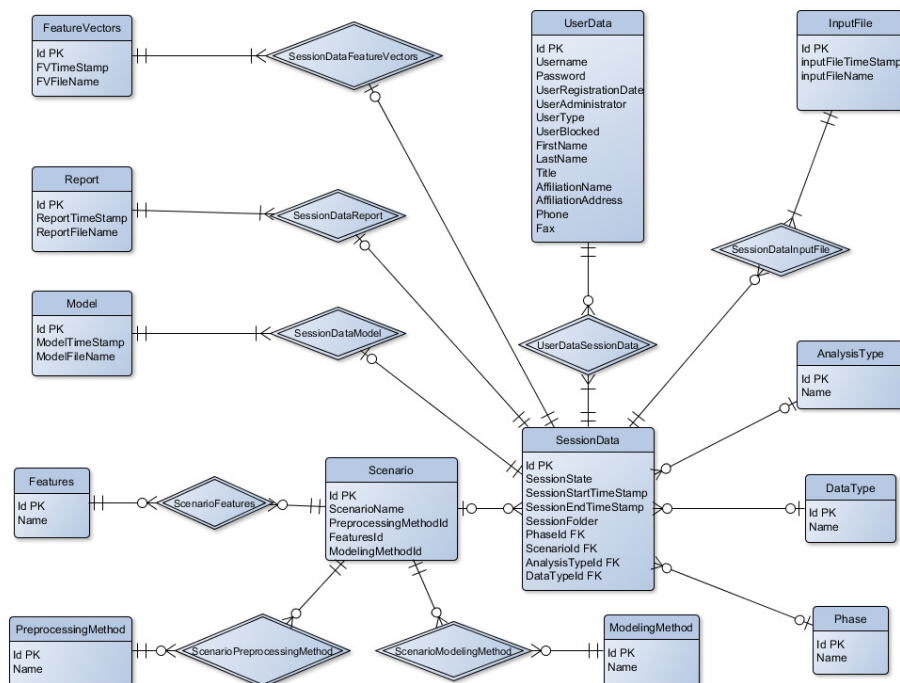
**Fig. 3.** MULTISAB relation database model

4 Signal visualization

Signal records are uploaded through the web browser's user interface. In Fig. 4, the user interface (UI) layout with an ECG image, plotted with a red grid (0.5 V and 0.5 s), is shown. Currently, user interface is in Croatian and the English version will be available at the end of the MULTISAB project. On the upper side of the UI, there are option boxes designed to specify parameters for generation of signal image. Parameters allow users to specify: data source ("Odaberi zapis"), width of signal segment ("Širina segmenta"), starting position in the signal ("Pozicija u signalu"), signals to show if there are multiple signals in a record ("Odaberi signale") and x- ("širina") or y-dimension ("visina") plot limit, specified with a width or height of the signal ("Fiksna dimenzija"). When the user clicks on the get image button ("Dohvati sliku"), the parameters are sent to the *backend* and processed under the *processing* subproject within the *SignalVisualization* framework. 2D image created in the *ImageCreating* class of the framework is sent back to the user via *backend* subproject and plotted in UI on the *frontend* (Fig. 5).

In the box: "Ukupna duljina signala," full signal length (in seconds) is shown. User can download a specific plotted image with a button ("Spremi sliku"). Scrolling through a signal is enabled with arrow buttons in the top right corner. An example of
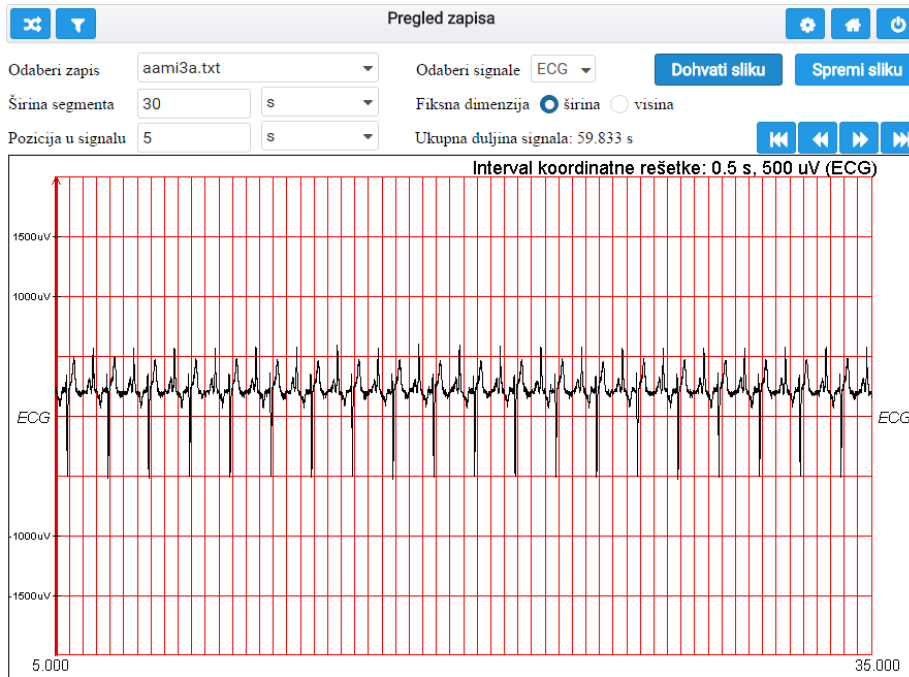
Fig. 4. Signal visualization in web browser user interface

multiple signal visualization (heterogeneous BTS) is shown in Fig. 6. Here, one can observe that the grid is scaled so that all the signals can be fitted into the same image.
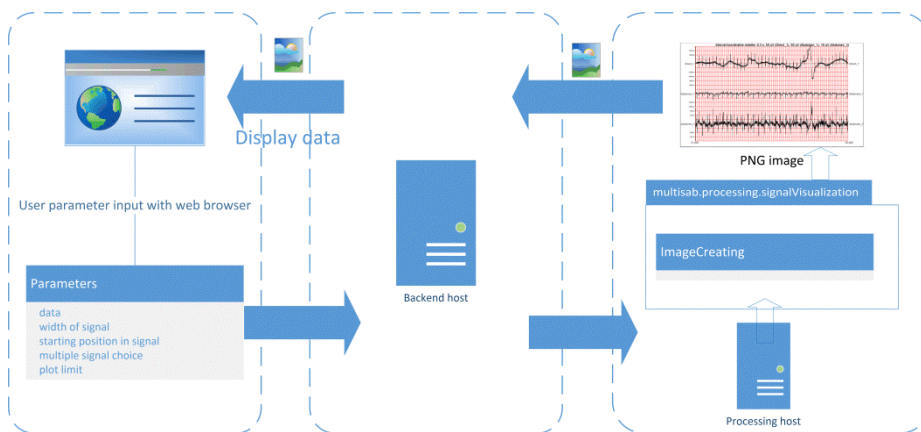
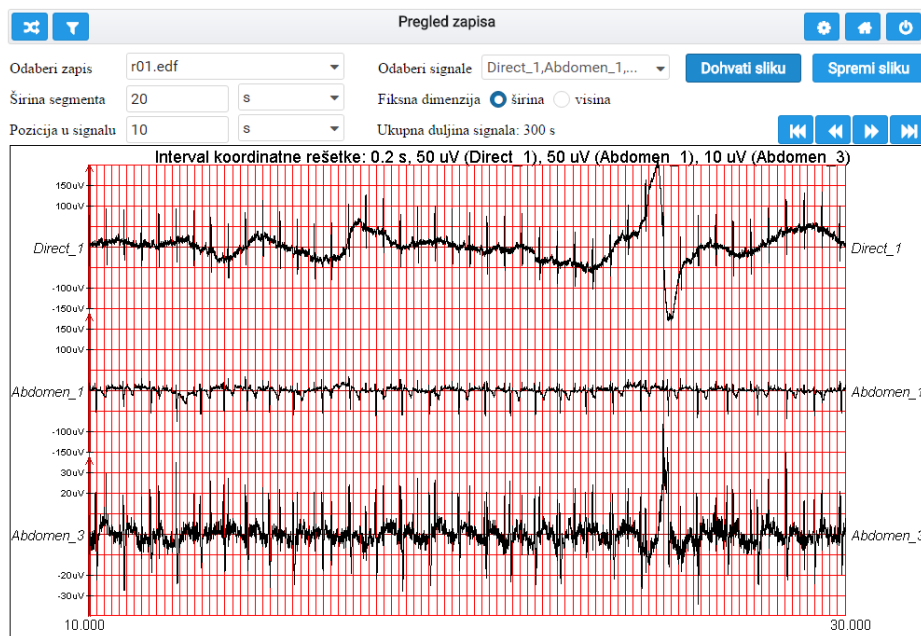**Fig. 5.** Sending image parameters and plotting image in a web browser

Fig. 6. Multiple signal visualization

5   Discussion and conclusion

As an ongoing effort, we are currently working on an implementation of a specialized framework for ECG analysis, which would, in collaboration with HRV analysis framework, be able to detect such important events as cardiac ischemia and atrial fibrillation with high accuracy. Characteristic waves (QRS complex, P and T waves [17]) detection is going to be implemented to help distinguish various indications of pathological states in ECG signal, such as ST depression or abnormal changes in T waves. Consensus documents for pathological states will be consulted (e.g. acute myocardial infarction diagnostic criteria [18]).

Data mining methods are currently being implemented in a separate framework, which also includes some frequently used open source tools (currently LibSVM [19] and Encog [20]). The machine learning algorithms implemented so far are: 1) support vector machine (SVM) from LibSVM, 2) various types of artificial neural networks (ANN) included in Encog: multilayer perceptron ANN, radial basis function ANN, probabilistic neural network (PNN) and NEAT ANN.

Furthermore, data mining framework is planned to include some ensemble algorithms (e.g random forest and AdaBoost), tree-based (e.g. C4.5) and rule-based algorithm (e.g. RIPPER), as these methods are fast and accurate, while some of them retain comprehensiveness of the obtained model. In order to speed up the model construction procedure and possibly improve accuracy, the framework will also include

typical filter (e.g. symmetrical uncertainty) and wrapper (e.g. Naïve Bayes) feature selection methods used to remove irrelevant and redundant features [21].

In this work, we have shown the current state of implementation of an advanced integrated web platform for heterogeneous time series analysis. Much work still needs to be done in order to accomplish the following goals: 3D visualization of body disorders, calculation parallelization on multiple physical resources, and analysis reporting. Furthermore, an expert system is planned, which will recommend a reduced number of features that are needed to be extracted in certain scenarios where medical guidelines recommend using some features, while disregarding others.

Conflict of Interest

The authors declare that they have no conflict of interest.

Acknowledgements

References

1. Jovic, A., Kukolja, D., Jozic, K., Horvat, M.: A Web Platform for Analysis of Multivariate Heterogeneous Biomedical Time-Series – a Preliminary Report. In: The 23rd Int. Conf. on Systems, Signals and Image Processing (IWSSIP 2016), p. 70. IEEE Press, Bratislava. (2016)
2. Jovic, A., Kukolja, D., Jozic, K., Cifrek, M.: Use Case Diagram Based Scenarios Design for a Biomedical Time-Series Analysis Web Platform. In: Proc. of the 39th Int. Convention MIPRO 2016, pp. 326–331. MIPRO, Opatija (2016)
3. Jovic, A., Bogunovic, N., Cupic, M.: Extension and Detailed Overview of the HRVFrame Framework for Heart Rate Variability Analysis. In: Proc. of the Eurocon 2013 Conf., pp. 1757–1763. IEEE Press, Zagreb. (2013)
4. Jovic, A., Suc, L., Bogunovic, N.: Feature extraction from electroencephalographic records using EEGFrame framework. In: Proc. of the 36th Int. Convention MIPRO 2013, pp. 1237–1242. MIPRO, Opatija (2013)
5. Fuchler, B.D., Little, A.M., Jones, N.S.: Highly comparative time-series analysis: the empirical structure of time series and their methods. J. R. Soc. Interface 2013 (10), p. 20130048. (2013)
6. Teich, M.C., Lowen S.B., Jost, B.M., Vibe-Rheymer, K.: Heart Rate Variability: Measures and Models. In: Akay, M. (ed.): Nonlinear Biomedical Signal Processing. IEEE Press, New York, pp. 159–213. (2000)
7. Wang, X.-W., Nie, D., Lu, B.-L.: EEG-Based Emotion Recognition Using Frequency Domain Features and Support Vector Machines. LNCS 7062, pp. 734–743. (2011)
8. Jovic, A., Jovic F.: Classification of cardiac arrhythmias based on alphabet entropy of heart rate variability time series. Biomed. Signal Process. and Control, in press, 31, pp. 217–230. (2017) doi:10.1016/j.bspc.2016.08.010

9. Waheed, K., Salam, F.M.: A Data-Derived Quadratic Independence Measure for Adaptive Blind Source Recovery in Practical Applications. In: Proc. 45th IEEE Int. Midwest Symposium on Circuits and Systems, pp. 473–476. IEEE Press. (2002)
10. Richman, J.S., Moorman, J.R.: Physiological time-series analysis using approximate entropy and sample entropy. Am. J. Physiol. (Heart Circ. Physiol.) 278 (6), pp. 2039–2049. (2000)
11. Stam, C. J., van Dijk, B. W.: Synchronization likelihood: an unbiased measure of generalized synchronization in multivariate data sets. Physica D: Nonlinear Phenomena 163 (3-4), pp. 236–251. (2002)
12. Sassi, R, Cerutti, S., Lombardi, F., Malik, M., Huikuri, H.V., Peng, C.-K., Schmidt, G., Yamamoto, Y.: 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 17, pp. 1341–1353. (2015)
13. Goldberger, A.L., Amaral, L.A.N., Glass, L., Hausdorff, J.M., Ivanov, P.C., Mark, R.G., Stanley, H.E.: PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. Circulation 101 (23), e215-e220. (2000)
14. Gerrard, P., Malcolm, R.: Mechanisms of modafinil: A review of current research. Neuropsychiatr. Dis. Treat. 3 (3), pp. 349–364. (2007)
15. Ermes, M., Pärkkä, J., Cluitmans, L.: Advancing from Offline to Online Activity Recognition with Wearable Sensors. In: Proc. of the 30th Ann. Int. Conf. of the IEEE Engineering in Medicine and Biology Society, pp. 4451–4454. IEEE EMBS. (2008)
16. Lin, C.-W., Wang, J.-S., Chung, P.-C.: Mining Physiological Conditions from Heart Rate Variability Analysis. IEEE Comput. Intell. M. 5 (1), pp. 50–58. (2010)
17. Sun, Y., Chan, K.L., Krishnan, S.M.: Characteristic wave detection in ECG signal using morphological transform. BMC Cardiovasc. Disord. pp. 5–28. (2005)
18. Thygesen, K., Alpert, J.S., Jaffe, A.S., Simoons, M.L., Chaitman, B.R., White, H.D.: Third universal definition of myocardial infarction. Eur. Heart J. 33 (20), pp. 2551–2567. (2012)
19. Chang, C.-C., Lin, C.-J.: LIBSVM : a library for support vector machines. ACM TIST 2(27), pp. 1–27. (2011)
20. Heaton, J.: Encog: Library of Interchangeable Machine Learning Models for Java and C#. JMLR 16, pp. 1243–1247. (2015)
21. Witten, I.H., Frank, E., Hall, M.A.: Data Mining: Practical Machine Learning Tools and Techniques, 3rd ed. Morgan Kaufmann, Burlington, CA (2011)