ONLINE PLATFORM FOR PROCESSING AND STORAGE OF INFORMATION IN THE FIELD OF MEDICINE: IMPROVING EDUCATION OF THE MEDICAL STUDENTS

G. Georgieva-Tsaneva¹, M. Gospodinov¹, G. Bogdanova²

¹Institute of Robotics, Bulgarian Academy of Sciences (BULGARIA)
²Institute of Mathematics and Informatics, Bulgarian Academy of Sciences (BULGARIA)

Abstract

Widely used advanced technologies today provide tools for creating modern information platforms that support software for processing and analyzing data in different areas of human life. The creation of web-based information platforms makes it possible to exploit the advantages of internet storage and maintenance of large volumes of information and access to this information in real time by different users. The information platforms can be successfully applied for educational purposes - they are an effective opportunity to create basic and additional knowledge and skills for the learners.

The report presents an information platform model for the collection, processing, analysis and storage of medical data. The software platform supports means to input incoming information for individuals undergoing medical research; physiological information; processing software and means of storing the results obtained. The software system performs mathematical analysis on input physiological data, calculates important parameters for the health of individuals and produces useful graphical results.

The information system model proposed in the report was created to store and analyze information cardiac records of patients obtained through Holter monitoring. Physiological data is large, take up much memory and therefore an effective solution for their storage is proposed through the use of cloud technologies rather than in the local servers of medical institutions. The information content of cardiac data is well structured and indexed to provide quick access to data.

The information platform is created to solve the tasks of a scientific project under the Scientific Research Fund of the Republic of Bulgaria, in which is provided conducting a study of patients with various cardiac diseases. The information platform makes it possible to store records of an individual made over a period of time to monitor the development of the disease and the effect of the treatment.

The platform information - input and results obtained through mathematical analysis can be used for the purpose of training of medical specialists at the universities of the Republic of Bulgaria. The information system can support cardiology training, providing real information on the development of cardiac disease and the effect of treatment. The platform may be developed and as a mobile application that students will be able to use on a daily basis to acquire new knowledge, improve their skills in the particular medical specialty in which they are trained, and improve their learning outcomes.

Keywords: Educational web based resources, Medical students, Information web-based platform, Cloud technologies, Research project, Mathematical analysis, Holter monitoring.

1 INTRODUCTION

The processing and analysis of Big Data [3] (large volumes of data) is a key factor for innovation and competitiveness. Due to population growth and aging in developed countries and the availability of a variety of data formats, the integration of these growing medical data over time is a great challenge. There is an acute need to create web-based information platforms for the collection, processing, analysis, and storage of medical data. Medical data is large in volume, occupy a lot of memory and therefore the local servers of the institutions are not the most appropriate solution for their storage. Better and more efficient solutions are provided through the use of cloud technologies [4].

The design and development of a web-based information platform for the collection, processing, analysis, and storage of medical data can be divided into several stages. When constructing the platform, it is good to explore and incorporate the next important features and functionalities.

Initially have explored previous research, technology, and solutions for web-based cloudy and interactive medical platforms. Based on successful past decisions, to create a life cycle scheme for the
project and select and develop the platform model. When have to choose data platform architecture, have to look for architecture that allows for quick data loading. A convenient choice is distributed computer architecture based on a database that processes and stores large volumes of information (e.g. HBase’s Hadoop). To generating a semantic data model, a suitable solution is an ontological model combined with a NoSql archive. Combining the semantic model with NoSQL technology make it easier to build a semantic network running large volumes of data (e.g., large RDF data).

To make it easier to analyze, search and filter large data, cloud computing [4] is best used. Cloud solutions allow the rapid integration of large volumes of medical data into repositories and the extraction of synthesized information. The construction of a telemedicine network ensures an overall efficiency of the medical system. The telemedicine solution, in comparison with traditional medical check-ups, ensures a better diagnosis, correct diagnosis, improved access and equality in the spread of health services. Important final steps in the life cycle of the system are the stages of implementation, digital marketing and the market realization of the established medical platform with the help of the use of modern technologies.

2 OVERVIEW OF RESEARCH AND SOLUTIONS FOR WEB-BASED MEDICAL PLATFORMS IN BULGARIA AND AROUND THE WORLD

Today the focus of software medical applications is telemedicine, management of treatment, respiratory and cardiac care, women’s health, care for diabetic patients, drug delivery, optimize keeping the review time, child care, mental health, delivery of medical marijuana.

An overview of previous research and some well-known web-based cloudy and interactive platforms in the field of medicine abroad is made. Taiwanese researchers have presented [1] a web-based cloud platform (Hadoop distributed computing platform) for processing Big Data which ensuring fast storage of a large number of medical records, data processing functions such as effective search, filtering and analyzing stored data. The created database uses distributed cloud architecture and is based on HBase’s Hadoop database (which loading and storing large volumes of information). The distributed computer architecture Hadoop MapReduce is used for data processing and analysis.

Software systems and solutions using NoSQL databases to store massive RDF data are described in [5-10]. The resource description framework for large RDF data called triplestore based on NoSQL data management systems such as HBase, Cassandra, Accumulo, and Couchbase is very perspective and allows for the storage and efficient management of a very large amount of web data compared to classic RDF databases. NoSQL technology can cope with the data explosion called Big Data. It provides advantages such as scalability and high availability compared to traditional triplestores.

Mobile monitoring of chronic diseases is particularly relevant. The mHealth platform (mobile health) is a common term for the use of mobile phones and other wireless technologies in medical care. The most common application of mHealth is the use of mobile devices for consumer education for preventive health services. The platform is also used for disease surveillance, support for treatment, epidemic epidemiology tracking and chronic disease management. The telehealth platform connects to a videoconference between a doctor or a nurse and a patient, including remote patient monitoring via wireless medical devices. HealthTap is an American digital platform where you can ask in a network of 100,000 doctors and get health advice. It is a digital system for liaison between physicians and patients, in which, besides answering specific questions and specific symptoms, adequate health advice can also be obtained for proper behavior in different conditions.

Cloud solutions for collections with patient data in health institutions are presented in [11, 13-16]. In [12] Koufi et al. show a medical portal for ubiquitous access to personal cloud health records.

There are several web-based interactive health platforms in Bulgaria that have been successfully used. One of them is the network of telemonitoring centers "Check Point Cardio" - a telemedicine system for long-term monitoring of patients with cardiac problems. The software of the telemedicine system was developed by a Bulgarian company [2], the system uses Polish Holter models. Models for evaluation of telemedicine systems are presented in [17].

A web-based platform for the Clinicubes CTMS Clinical Investigation Management System and a web-based information system HARP has been developed. Clinicubes (https://www.clinicubes.com/) offers budgeting features, patient reviews, and physician activities. Clinicubes CTMS offers a data management system and complex processes related to medical research. The Bulgarian Aptuso Healthcare Software Ecosystem (http://aptuso.com/) is an integrable and intuitive web-based
healthcare platform based on the Big Data structure and machine learning algorithms that ensures integrity and data security. The system offers comprehensive healthcare solutions including a hospital information system, an electronic medical record, laboratory software, electronic prescriptions and therapies, electronic directions, home care, and telemedicine applications, health management. Sqilline (https://sqilline.com/support) has developed the Danny Platform analytical platform in the field of oncology and cardiology. The platform retrieves information from hospital information systems, registers, etc., and generates aggregated real-time analyzes. Through built-in machine learning algorithms, sophisticated tools for comprehensive searches and analysis of heterogeneous patient data, the system generates possible diagnoses, therapies and specific clinical metrics.

3 MODEL OF INFORMATION PLATFORM FOR STORAGE, PROCESSING, AND ANALYSIS OF CARDIOLOGICAL DATA

The standard analysis of ECG records is routine, informative, calculates only a few parameters, provides no predictive prognosis for disease progression, is based on the detection of some signs of cardiovascular disease and examines static research records. HRV-based studies (non-invasive, based on time intervals between heart beats) allow the analysis of cardiac dynamics and prognosis of disease progression in the near and distant future. HRV investigation can be performed in the time domain, frequency domain, time-frequency domain, and a total of over 20 parameters in these domains are surveyed. The HRV standard established in 1996 [23] provides guidance for carrying out these studies and identifies the normal values (characteristic of a healthy individual) of a large number of parameters. By conducting a non-invasive study of patients through HRV, we can determine based on the parameters determined the individual's state of health.

3.1 Model of information system for storage, treatment, and analysis of cardiological data

The information system provides opportunities for long-term storage of patients’ cardiac recordings using cloud technologies. The saved records can be used in different locations, which allows consultation with cardiologists from different localities, including from abroad. The information platform comprises means for storing incoming information obtained by conducting a survey of patients (ECG data, Holter data, data obtained by means of photoplethysmographic methods, etc.), a software system for processing and analysis of the input data and a recording system and storage of some of the more important results obtained.

The goals set with the creation of the information cardiology platform are the creation of an efficient software system for processing and analysis of prolonged cardiological data with new, modern mathematical methods; organizing the storage of input Holter data in an information database and storing the results obtained in the electronic history of the disease of each patient. The information platform can be used in the training process of the future cardiologists, in order to get acquainted with the occurring parametric changes in the HRV in the various cardiac diseases.

The model for an information system for the processing, analysis, and storage of cardiac data is presented in Fig. 1.

ECG data can be obtained by an electrocardiograph (5-20 minute stationary method) or Holter (long-term patient monitoring device - from 24 to 72 hours, with modern Holters up to 2 weeks).

3.2 Description of the system

The software system performs pre-processing of input cardiac data, including reduction of input data interference, detection of cardiac complexes, the formation of time sequence from real cardiac complexes (excluding amplitude deviations due to side factors), the formation of the time sequence from normal cardiac intervals. On the last sequence, analyzes are performed using various mathematical methods: linear and non-linear methods, fractal analysis methods and methods of wavelet theory. The established software system aims to integrate all new mathematical methods, the application of which is effective scientific research and clinical application for the study of cardiac diseases through the non-invasive HRV method. Each software procedure is optimally designed to achieve effective results while minimizing the time required to complete the procedure. The developer's strive for the software system is to enable the system to run in real and delayed time.
The following mathematical methods have been implemented: statistical analysis of the HRV: time domain analysis (HRV Time Domain Measures); frequency analysis (obtaining numerical and graphical results): PSD (power spectral density) determination: Burg, Lomb-Scargle and wavelet method; nonlinear methods: Detrended Fluctuation Analysis, Poincare method; methods of time-domain analysis by calculating a periodogram (Burg methods, Lomb-Scargle and wavelet method).

3.3 Description of the database and results

A cardiological basis of long-term data obtained through Holter monitoring for solving the tasks of a project funded by the Scientific Research Fund of the Republic of Bulgaria is being established. The data is 24 hours long and is obtained from specially purchased Holter devices. It is planned to make about 2000 Holter records within the project for three years (daily patient recording is monitored). Holter records are performed on pre-selected patient groups, each group having the same cardiac disease. It is planned to carry out planned studies of approximately 20 patients in each group with the following diseases: myocardial infarction, heart failure, ischemic heart disease, arterial fibrillation, syncope, etc. Each patient will be examined initially every 3 months, and then every 6 months. The records will be stored in an electronic history of the disease and will serve to monitor the development of cardiac disease as well as the medications that have been applied and the effect of their use.
The results that result from the applied mathematical methods are numerical (tabulated) and graphical (histograms, graphs, spectrograms). The user can specify which of the results to be saved in a text file and to be included in the history of the subject's disease. The results obtained from the current study of the individual may be compared with the results of previous studies (when available). Comparative analysis can show how the patient's illness develops: there are signs of improvement in health or a tendency to deterioration, or the results do not show a sustained improvement in the condition of the individual under study. According to the results of the comparison of the results obtained from the analysis of the performed researches, appropriate measures can be taken to improve his / her health: to appoint new studies in order to detect other factors influencing the disease, changing the dose of the drugs or changing them medicines; dropping some of the medicines or appointing additional.

**Application of the information platform for education purposes**

The described information platform can be used for the purposes of medical education of the students - future cardiologists. Students can start the software system with different input cardiac data and monitor the results of the various mathematical analyses performed on the data. Exploring the resulting numerical and graphical results may be a preliminary experience from which they can be used in their future physician practice.

### 4 CLUSTERING OF CARDIAC DATA

The clustering of the cardiac data processed by the information platform is expected to play an important role in cardiology, both for the proper study of heart disease and for the application of effective patient treatment methodologies. [18] This is also the primary purpose of developed and presented above in this article information medical portals, WEB-based cloud, and interactive platforms. Clustering involves grouping the accumulated research experience in the field of diagnosis and treatment of cardiac diseases such as heart failure, atrial fibrillation, ischemic disease, and other cardiac diseases, defined by the World Health Organization. Such medical status could be described by measurable physiological parameters in specialized cardiological databases included in the platforms. For every group of diseases could be described effective treatment methods, recommended by cardiac centers [19]. In this paper are introduced the basic points of the methodology of extraction the useful information (pattern) from raw data by using different investigation and learning techniques. Such patterns can help the medical practitioner to understand the hidden relation (dependency) among the cardiac data and the effective treatment methods. With the using of the clustering can be forecasting the hidden trends in patients. The using of the correlation between the cardiac pattern and patients' status will help the practitioner to treat the patient wisely [20]. According to the nature of the cardiac data are developed three basic clusters of the patients’ cardiac status with medical information - normal, risk and critical patients. The information platform integrates different clustering algorithms and analytical methodologies for the performance of each algorithm in the cardiac dataset.

#### 4.1 Classification of patients with heart diseases

This article provides only a brief description of the clustering methodology as well as classification techniques where the data for patients with cardiovascular disease are grouped according to their condition. For the purpose of clustering, the information platform described in this article consists of information on patients with heart disease, which can be classified into three categories: Cardiac arrest, Heart attack, Heart failure [22]. In the process of diagnosis of cardiac diseases, the treating team can compare medical data about the patients studied with the existing classification data in the relevant cluster of the disease. The first step of classifying a patient with cardiovascular disease is when, based on medical research, the concerned patient is placed in one of the three mentioned categories. After entering medical data in the information platform, the classification is performed automatically by the developed software. Depending on the complexity of the disease, additional medical examinations are performed for the detail of the pathology [21].

As a partial example of the process, Table 1 presents the results of the classification of 92 patients with the following diseases: heart failure (Group 2), arrhythmia (Group 3), syncope (Group 4). The results are compared with healthy controls (Group 1). Study of the parameters of cardiac activity was performed by time domain analysis, frequency domain and nonlinear: Poincare plot, DFA, R/S method. The results are presented as mean ± standard deviation (mean ±sd). The differences between the values of the parameters of the test groups were tested by the ANOVA test and were considered reliable at a level of significance p<0.05.
The calculated time domain parameters for Heart Rate Variability (HRV) analysis are well known:

- mean RR interval value;
- SDNN - standard deviation of normal RR intervals;
- SDANN - calculate the mean values of normal RR intervals every 5 minutes of the study and calculate their standard deviation;
- RMSSD - the difference between each two adjacent normal NN intervals is squared and summed as the sum is divided by the number of intervals and the index is a square root of that number;
- pNN50 - the ratio of the number of adjacent NN intervals differing by more than 50ms (NN50) from the total number of NN intervals and multiplied by 100.

Table 1. Comparison of HRV measures between healthy controls (Group 1) and patients with heart failure (Group 2), arrhythmia (Group 3), syncope (Group 4).

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Group 1 (n=15)</th>
<th>Group 2 (n=38)</th>
<th>Group 3 (n=23)</th>
<th>Group 4 (n=16)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean±sd</td>
<td>mean±sd</td>
<td>mean±sd</td>
<td>mean±sd</td>
<td>Group1 vs Group2</td>
</tr>
<tr>
<td>Mean RR (ms)</td>
<td>1201.8±48</td>
<td>886±195</td>
<td>887.1±127</td>
<td>951.2±146</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Mean HR (bpm)</td>
<td>74.3±6</td>
<td>72.8±18</td>
<td>65.4±11</td>
<td>66.6±10.7</td>
<td>0.08</td>
</tr>
<tr>
<td>SDNN (ms)</td>
<td>141±48</td>
<td>121±51</td>
<td>122.9±49</td>
<td>140.9±38</td>
<td>0.4</td>
</tr>
<tr>
<td>SDANN (ms)</td>
<td>123±52</td>
<td>94.4±47.6</td>
<td>102±48</td>
<td>120.4±42</td>
<td>0.1</td>
</tr>
<tr>
<td>SDi (ms)</td>
<td>64±12</td>
<td>63.9±35.1</td>
<td>60.1±29.1</td>
<td>45.25±15</td>
<td>0.1</td>
</tr>
<tr>
<td>RMSSD</td>
<td>30.3±16.4</td>
<td>66.4±42.1</td>
<td>56.2±36.6</td>
<td>40.1±13.5</td>
<td>0.003</td>
</tr>
<tr>
<td>pNN50 (%)</td>
<td>14.2±11.1</td>
<td>28.6±25.1</td>
<td>22.7±22.2</td>
<td>10.2±8.8</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

Time domain analysis

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Group 1 (n=15)</th>
<th>Group 2 (n=38)</th>
<th>Group 3 (n=23)</th>
<th>Group 4 (n=16)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>VLF (ms²)</td>
<td>1324.5±120</td>
<td>1470±987</td>
<td>1175±908</td>
<td>863±99</td>
<td>0.7</td>
</tr>
<tr>
<td>LF (ms²)</td>
<td>1418.3±127</td>
<td>554.5±445</td>
<td>598.2±419</td>
<td>318.4±53</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>HF (ms²)</td>
<td>805.3±28.8</td>
<td>769.1±843</td>
<td>695.3±854</td>
<td>343.9±86</td>
<td>0.7</td>
</tr>
<tr>
<td>LF/HF</td>
<td>1.7±0.18</td>
<td>1.15±1.15</td>
<td>1.58±1.26</td>
<td>1.07±0.30</td>
<td>0.2</td>
</tr>
</tbody>
</table>

Frequency domain analysis

Nonlinear analysis

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Group 1 (n=15)</th>
<th>Group 2 (n=38)</th>
<th>Group 3 (n=23)</th>
<th>Group 4 (n=16)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>SD1 (ms) Poincare plot</td>
<td>26.9±13.1</td>
<td>46.9±29.8</td>
<td>39.7±25.9</td>
<td>28.3±9.6</td>
<td>0.2</td>
</tr>
<tr>
<td>SD2 (ms) Poincare plot</td>
<td>197.2±68.3</td>
<td>162.2±71</td>
<td>167.9±68</td>
<td>197.2±53</td>
<td>0.5</td>
</tr>
<tr>
<td>SD1/SD2 Poincare plot</td>
<td>0.14±0.02</td>
<td>0.32±0.19</td>
<td>0.25±0.13</td>
<td>0.15±0.04</td>
<td>&lt;0.05</td>
</tr>
<tr>
<td>Alpha 1 DFA</td>
<td>1.23±0.20</td>
<td>0.75±0.26</td>
<td>0.88±0.26</td>
<td>0.84±0.17</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Alpha 2 DFA</td>
<td>1.02±0.07</td>
<td>1.06±0.13</td>
<td>1.11±0.10</td>
<td>1.20±0.05</td>
<td>&lt;0.05</td>
</tr>
<tr>
<td>Alpha All DFA</td>
<td>1.06±0.10</td>
<td>0.98±0.15</td>
<td>1.07±0.13</td>
<td>1.12±0.05</td>
<td>&lt;0.05</td>
</tr>
<tr>
<td>Hurst exponent (R/S)</td>
<td>0.93±0.09</td>
<td>0.94±0.17</td>
<td>0.95±0.10</td>
<td>0.97±0.12</td>
<td>0.9</td>
</tr>
</tbody>
</table>

The frequency domain analysis gives the distribution of each of the multiple frequencies present in the RR intervals. Studies have shown that there are three discrete components (VLF, LF, HF) in the frequency domain, each related to certain physiological reasons.

HRV analysis based on non-linear mathematical methods (Poincare plot, DFA, R/S method) can provide important information about the physiological status of patients as well as the risk of sudden death. The short example presented through Table 1 shows the importance of rapid and accurate classification of patients with cardiovascular disease, in terms of taking proper treatment. This is one of the reasons for the continuous scientific development of mathematical and software methodologies.
for the analysis of cardiological status with the assistance of information platforms, supporting correct medical diagnoses.

5 RESULTS AND DISCUSSIONS

Currently, using methods for mathematical computerized analysis of cardiovascular parameters and evaluation of HRV parameters, used in the process of clustering in cardiology and clinical medicine, the described approaches for the interpretation of physiological and clinical data through models and patterns allow an effective solution of diagnostic and prognostic problems, based on accurate evaluation of the functional state, monitoring the effectiveness of therapy and preventive medicine, studied in Medical Universities.

In conclusion, the described classification and clustering methodology should consider only the aspects that can be attributed to the analysis of cardiac activity lasting from a few minutes to several hours. The research methodology and the principles for analysing registered medical data differ at higher levels in case of treatment of 24 or 48-hour HRV records obtained through Holter's monitoring. Various data from daily monitoring are available, which allows a mandatory assessment of the state of the mechanism of neuroendocrine regulation of cardiovascular circulation. The analysis of 24-hour studies is much longer and more expensive, so the process of categorizing and clustering of HRV is still not developed enough. The undisputed advantage of short analysis is the wider use of the method, the simplicity of the equipment and the software, the ability for quickly getting the results. All of this determines the prospects for the widest possible spread of HRV analysis in the process of study of the applied neurophysiology.

6 CONCLUSIONS

The report presents an overview of information platforms in the field of healthcare and medicine in Bulgaria and in the world. A model of an information platform for storing, processing and analyzing prolonged cardiac data is presented. The information platform includes a software system for the processing and analysis of cardiac data, a physiological database containing patient records. The information database is kept long-term and effective through the use of cloud technologies. The software cardiology system analyzes patients' cardiac data by linear and non-linear mathematical methods, fractal analysis and time-frequency analysis. The paper presents results obtained from the analysis of real data of patients with various cardiac diseases. The information platform described can be used to train students in cardiology.

ACKNOWLEDGEMENTS

This research was carried out as part of the project "Investigation of the application of new mathematical methods for the analysis of cardiac data" № KP-06-N22/5, 07.12.2019, founded by National Science Found, Bulgaria.

REFERENCES


