The Smart House based System for the Collection and Analysis of Medical Data

Yurii Kryvenchuk 1[0000-0002-2504-5833], Nataliya Shakhovska 1 [0000-0002-6875-8534],

Iryna Shvorob 1 [0000-0001-5569-2647], Sergio Montenegro 2 [0000-0002-0636-5866],

Maksym Nechepurenko 3 [0000-0002-9515-8841]

[1]Lviv Polytechnic National University, Lviv 79013, Ukraine,
[2]Julius-Maximilians-University Würzburg, Am Hubland, D-97074 Würzburg, German
[3]N-iX , Lviv, Ukraine
yurii.p.kryvenchuk@lpnu.ua, nataliya.b.shakhovska@lpnu.ua

Abstract. The analysis of personal patient information is one of critical factors for well-been paradigm. The collection and processing system of the medical data based on cloud computing, IoT in medicine and anticipation system for the deterioration of the patient's state on the basis of AI systems were constructed. The time series are used for future state prediction.

Keywords: smart house, medical data, time series, data analysis

1 Introduction

A problem of producing the so-called “Smart house for older persons and persons with physical disabilities” that provides a 24/7 health monitoring is intensively researched nowadays in many leading countries of the world. A rapid development of smart houses has become possible because of the recent fast progress in various computational intelligence (CI) techniques such as fuzzy logic, artificial neural networks and hybrid systems of CI. Thus, the smart houses are called sometimes “intelligent houses”.

The object of the investigation is to change the patient's condition at a certain time interval and to predict his condition with different methods of treatment. Also, an important point is the recognition of streaming video and the prevention of medical personnel about a particular case that occurred to the patient, for example, the patient fell. Based on the data obtained for a certain period, one can conduct an analysis and construct associative rules for each particular case. The conducted analytical survey showed that the following parameters, such as changes in body temperature, changes in blood pressure, cardiological data, are most often required to be monitored [1-8]. Since these indicators most accurately reflect changes in the human body, regardless of whether it is rapid changes, or slow. The aggregate of such data and their change in time will allow constructing associative rules by which one can predict a patient's condition for a certain period [9-11].

2 State of arts

Nowadays, Computational Intelligence (CI) methods and systems have been widespread to address a variety of Exploratory Data Analysis and Data Mining tasks, including as traditional: pattern recognition, classification, segmentation, and clustering, and are not directly related to this area of research. Intelligent management, defect detection, etc.

It is within this framework that the technology of computing intelligence can effectively process information (images and video streams from surveillance cameras) in conditions of uncertainty, nonlinearity, stochasticity, chaos, different types of disturbances and obstacles due to its universal approximating properties and learning opportunities based on experimental data characterizing functioning of the investigated phenomenon or object.

Nowadays, new areas such as Dynamic Data Mining, Data Stream Mining, Temporal Data Mining are based on the classic Data Mining. For these technics information is delivered in real time in the form of multidimensional time series, video streams, etc. Classical Neural Networks, Fuzzy Systems (fuzzy systems), the evolutionary algorithms were ineffective. Particularly, in case of Big data paradigm and information collection from different sources [12 – 13] it is very important to analyze the data in historical retrospective.

In [14] the patient state formalism is provided. The semantic search in heterogeneous environment was proposed.

The study of “Smart house for older persons and persons with physical disabilities” systems is carried out mainly on the analysis of multidimensional time series with the dynamics of specific indicators. The time series include methods of nonlinear analysis, determination of fractal dimension, phase portrait construction, phase portraits of decomposition on separate cycles, Lyapunov's indexes etc. [15 – 16]. Time series modeling and prediction widely used abroad, particularly in medical applications. The most famous approaches are: AR, MA, or ARMA univariate models. Approaches that are more sophisticated rely on nonlinear modeling and state space projection of the time series [22], Ralaivola et al. [22] present an approach for time series prediction based on kernel trick and supportvector regres-sion. In comparison, our approach is based on delay embedding and kernel regres-sion. The interesting for us is phase portraits and usage of constructed portrait for data prediction.The papers [17 – 20] emphasizes the need to protect personal data. In [22 – 25] the uncertainty in medical records are described.

The paper describes hardware-software system for personal medical data collection and processing.

3 The system architecture

3.1 Data collection

The complex of data collection and analysis consists of two subsystems. The first one is the local one, consists of the necessary sensors which allow measuring the body temperature, cardiological parameters, as well as the blood pressure index, if necessary, it is possible to expand the range of measuring parameters. Measured by sensor information enters the ESP8266 microcontroller, which collects and transmits data through a Wi-Fi router to a local server at intervals of 1 minute. If necessary, the system allows to change the time of the survey of sensors in the range from 1 second to several hours, therefore, depending on the patient's condition, the doctor can select the required time for updating and get the most actual information. Thus, it is possible to minimize the size of the information gathering unit to the size of the smartphone. In order to prevent the occurrence of heterogeneous problems associated with the system de-energized, an internal source of uninterruptible power is predicted to provide the device work for 60 minutes. Also, the system is equipped with a video camera or a video camera cascade, which makes it possible to see what happens to the patient in real time. The video stream is transmitting through a local server to the cloud, where it is storing for a certain period, which is requiring in each individual case (Fig. 1).

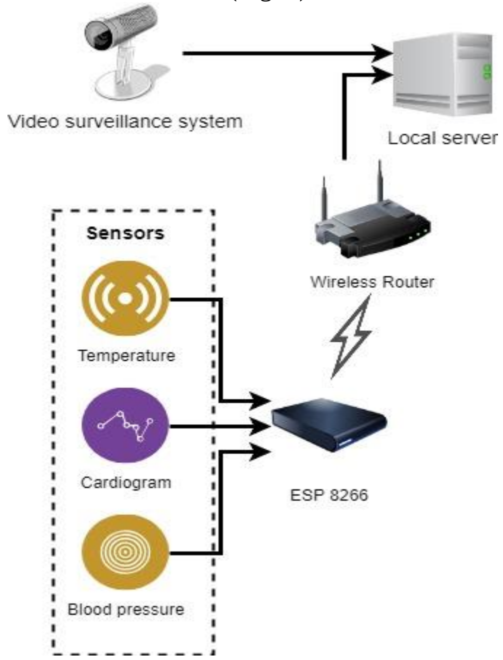

Fig. 1. The medical data collection system

Indicators of the patient's condition are quickly transmitted to the cloud where they accumulated into the database. A healthcare worker can analyze the patient’s state using the tablet or the smartphone. Also, it is possible to teach the system to notify a healthcare worker when the patient's condition changes, whether it is for the better or worse (Fig. 2).

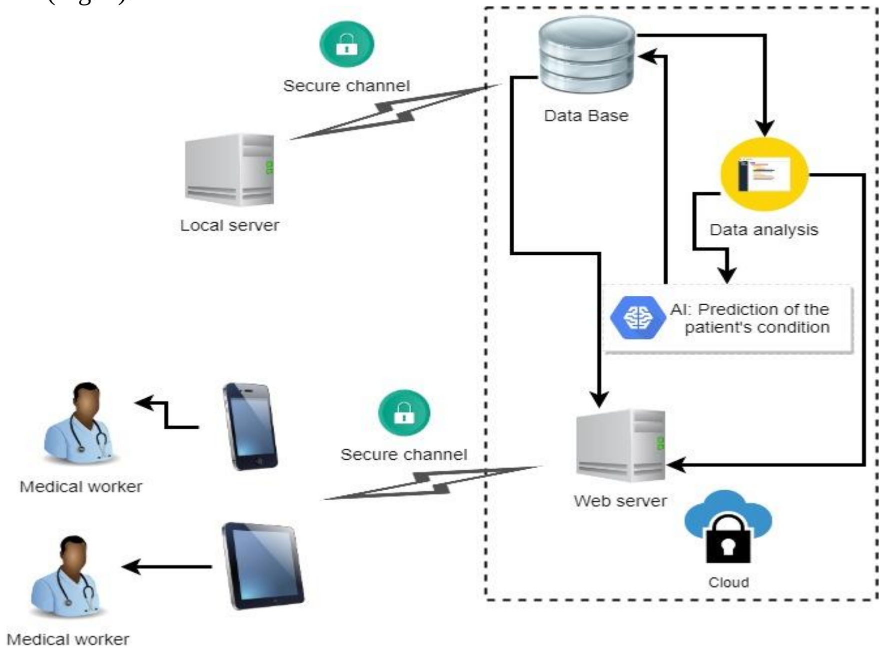

Fig. 2. The system for processing medical data and predicting the patient's condition

Measured information goes through a secure channel via the server to the database, where records are storing with information that has been receiving before. The healthcare worker can choose the period for which it is necessary to review the patient's medical data. In this case, the necessary information extracted from the database is sending to the web server where the data is visualized and returns the response through a secure channel on the healthcare worker's request. Also worth noting that the system has an AI block, which is responsible for both recognizing images from the video stream and for predicting the patient's condition. The video stream that reaches the cloud is immediately processed to identify the contents of the templates in it. It could be specified the number of templates, for example, if the patient fell from the bed, then the system signals the healthcare worker about this situation.

Prediction of the patient's condition is basing on the obtained data: if the patient does not intensively move, and he has abruptly increased blood pressure, then most likely he has a predominant state. In the case when the cardiogram shows problems in the work of the heart, the AI unit recognizes this, as well as assesses the condition and immediately warns the healthcare worker of the threat to the patient's life. In the same way one can foresee, and as a consequence, to warn a lot of the threats of the patient's life which need immediate medical intervention. AI unit can also be implemented as a system for predicting the patient's condition, taking into account his current state, and the choice of one or another method of treatment. The advantages of this method of implementation are that the doctor can more accurately predict the effectiveness of one or another method of treatment, based on his experience, data of the AI block, the dynamics of the growth of a disease.

3.2 Data processing

The future patient state prediction is made using time series [23, 24].

This database consist of a cell array of matrices, each cell is one record part consists of blood pressure and cardiological parameters. In each matrix, each row corresponds to one signal channel: PPG signal, FS=125Hz; photoplethysmograph from fingertip; ABP signal, FS=125Hz; invasive arterial blood pressure (mmHg; ECG signal, FS=125Hz; electrocardiogram. The data is consolidated from different sources such as sensors and medical records [25].Also the adequacy determination of personal medical profiles is provided [21] The data uncertainty is evaluated [22].

The prediction model is built using open database https://archive.ics.uci.edu/ml/datasets/Cuff-Less+Blood+Pressure+Estimation. Using this dataset the time series chars is built. The results are shown in Fig.3. To obtain data in the form of uniform data of a time series, it is tied to the number of test images. The fig. 3 shows the time series of the individual patient. A two-dimensional spatial matrix of distances is used. The purpose of the analysis is to search for repeating elements of a series.

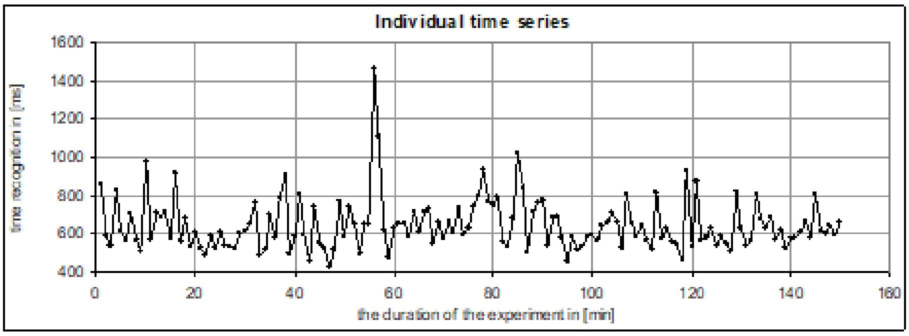

Fig. 3. The individual time series view of patient

The values of the elements of this time series actually characterize the common state of patient. Value is calculated as additive average number of the set of parameters.

The resulted data can be presented as discrete stochastic process [15]. The main indicators of descriptive statistics are analyzed: arithmetic average $\bar{y}_{mean}$, standard deviation, minimum $y_i = y_{\min}$ and the maximum $y_i = y_{\max}$ value, range $R = y_{\max} - y_{\min}$. Also, median, fashion, asymmetry and kurtosis are defined.

The model of the time series is presented as the additive model:

$$y_i(t) = m(t) + \phi(t) + \xi_t \,, \tag{1}$$

where $m(t)$ is the trend. It is slowly variable component. The next element is oscillatory component $\phi(t)$. This component presents the noise. This is the random variables normally distributed with a mean $m_\tau = 0$ and dispersion $s^2 = 1$, and they include measurement and calculation errors.

For a stationary time series with an increase in the lag, the values of the coefficients of autocorrelation should show a rapid monotonous decrease in absolute magnitude.

If case of non-stationary time series or closely to a stationary, the model is simplified like on equation:

$$y(t_i) = m(t_i) + \xi_t \tag{2}$$

Phase space reconstruction is performed by the delay embedding of the observed data into phase space vectors. The phase space that is constructed from $y(t)$ and $m(t)$ phase space of a time series based on Taken theory is constructed from a vector $[m(t - l), m(t - 2 * l), \ldots., m(t - n * l)]$ in which l is the delay and m is minimum Embedded Dimension of the time series [15, 21].

Phase analysis is realized as a reflection of differential functions in the original function. The series is uniform, and the bypass is smooth and monotonous function. To find the differential, the equation is used:

$$y' = (y_{i+1} - y_{i-1})/2h \,, \; y' = (-y_{i+2} + 8y_{i+1} - 8y_{i-1} + y_{i-2})/12h \,, \tag{3}$$

where i is the ordinal number of the level of the time series, h is the step between adjacent levels.

One solution is dual median smoothing, in which the sequence of smoothed data $\{y_i^{'}\}$ obtained in the first stage of processing the output timeline (with a sufficiently large m) is once again "matched" by the median filter with $m_2 \langle m_1$. An example of such processing is the so-called "Tuki 53" procedure, where $m_1 = 2$, $m_2 = 1$. After receiving the second smoothing series $\{y_i^{''}\}$, the final smoothed assessment is recommended to calculate by the formula:

$$\tilde{y}_i = 0,5 y_i^{''} + 0,25\left(y_{i-1}^{''} + y_{i+1}^{''}\right).$$

The polynomial trend model is in fact the multiple regression equation, so for its identification, fully applicable regression analysis methods and procedures. In particular, an expanded matrix of data will look like:

$$[Y+E,T]=[Z,T]=\begin{bmatrix} z_1 & 1 & 1 & 1 & \ldots & 1 \\ z_2 & 1 & 2 & 4 & \ldots & 2^p \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ z_t & 1 & t & t^2 & \ldots & t^p \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ z_n & 1 & n & n^2 & \ldots & n^p \end{bmatrix}=\begin{bmatrix} z_1 & \tau_1 \\ z_2 & \tau_2 \\ \ldots & \ldots \\ z_t & \tau_t \\ \ldots & \ldots \\ z_n & \tau_n \end{bmatrix}, \tag{4}$$

The vector of MSE-values looks like:

$$\tilde{A}=[\tilde{a}_0,\tilde{a}_1,\ldots,\tilde{a}_p]^T=(T^T T)^{-1} T^T Z. \tag{5}$$

3.3 Results

The analysis is made using R and RHRV package [26]. It allows us to load heart beat positions from sensors stream, to build the instantaneous Heart Rate (HR) series and filter it to eliminate spurious points. The next is the plot building for the the instantaneous HR series etc.

```
hr = CreateTimeAnalysis(hrv.data, size = 300,
        interval = 7.8125)
hr = CreateHRVData()
hr = SetVerbose(hrv.data,FALSE)
hr = LoadBeatAscii(hrv.data,"sensor.beats")
hr = BuildNIHR(hr)
hr = FilterNIHR(hr)
hr = SetVerbose(hr,TRUE)
hr = CreateTimeAnalysis(hr,size=400,interval = 7.7125)
PlotPowerBand(hr, indexFreqAnalysis = 1, ymax = 200,
ymaxratio = 1.7)
PlotPowerBand(hr, indexFreqAnalysis = 2, ymax = 700,
ymaxratio = 50)
```

The result is shown on Fig. 4.

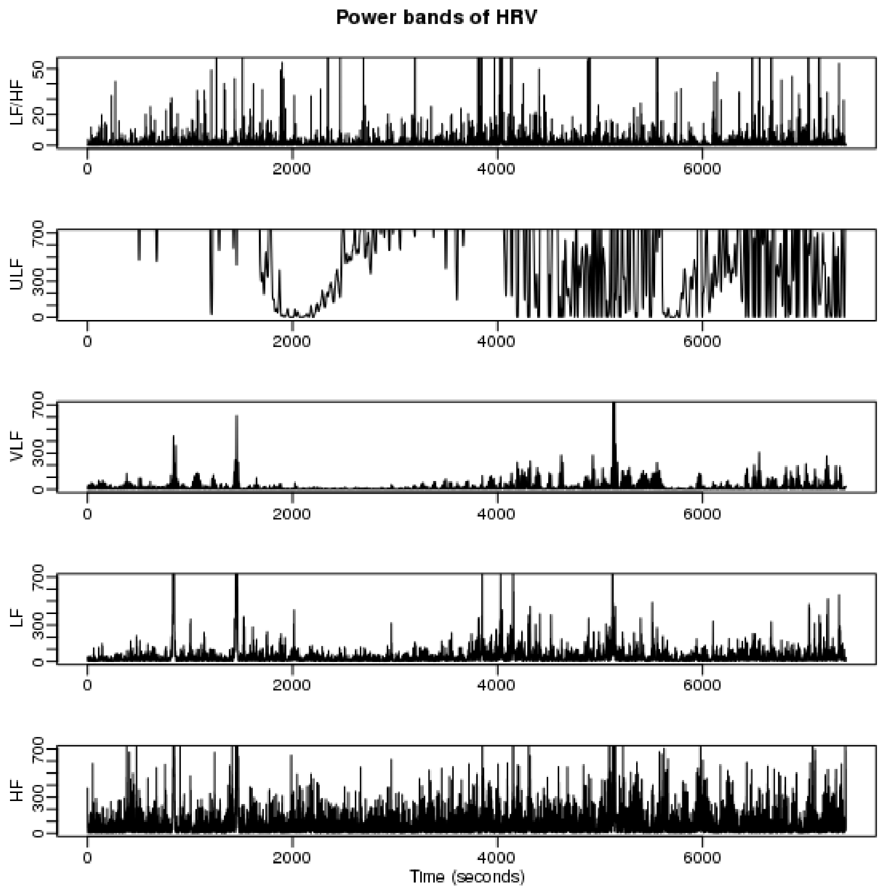

Fig. 4. The result of power bands of HRV plotting

The power spectrum obtained on the basis of Fourier analysis has fewer samples than the initial signal. In this regard, the power spectrum obtained from the wavelet analysis has the same number of samples as the original RR time series.

4 Conclusions

The paper presents hardware-software module for collecting and processing of personal medical information. The program is implemented in R language. Two approaches of time series are used. The SMA() function is used for data processing. The main purpose of this function is to smooth time series data with a simple moving average. The CreateTimeAnalysis is used for heart beats analysis. The Arima and own nonlinear dynamical system were used for prediction.

References

1. Godman, B., Malmström, R. E., Diogene, E., Gray, A., Jayathissa, S., Timoney, A., & Campbell, S. M.: Are new models needed to optimize the utilization of new medicines to sustain healthcare systems?. Expert review of clinical pharmacology, 8(1), 77-94 (2015).
2. Drakopoulos, G., Megalooikonomou, V.: A graph framework for multimodal medical information processing. In: 8th International Conference on eHealth, Telemedicine, and Social Medicine, pp. 278-282 (2016).
3. Ben-Assuli, O.: Electronic health records, adoption, quality of care, legal and privacy issues and their implementation in emergency departments. Health Policy, 119(3), 287-297 (2015).
4. Lorenzoni, L., Belloni, A., Sassi, F.: Health-care expenditure and health policy in the USA versus other high-spending OECD countries. The Lancet, 384 (9937), 83-92 (2014).
5. van Staa, T. P., Klungel, O., Smeeth, L.: Use of electronic healthcare records in large-scale simple randomized trials at the point of care for the documentation of value-based medicine. Journal of internal medicine, 275(6), 562-569 (2014).
6. Bender, D., Sartipi, K.: HL7FHIR: An Agile and RESTful approach to healthcare information exchange. In: Proceedings of the 26th IEEE International Symposium on Computer Based Medical Systems, pp. 326–331 (2013).
7. Vitek, C. R. R., Nicholson, W. T., Schultz, C., Caraballo, P.: Evaluation of the use of clinical decision support and online resources for pharmacogenomics education. Pharmacogenomics, 16(14), 1595-1603 (2015).
8. Qassas, M. A., & et al.: ArgMed: A Support System for Medical Decision Making Based on the Analysis of Clinical Discussions. Real-World Decision Support Systems. Integrated Series in Information Systems, vol 37. Springer, pp. 15–41 (2016).
9. Liu, S., Ma, W., Moore, R., Ganesan V., Nelson, S.: RxNorm: prescription for electronic drug information exchange. IT Professional, vol. 7, no. 5, pp. 17-23 (2005).
10. Kern, L. M., Edwards, A. M., Pichardo, M., Kaushal, R.: Electronic health records and health care quality over time in a federally qualified health center. Journal of the American Medical Informatics Association, 22(2), 453-458 (2015).
11. Giannini, B., & et al.: A web based tool to enhance monitoring and retention in care for tuberculosis affected patients. Stud Health Technol Inform, 237, 204-208 (2017).
12. Melnykova, N.: Semantic search personalized data as special method of processing medical information. Advances in Intelligent Systems and Computing, pp. 315-325. Springer, (2017).
13. Shakhovska, N., Nych, L., Kaminskyj, R.: The identification of the operator's systems images using the method of the phase portrait. Advances in Intelligent Systems and Computing, pp. 241-253. Springer (2017).
14. Nguyen, L., Bellucci, E., Nguyen, L. T.: Electronic health records implementation: an evaluation of information system impact and contingency factors. International journal of medical informatics, 83(11), 779-796 (2014).
15. Krist, A. H., & et al.: Electronic health record functionality needed to better support primary care. Journal of the American Medical Informatics Association, 21(5), 764-771 (2014).
16. Pearce, C., Bainbridge, M.: A personally controlled electronic health record for Australia. Journal of the American Medical Informatics Association, 21(4), 707-713 (2014).
17. Clegg, A., & et al.: Development and validation of an electronic frailty index using routine primary care electronic health record data. Age and ageing, 45(3), 353-360 (2016).
18. Hassanalieragh, M., & et al.: Health monitoring and management using Internet-of-Things (IoT) sensing with cloud-based processing: Opportunities and challenges. In: IEEE international conference on services computing, pp. 285-292, (2015).

19. Korzh R., Peleshchyshyn A., Fedushko S., Syerov Y. Protection of University Information Image from Focused Aggressive Actions. Recent Advances in Systems, Control and Information Technology. SCIT 2016. Advances in Intelligent Systems and Computing, vol 543. Springer, 104-110 (2017).
20. Ralaivola, L.: Dynamical modeling with kernels for nonlinear time series prediction. In Advances in neural information processing systems, pp. 129-136 (2004).
21. Syerov, Y., Shakhovska, N., Fedushko S.: Method of the data adequacy determination of personal medical profiles (in press).
22. Shakhovska, N., Vovk, O., Kryvenchuk, Y.: Uncertainty reduction in Big data catalogue for information product quality evaluation. Eastern European Journal of Enterprise Technologies, Volume 1, Issue 2-91, 12-20 (2018).
23. Alanazi, H. O., & et al.: Meeting the security requirements of electronic medical records in the ERA of high-speed computing. Journal of medical systems, 39(1), 165 (2015).
24. Miotto, R., Li, L., Kidd, B. A., & Dudley, J. T.: Deep patient: an unsupervised representation to predict the future of patients from the electronic health records. Scientific reports, 6, 26094 (2016).
25. Shakhovska, N.: Consolidated processing for differential information products. In: Proceedings of VIIth International Conference on Perspective Technologies and Methods in MEMS Design (MEMSTECH), pp. 176-177 (2011).
26. RHRV Quick Start Tutorial, https://cran.r-project.org/web/packages/RHRV/vignettes/RHRV]-quickstart.html, last accessed 2018/10/3.