Identifying Chronic Disease Complications Utilizing State of the Art Data Fusion Methodologies and Signal Processing Algorithms

John Gialelis, Petros Chondros, Dimitrios Karadimas, Sofia Dima, and Dimitrios Serpanos

Industrial Systems Institute/RC Athena,
PSP Bld, Stadiou St., GR26504 Platani Patras, Greece
{gialelis,serpanos}@isi.gr,
{pchondros,karadimas,sdima}@ece.upatras.gr

Abstract. In this paper a methodology for identifying patient's chronic disease complications is proposed. This methodology consists of two steps: a. application of wavelet algorithms on ECG signal in order to extract specific features and b. fusion of the extracted information from the ECG signal with information from other sensors (i.e., body temperature, environment temperature, sweating index, etc.) in order to assess the health state of a monitoring patient. Therefore, the objective of this methodology is to derive semantically enriched information by discovering abnormalities at one hand detect associations and inter-dependencies among the signals at the other hand and finally highlight patterns and provide configuration rulesets for an intelligent local rule engine. The added value of the semantic enrichment process refers to the discovery of specific features and meaningful information with respect to the personalized needs of each patient.

Keywords: patient monitoring, algorithms, data fusion.

1 Introduction

Heart disease is the most important cause of death in many countries. Thus an automated solution of pervasive heart monitoring is required in order to take care of senior chronic heart patients. The electrocardiogram is an important signal for providing information about functional status of the heart. It shows how fast the heart is beating, whether the heart beat is steady or irregular and the strength and timing of electrical signals as they pass through the heart. The processing of the ECG signal in order to extract specific features of the P wave, the QRS complex, the T wave and ST segment, which in conjunction with other physiological parameters such as blood pressure and sweating index, is of great importance in the detection of cardiac anomalies. Furthermore, it is essential to acquire physiological signals from any type of environment - clinical, domestic, rural and urban - accurately and properly classify them so physicians or medical experts are able to correctly evaluate the patient's

K.S. Nikita et al. (Eds.): MobiHealth 2011, LNICST 83, pp. 256–263, 2012.

health status and perform the appropriate actions. Therefore it is essential to follow an approach using various signals and features in order to measure the same underlying clinical phenomenon: the gradually worsening condition of a chronic patient. The aim is to improve the quality and robustness of the indicator, thus improving sensitivity without causing false alarms. The final goal is to come up with a system that measures a number of parameters from easily applicable low-cost sensors, and simultaneously forms a powerful diagnostic tool. The envisaged high level architecture of such a system is depicted in Fig. 1.

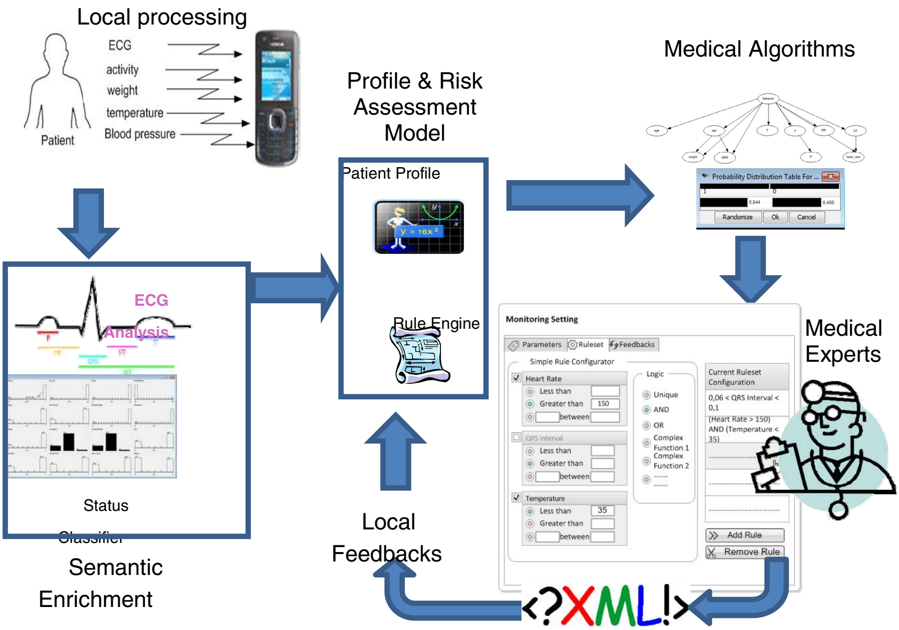

Fig. 1. The envisaged high level architecture of the proposed system

2 Signal Processing Algorithms for ECG Features Extraction

In the clinical domain of patients with chronic heart disease physicians can evaluate patients' health status based on ECG signal analysis. Furthermore, the extraction of ECG features, as they are depicted in Fig. 2, enables the medical experts not only to evaluate heart problems, such as arrhythmia and cardiovascular diseases, but we argue that additional features and meaningful information with respect to the personalized needs of a patient could be discovered.

ECG signals are gathered through appropriate sensors. The measurement of an ECG signal always imposes noise and artifacts within the frequency band of interest.

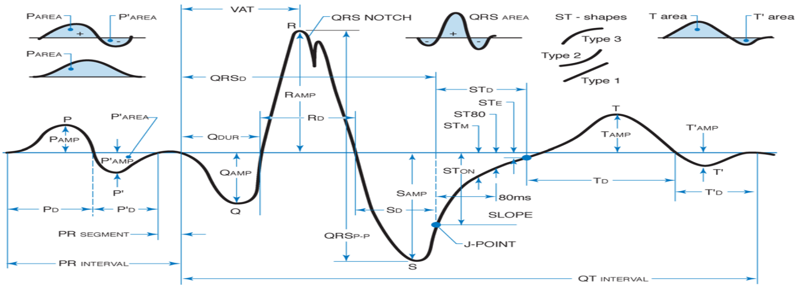

Fig. 2. ECG waveform features

Therefore suitable methods and algorithms have been developed in order to derive the useful information from the noisy ECG signals. Specifically, a preprocessing stage removing noise and/or artifacts from the raw ECG results the starting point of ECG analysis. Wavelet based algorithms can provide efficient localization in both time and frequency [1] thus constitute the most efficient and commonly used approach. The detection of fiducial points [2] which are relevant and strongly correlated to several heart diseases [3] is performed through a parameter extraction stage, as depicted in Fig. 3. This task becomes even more challenging due to the presence of artifacts and time-varying morphology. The key point is to provide a description of the ECG signal in the time-scale domain allowing the representation of the temporal features of a signal at different resolutions.

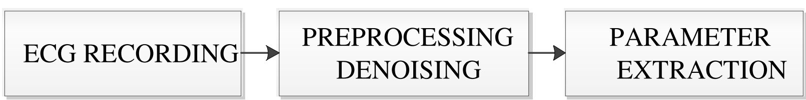

Fig. 3. ECG signal processing for parameter extraction

Furthermore the basic steps of the ECG signal processing procedure, starting with the QRS complex detection, are depicted in Fig. 4. With respect to the QRS complex, Q and S peaks occur before and after R peak, respectively. R peak is detected on the first scale of Wavelet Analysis. Multi resolution analysis of Wavelet Transform on time-domain ECG is utilized at different Wavelet scales, and the modulus-maxima and zero-crossing approach has been exploited at each resolution level. At the second resolution level, QRS onset and offset (J point) can be identified by using modulus-maxima on the wavelet coefficients. Since wavelet analysis preserves the time domain information of the original signal, the location of these characteristic points (Q, R and S) can be easily obtained from the wavelet domain. Finally, T wave analysis is achieved by considering details at lower frequencies. At the fifth resolution level of wavelet analysis, T wave is detectable using the modulus-maxima based approach on the wavelet coefficients. Appropriate time windows and amplitude thresholds are required in order to get rid of irregular points and misdetections. This is essential so as to provide a robust ECG signal processing methodology.

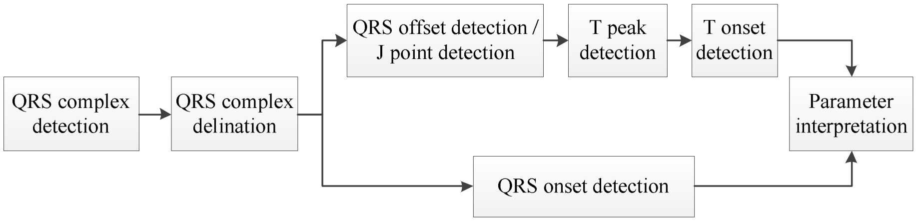

Fig. 4. ECG signal processing procedure

3 Data Fusion Model

An appropriate model describing a fusion process of data obtained from body area sensors, EHR (Electronic Health Records) and/or other historical data has been utilized. Based on these raw data, patient's characteristics and expert's knowledge, a mathematical framework is proposed to extract knowledge. The framework's basic goal is to determine the patient's health status and potentially to be used as a medical decision support tool.

Specifically, medical data, as they may come from different sources, are in different format and types. Typical examples are ECG, medical images, historical data and also habits, such as smoking, exercising etc. Due to the unique nature of these data and in order to extract knowledge, data mining approach is needed as any patterns found should be capable of human interpretation. Moreover relationships or patterns that are extracted may not be commonly accepted or conform to current medical knowledge. Results may indicate an association between the physiological parameters and the class (medical condition of the patient) but there is no implication of cause and effect. The interpretation of these associations is fully up to the experts, so additional visualization and statistical processing is needed to present results in a human interpretable format. Moreover, as the evaluation of the results fully depends on the experts, the proposed methodology can only serve as a decision support tool regarding the prognosis and the diagnosis of the monitoring patient.

There are two main approaches during the design process: supervised learning and unsupervised learning. The basic requirement of classification of patient's health status using supervised learning algorithms is that the dataset must be annotated. The most commonly used algorithms are C4.5 (decision tree approach), Multilayer Perceptron and Naïve Bayes. Unsupervised learning's goal is to determine how a set of data is structured. Clustering and blind source separations are typical unsupervised learning techniques.

Having WEKA [4] available, it was decided to use both supervised and unsupervised machine learning techniques. Besides the typical classification problems, there is an additional need to discover any associations between attributes (physiological parameters). The resulting relations and rules may enlighten hidden interdependencies and provide the experts with meaningful information regarding the patient's health status.

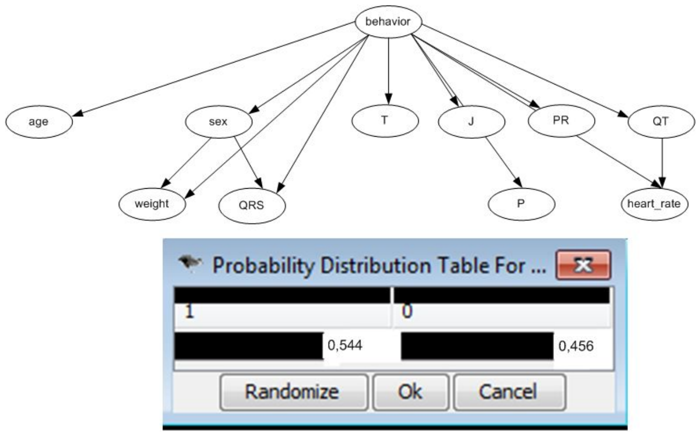

Fig. 5. Screenshot of the Weka tool depicting the developing classifier

In Fig. 5, a screenshot of the Weka tool is depicted. Specifically, the classification process is implemented using NaiveBayes algorithm with 3-parent configuration. The class 'behavior' represents the health status of the patient that can be either normal or abnormal, while the selected attributes that form the annotated dataset are general personal data (age, sex, weight), physiological monitoring parameters (body temperature), as well as extracted ECG features (QRS duration, J point, P wave, Q-T duration, Heart Rate, P-P duration). The training of the model ended up to the structure depicted and indicates some kind of relationship between P wave and P-P duration as well as QT duration and heart rate regarding the ECG extracted features. During testing and run of different algorithms and approaches, it was made clear that results were strongly depended on the dataset we were using, meaning that the algorithms' performance is highly correlated to the size and the quality (missed values) of the used dataset. Therefore, the above findings may trigger the expert for further investigation of the proposed associations outside the scope of the machine learning approach.

4 Methodology for Patient's Personalized Monitoring

The aforementioned data fusion techniques aim to support the doctors and the medical experts in general and also to potentially contribute in discovering additional personalized medical features that will provide information about patients' health status. Under this scope, personalized monitoring of patients could be conclusively accomplished as a workflow mechanism, and lead the doctors to a decision driven by patient's unique parameters as depicted in Fig. 6.

As previously described, the applied data fusion models regularly investigate sensor data along with ECG extracted features for irregularities, associations and inter-dependencies so as to detect abnormal patterns. Once an irregularity or an

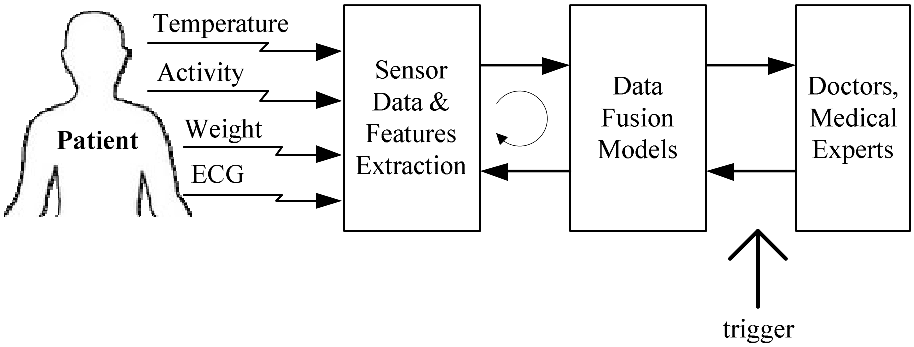

Fig. 6. Workflow mechanism exploiting data fusion assets for conducting doctors personalized monitoring

abnormal pattern is discovered, a trigger signal is provided to the medical expert. After this, the medical expert is able to re-assess the set of rules for the specific patient through a simple set of user interfaces (UIs) as depicted in Fig. 7.

The first step for the re-assessment involves the adding or the subtracting of physiological parameters to be monitored depending on the findings of the data fusion model, designated as “Machine Learning (ML) Findings”.

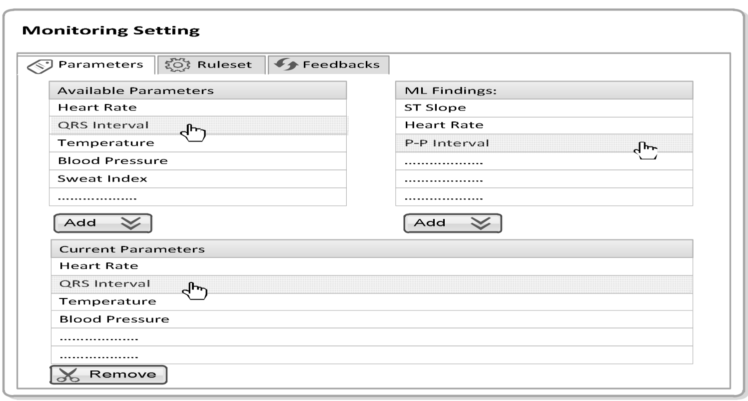

Fig. 7. Adding physiological parameters to be monitored

At a second step, the definition of a set of rules, according to the doctor’s expertise and experience, complements patient’s treatment towards personalization. The personalization concept lies on the definition of different physiological parameters ranges that correspond to the normal state for each patient. The key feature of personalization is associated with the ability to create rules with the selected physiological parameters as presented in Fig. 8. So far the proposed methodology will provide the doctors and medical experts a ruleset configuration covering almost any

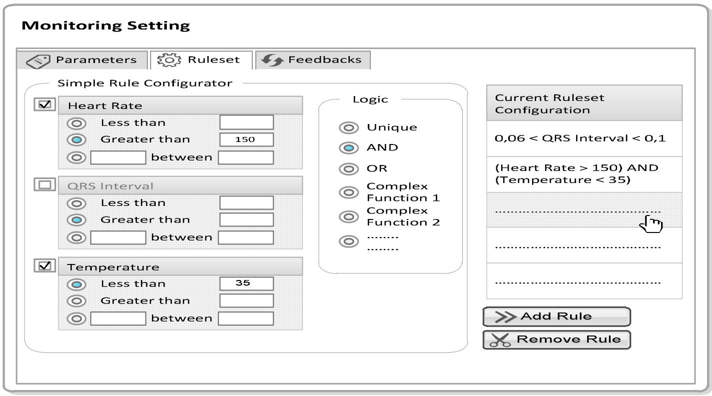

Fig. 8. Creation of simple rulesets

potential irregularity for a given patient, since this ruleset will be constructed based not only on doctors' aspect but also on identified abnormalities issued by the "Machine Learning Findings" as described previously.

Eventually, the constructed ruleset configuration, derived in an interoperable XML-based format such as PMML v4.0 Standard [5], is forwarded to an appropriate, operated on patient's portable device, application that monitors and records his health status. Moreover, doctors and medical experts are also able to predefine a set of feedback actions that will be executed from that application in case of a hit in any of the rules within the defined ruleset, as depicted in Fig. 9.

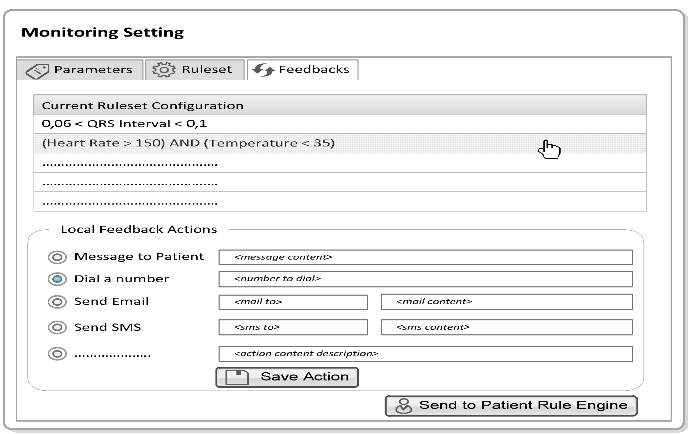

Fig. 9. Defining feedback actions and exporting configuration rulesets

5 Conclusion

In this paper high level system architecture for identifying patient's chronic disease complications is depicted. Furthermore, a methodology was proposed. This methodology which consists of two tasks: application of algorithms on primary sensor readings (e.g. ECG) in order to derive semantically enriched information and a data fusion model that uses the above extracted information as well as data from other sensors (body temperature, environment temperature, sweating index, etc.) in order to assess the health state of a chronic heart disease person. Our goal is to investigate data irregularities, associations and inter-dependencies as well as to detect abnormal patterns and provide interpretation of continuous data which can support an intelligent rule engine machine efficiently.

Acknowledgment. The presented work is taken place in the framework of the ARTEMIS JU Project CHIRON (ARTEMIS - 2009-1 100228 - http://www.chiron-project.eu).

References

1. Li, C., Zheng, C., Tai, C.: Detection of ECG characteristic points using wavelet transforms. IEEE Transactions on Biomedical Engineering 42(1), 21–28 (1995)
2. Martinez, J.P., Almeida, R., Olmos, S., Rocha, A.P., Laguna, P.: A wavelet-based ECG delineator: evaluation on standard databases. IEEE Transactions on Biomedical Engineering 51(4), 570–581 (2004)
3. Choi, S.: Detection of valvular heart disorders using wavelet packet decomposition and support vector machine. Expert Systems with Applications 35(4), 1679–1687 (2008)
4. WEKA, `http://www.cs.waikato.ac.nz/ml/weka/`
5. PMML v4.0 Standard, `http://www.dmg.org/v4-0-1/GeneralStructure.html`