

Detection of arrhythmic events by means of trend prediction of physiological time series in pHealth systems

Teresa Rocha*, Simão Paredes*, Paulo de Carvalho**, Jorge Henriques**

* *Instituto Politécnico de Coimbra, Departamento de Engenharia Informática e de Sistemas, Coimbra Portugal, {teresa@isec.pt, sparedes@isec.pt}.*

** *CISUC, Departamento de Engenharia Informática, Universidade de Coimbra, Coimbra, Portugal, {carvalho@dei.uc.pt, jh@dei.uc.pt}*

Abstract: This work addresses the research and implementation of methodologies for time series prediction, mainly to support the early detection of critical events. To this aim, a wavelet decomposition based scheme is proposed to estimate the evolution trend of physiological signals. The scheme does not involve the explicit development of a model and is essentially supported on the hypothesis that future evolution of a biosignal can be estimated from similar historic patterns. The strategy considers an à-trous wavelet decomposition, where the most representative trends are extracted from the historic similar patterns. Then, a set of distance-based measures able to assess the prediction likelihood of each representative trend is introduced. From these measures and through an optimization process, a subset of these trends is selected and aggregated to derive the required time series evolution trend. The performance of the scheme was, in a first phase, compared with other typical prediction strategies, namely the autoregressive integral moving average model and the generalized regression neural network. Then, in a second phase, the effectiveness of the methodology was tested in the assessment of arrhythmic episodes using heart rate measurements computed from ECG signals collected in the context of TEN-HMS telemonitoring study.

Keywords: telemonitoring systems; arrhythmias: prediction; wavelets.

1. INTRODUCTION

Cardiovascular disease (CVD), a general name for a wide diversity of diseases, disorders and conditions that affect the heart and often the blood vessels, is a major cause of disability and premature death throughout the world. Cardiovascular diseases can take many forms, such as raised blood pressure (hypertension), cardiac arrhythmias (abnormal heart rhythms), coronary artery disease, heart attack (myocardial infarction) and cerebrovascular disease (stroke). On the other hand, it is well known, the heart health is linked to behavior and lifestyle. Therefore, the focus should be on prevention, recognized by clinical professionals as the best method to avoid diseases from happening. In fact, preventive medicine is more cost effective, that is, able to obtain the same outcome in terms of health, with smaller costs, less pain, and within a short time.

In the context of preventive medicine, telemonitoring solutions are making a huge impact by enabling the remote monitoring of patients. In effect, the information collected during long periods allows for advances in the diagnosing of a disease, for the description of its evolution and for the prediction of possible complications, including the early prevention of the occurrence of severe events that may require hospitalization.

Cardiac arrhythmias are among the top most factors associated with cardiovascular diseases. In fact, abnormal rhythms can result in a multiplicity of alterations in the myocardial structure, coronary vasculature, and conduction system of the heart, which can lead to the development of left ventricular hypertrophy, coronary artery disease, myocardial infarction, cardiac arrhythmias, heart failure, among others (Sign, 2007). Therefore, the development of methodologies able to detect not only the occurrence of these arrhythmias but also their evolution trends is of extreme importance for the conception of early prevention systems.

This work aims at the development of a predictive methodology to estimate biosignals' future values. In particular, the research addresses the trend analysis, in order to predict if the HR of a given patient evolves towards high values or, on the contrary, is maintaining or decreasing to normal values. The strategy is founded on the hypothesis that the estimation of biosignals' future evolution can be supported on current and past measurements taken from historical data of a group of patients, including the patient under study. For this purpose three main stages are considered: *i*) the selection of patients who display similar behaviors in their physiological time series is carried out by means of a similarity analysis process; *ii*) then, an estimation of the biosignal's future values is performed, based on the similar time series identified in the first phase; *iii*) finally, the estimated values are used in the assessment of the arrhythmias risk for the patient under study.

Regarding methodologies for similarity search in time series, significant advances have been made. The simplest algorithms used the Euclidean distance between raw time series of the same length. Others proposed dynamic time warping for time series of different lengths (Park et al, 2000). However, due to the high dimensionality of time series, most of the approaches perform dimension reduction on data. Among them, some used discrete Fourier transform (Agrawal et al, 1993), principal component analysis (Karhunen-Loève transform) (Yang and Shahabi, 2004), and wavelet transform (Saeed and Mark, 2006). Rocha et al (2012) presented an innovative scheme able to efficiently evaluate the similarity between two physiological time series. It combines the Haar wavelet with the Karhunen Loève transforms, allowing for the representation of signals as a linear combination of an optimal reduced set of basis functions. Through an iterative algorithm for computing the expansion coefficients, the computational complexity of the method is significantly decreased.

Among prediction techniques, linear regression methods, such as autoregressive structures, have been the most used in practice. However, linear models are not always adequate for biosignals, since they are non-linear to some extent. Among the non-linear methods, neural networks became very popular mainly due to their universal approximation properties. Many different types of neural networks, such as time delay and recurrent neural networks, demonstrated to be effective for time series modeling and prediction (Haykin, 2008). On the other hand, in most clinical cases, an assumption of global stationarity can not be considered. Among time-frequency methods, wavelet transform that provides a good local representation of the signal in both the time and frequency domains, offering an appropriate framework to deal with non-stationarities, has been frequently applied. Although the wavelet transform itself is not a forecasting methodology, it may be incorporated in hybrid prediction schemes involving the multi-resolution decomposition of signals (Fryzlewicz et al, 2003).

This work, starting from the similarity measure and indexing scheme proposed by the same authors (Rocha et al, 2012), presents a strategy based on Haar “à-trous” wavelet decomposition for the prediction of biosignals. The proposed approach does not intend to perform an accurate prediction, but to obtain a reasonable forecast of the future trend. Basically, from the wavelet decomposition of similar signals retrieved from the historic, the most representative trends at the several decomposition levels are identified and combined through an optimization process, directly providing an estimation of the current time series evolution.

The structure of the paper is as follows: section 2 describes the proposed wavelet decomposition scheme and section 3 presents its application to heart rate signals for the assessment of arrhythmia risk, using data collected during TEN-HMS (Cleland et al, 2005) tele-monitoring study. Finally, in section 4, some conclusions are drawn.

2. METHODOLOGIES

Figure 1 illustrates the idea behind the proposed prediction approach. Basically, the process starts by considering the current signal to be predicted, designated here as the template, $X(t) \in \mathbb{R}^{1,N}$. Using the template and from a similarity analysis procedure **, the set of the M most similar conditions (patterns) $\mathbf{X}(t) \equiv \{X_m(t) \in \mathbb{R}^{1,N}\}$, $m=1, \dots, M$, is identified. From these, the corresponding subsequent P future values, $\mathbf{Y}(t) \equiv \{Y_m(t) \in \mathbb{R}^{1,P}\}$, are straightforwardly obtained (known past values from historic dataset). Then, the known “future” evolution of the identified patterns, $\mathbf{Y}(t) \equiv \{Y_m(t)\}$, can be used in a prediction mechanism to estimate the evolution of the current template, $\hat{Y}(t) \in \mathbb{R}^{1,P}$.

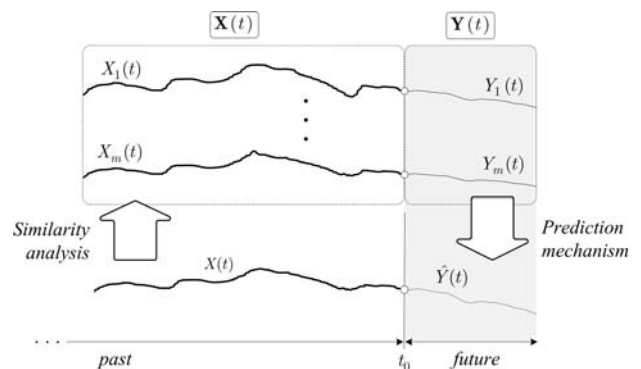


Fig. 1. Prediction strategy based on similar signals.

The global set of patterns, $\mathbf{Z}(t) \in \mathbb{R}^{M,N+P}$, is thus composed of two components, $\mathbf{X}(t)$ and $\mathbf{Y}(t)$, in the form of (1).

$$\mathbf{Z}(t) = [\mathbf{X}(t) \ \mathbf{Y}(t)] \quad (1)$$

Figure 2 depicts the main steps involved in the estimation of $\hat{Y}(t)$, based on the similar patterns $\mathbf{Z}(t)$ and through a wavelet decomposition scheme.

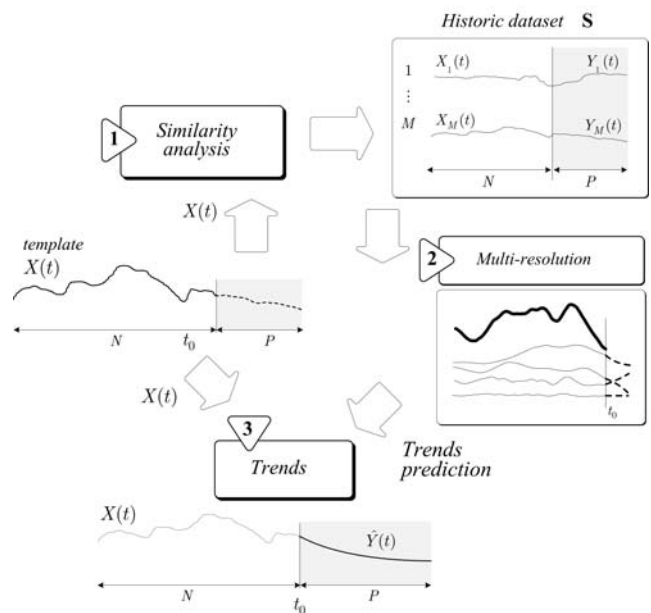


Fig. 2. Prediction methodology.

2.1. Template decomposition

The second step involves the determination, at each decomposition level, of the most representative time series (trends) from the retrieved similar historic signals. To achieve this goal, the historic signals are in a first phase decomposed using the “à-trous” wavelet (Renaud et al, 2003), according to (2).

$$W\{Z_m\} = \{d^l Z_m(t), a^l Z_m(t)\}, l = 1, \dots, L \quad (2)$$

The variables $a^l Z_m(t)$ and $d^l Z_m(t)$, $m = 1, \dots, M$, represent, respectively, the approximation and the details. It is important to note that, in this case, the decomposition can be extended to the “future” (time instants from $N+1$ to $N+P$), with length $N+P$, that is, $W\{Z_m\} \in R^{L+1, N+P}$.

Then, in a second phase, the representative decomposition trend at each level is determined through a clustering process. In this case the subtractive method was employed (Kriegel et al 2003).

$$d^l \bar{Z}(t) = \text{subClustering} \{d^l Z_m(t)\}, m = 1, \dots, M, l = 1, \dots, L \quad (3)$$

$$a^l \bar{Z}(t) = \text{subClustering} \{a^l Z_m(t)\}, m = 1, \dots, M \quad (4)$$

The variables $d^l \bar{Z}(t) \in R^{1, N+P}$ and $a^l \bar{Z}(t) \in R^{1, N+P}$ denote, respectively, the representative details and approximation.

2.2 Optimal Trends

In this step, the representative trends are reduced to an optimal set, that is, to a set of trends (decomposition levels) that have the potential to contribute to a consistent prediction. To this purpose, a combination process comprising the minimization of a set of distance-based measures, that assess the likelihood that a representative trend will contribute to a correct estimation, is implemented.

a. Distance-based measures

The distance-based measures are computed for each decomposition level $l = 1, \dots, L+1$, where $L+1$ stands for the approximation, using: i) the template $X(t) \in R^{1, N}$; ii) the corresponding wavelet decomposition at l level, $d^l X(t) \in R^{1, N}$; iii) the wavelet decomposition of similar patterns at the same level, $d^l Z_m(t) \in R^{1, N+P}$, $m = 1, \dots, M$; iv) the corresponding clustering, that is, the representative trends $d^l \bar{Z}(t) \in R^{1, N+P}$. Using these signals, a set of distance-based measures θ_i^l is computed as follows, for $m = 1, \dots, M$:

$$\theta_1^l = S(X(t), d^l X(t)), t = 1, \dots, N \quad (5)$$

$$\theta_2^l = S(d^l X(t), d^l \bar{Z}(t)), t = 1, \dots, N \quad (6)$$

$$\theta_3^l = \text{mean}[S(d^l X(t), d^l Z_m(t))], t = 1, \dots, N \quad (7)$$

$$\theta_4^l = \exp[-\text{std}\{S(d^l X(t), d^l Z_m(t))\}], t = 1, \dots, N \quad (8)$$

$$\theta_5^l = \text{mean}[S(d^l \bar{Z}(t), d^l Z_m(t))], t = N+1, \dots, N+P \quad (9)$$

$$\theta_6^l = \exp[-\text{std}\{S(d^l \bar{Z}(t), d^l Z_m(t))\}], t = N+1, \dots, N+P \quad (10)$$

As result, a vector composed of six measures is obtained, $\Theta^l = [\theta_1^l, \theta_2^l, \dots, \theta_6^l]$, where each $\theta_i^l \in [0, 1]$.

The measure $S(X_1(t), X_2(t)) \in [0, 1]$, is a normalized similarity measure, where a value of 1 means a total agreement between the signals. The operators $\text{mean}(\cdot)$ and $\text{std}(\cdot)$ denote, respectively, the mean and standard-deviation operators.

b. Selection of the optimal trends

The optimization strategy assumes that each of the parameters θ_i^l defines a measure that enables to assess the quality of each representative trend. Therefore, according to the obtained measures, a decision regarding the inclusion or exclusion of a specific representative trend in the optimal set can be taken. To support this decision the operators $\text{maximum}(\cdot)$ and $\text{product}(\cdot)$ are employed, respectively, as the aggregation and conjunction operators. As result, the quality of a specific prediction is assessed according to (11).

$$p(\Theta) = \max\{p(\Theta\sigma_1), \dots, p(\Theta\sigma_i), \dots, p(\Theta\sigma_n)\} \quad (11)$$

Each variable $p(\Theta\sigma_i)$ denotes the possible decomposition level combinations, resulting from the operator $C(nL, nN)$ (combinations of nL taken nN at a time). In this process, the conjunction of the metrics corresponding to a specific level is given by the $\text{product}(\cdot)$ operation (12).

$$p(\Theta^l) = \prod_{i=1}^6 \theta_i^l, l = 1, \dots, L+1 \quad (12)$$

If two levels L_i and L_j are combined, the corresponding aggregation $p(\Theta\sigma_i)$ is obtained according to (13).

$$p(\Theta\sigma_i) = \max\{p(\Theta^{L_i}), p(\Theta^{L_j})\} \quad (13)$$

2.4 Trend Prediction

Finally, the optimal trends resulting from the optimization process are combined to obtain the trend prediction corresponding to the template $X(t)$, as (14).

$$\hat{Y}(t) = a^\sigma \bar{Z}(t) + \sum d^\sigma \bar{Z}(t) \quad t = N+1, \dots, N+P \quad (14)$$

In the previous equation the subscript σ denotes the optimal trends identified by the optimization process.

3. RESULTS

3.1 Introduction

a. Tachycardia

The precise definition of tachycardia is difficult to find and, therefore, the threshold for this condition should be considered flexible, based on the level and profile of the patient’s cardiovascular risk. For example, a heart rate (HR) value may be considered as unacceptably high for patients in high risk state, but still acceptable for low risk patients. Nevertheless, it is the responsibility of the clinician to decide if the individual suffers from tachycardia not only based on heart rate measurements, but also on the patient’s history. To show the feasibility of the approach, in this work a threshold of 100 bpm was considered as the limit value for tachycardia.

b. TEN-HMS dataset

The Trans-European Network Homecare Monitoring Study (TEN-HMS) was designed to assess whether home based telemonitoring could reduce morbidity and mortality in patients with heart failure, compared with usual care or regular telephone contact (a total of 426 patients). Particularly, home telemonitoring consisted of twice-daily self-measuring of weight, blood pressure and heart rate, during the period of two years. For the present work, a subsection containing data from 83 patients was made available. In terms of strategy validation, only HR signals were employed. Furthermore, only patients for whom there were HR measurements in, at least, 150 days (5 months) were selected for this purpose, resulting in a total of 51 patients, with a heart rate value per day (sampling rate of 1 day).

c. Time horizon

The selection of a forecasting horizon is an essential step when predicting time-series data. From a clinical perspective, this period should be longer enough to allow a timely intervention in order to avoid an undesirable outcome. From the prediction perspective, the period should be as shorter as possible, since trends in data may not persist for too long. Taking into account these aspects, a forecast period of approximately one week (eight days, $P=8$) was stipulated. In terms of the length of the template, that is, the past information used in the prediction, the value suggested by the clinical partners was about a month ($N=32$).

d. Experiments

Two groups of experiments were carried out. The first group assesses the capacity of the proposed wavelet multi-resolution scheme (WMM) in the trend prediction of heart rate signals. Moreover, the performance of this scheme is compared with other typical prediction strategies, namely a linear regression model, the autoregressive integral moving average model - ARIMA, and a non-linear regression models, the generalized regression neural network – GRNN. Other prediction method (AVP) simply considers the average value of predictive signals $Y_m(t)$, as an estimation for the prediction of $\hat{Y}(t)$.

The second set of experiments selects patients with HR values in a critical range (around the threshold of tachycardia), and uses the previously estimated trend to determine the risk of tachycardia. Specifically, the goal is to evaluate whether during the following week the HR signal of a given patient evolves towards tachycardia values or, on the contrary, is maintaining or decreasing to normal values.

3.2. Experiment 1: trend prediction comparison

a. Parameters

With respect to the ARIMA model, the examination of the autocorrelation and partial autocorrelation functions of the differenced series, was used in the estimation the order of the model $ARIMA(n_a, d, n_c)$. The parameters n_a , d and n_c identify, respectively, the number of autoregressive terms, the degree of differencing and the number of lagged forecast

errors in the prediction equation. As result, the ARIMA structure was $ARIMA(2,1,2)$. The estimation of parameters was carried out with the *armax(.)* Matlab command.

Regarding GRNN structure, it can be seen as normalized radial basis function networks, where there is a hidden unit centred at every training case. These units are called "kernels" and, usually, are probability density functions, such as Gaussian functions. The weights from the hidden to output layer are just the target values, so the output is simply a weighted average of the target values of the training cases, close to the given input case. As a consequence, the only parameters to be learned are the widths of the units. In the experiments using the heart rate signals, the width of the kernels was experimentally determined as $\lambda = 0.2$. The *newgrnn(.)* Matlab command was used to implement this neural model. Moreover, a different neural network had to be trained for each template.

With respect to AVP, the average prediction $\bar{Y}(t)$, of the identified patterns was computed using an weighted average, taking into account the similarity measure evaluated for each pattern.

The last approach (WMM) put into practice the proposed wavelet strategy, considering the following parameters:

Similarity analysis: $N=32$, $P=8$, where N and P denote, respectively, the time intervals before and after the current time instant; $M=5$, number of patterns retrieved from the historic dataset; $L=5$, wavelet decomposition level.

Selection of the optimal trends: Number of decompositions considered in the optimal trend selection $l = 3,4,5,6$ (the details are the levels $l = 3,4,5$; the approximation is the level $l = 6$); the first two levels of detail ($l = 1,2$) were neglected; conjunction and aggregation operators were, respectively, the *maximum(.)* and the *product(.)* operators.

b. Prediction metrics

The accuracy of the forecasting methods was determined in terms of four performance metrics: *i)* the proposed similarity measure based on the wavelet decomposition+KLT (SWK), (15); *ii)* the Pearson's correlation coefficient (CORR), (16); *iii)* the normalised root mean squared error (NRMSE), (17) and *iv)* the mean absolute percentage error (MAPE).

$$SWK = S(Y(t), \hat{Y}(t)) \quad t = N + 1, \dots, N + P \quad (15)$$

$$CORR = \frac{\sum_{t=N+1}^{N+P} (Y(t) - \bar{Y}) (\hat{Y}(t) - \bar{\hat{Y}})}{\sqrt{\sum_{t=N+1}^{N+P} (Y(t) - \bar{Y})^2} \sqrt{\sum_{t=N+1}^{N+P} (\hat{Y}(t) - \bar{\hat{Y}})^2}} \quad (16)$$

$$NRMSE = \frac{1}{P} \frac{\sum_{t=N+1}^{N+P} (Y(t) - \hat{Y}(t))^2}{\sum_{t=N+1}^{N+P} (Y(t) - \bar{Y})^2} \quad (17)$$

$$MAPE = \frac{1}{P} \sum_{t=N+1}^{N+P} \left| \frac{Y(t) - \hat{Y}(t)}{Y(t)} \right| \quad (18)$$

In the previous equations, $Y(t)$ is the actual HR value, $\hat{Y}(t)$ is the forecasted HR, \bar{Y} and $\hat{\bar{Y}}$ are, respectively, the means of the actual and the estimated signals. The metrics NRMSE and MAPE were transformed to $NRMSE = \exp(-\kappa_N NRMSE)$ and $MAPE = \exp(-\kappa_M MAPE)$, in order to guarantee that their values are in the range $[0,1]$. The parameters κ_N and κ_M are constants, respectively, $\kappa_N = 0.25$ and $\kappa_M = 10$.

c. Comparison of prediction methods

Among the available parametric and nonparametric tests, the Friedman test is a nonparametric one that enables to perform multiple comparisons in experimental studies. This test (Friedman, 1937), (Friedman, 1940) is equivalent to ANOVA and is particularly adequate for machine learning studies when the assumptions (independency, normality and homoscedasticity) do not hold or are difficult to verify for a parametric test (Garcia et al, 2010).

The objective of the Friedman test is to determine if it is possible to conclude, from a set of results, that there is a difference among the several methods. Basically, the Friedman test then compares the average ranks R_j of each method, to decide about the null hypothesis, which states that “*Ho: all the algorithms behave similarly and thus their ranks R_j should be equal*”. The Friedman statistics, is distributed according to χ^2_F , with $k-1$ degrees of freedom. From the computation of the corresponding *p-value*, the null hypothesis can be or not rejected at a given level of significance.

The Nemenyi test enables a pairwise comparison of the methods, based on the average ranks computed in the Friedman test. Basically, by means of the Nemenyi test, two methods can be significantly different at a several levels, namely $\alpha = 1\%$, $\alpha = 5\%$, or $\alpha = 10\%$, if their average ranks differ at least the critical value. In this case ($k=4$) the thresholds for the critical values are, respectively, $CD_1 = 1.4675$, $CD_5 = 1.2110$ and $CD_{10} = 1.080$.

d. Results

For the comparison of the proposed prediction method (WMM) against the other strategies a total of 300 random experiments were performed. The Figure 3 depicts the box-plot resulting from the comparisons.

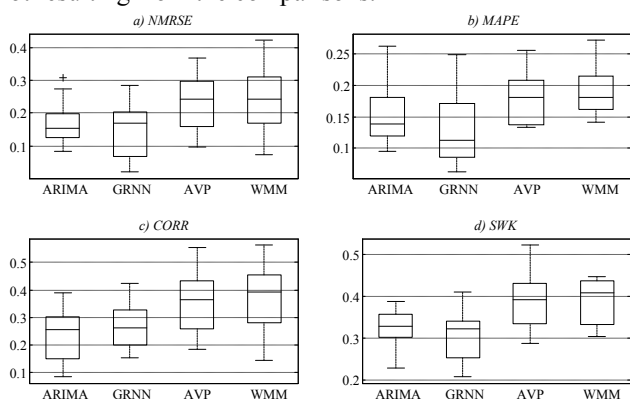


Fig. 3. Comparison of the prediction methods (NRMSE, MAPE, CORR, and SWK metrics) using TEN-HMS dataset.

From the analysis of Figure 3 and, in global terms, it appears that the proposed method is slightly superior to the others. In effect, the wavelet based prediction method (WMM) presents the highest median for all the metrics showing, however, a higher variability for some of these metrics.

The methods ARIMA and GRNN compute the prediction based on an iterative approach: a one-step ahead model is iteratively applied during P times, being the current predictions used by the model in order to obtain the next forecast. The last two methods (AVP and WMM) do not involve the explicit computation of a model, thus, they are, to some extent, similar to a direct approach. This fact can justify why GRNN and ARIMA present poor.

In order to accurately compare the predictive methods, the Friedman test was implemented, considering the four metrics. The Table 1 and Table 2 summarize the average ranks and the respective values of qui-square and *p-value*.

Table 1. Comparison of the prediction methods

a) Average ranks.				
	ARIMA	GRNN	AVP	WMM
NRMSE	2.066	1.40	2.600	3.933
MAPE	1.733	1.333	3.066	3.866
CORR	1.933	1.266	3.200	3.600
SWK	1.800	1.333	3.466	3.400

b) Qui-square and p-value.		
	χ^2_F	<i>p-value</i>
NRMSE	31.16	7e-7
MAPE	37.24	4e-8
CORR	31.88	5e-7
SWK	32.36	4e-7

From the analysis of Table 1b) the null hypothesis has to be rejected for all the metrics. Moreover, from the previous comparison using individual metrics using the average of ranks (Table 1a), it can be concluded that the proposed method is globally superior to the others, except when the SWK metric was used.

In a second phase the Nemenyi test was used to compare the methods based on the computed average ranks. Table 2 presents this comparison, for the particular case of the CORR metric (Figure 3, left, bottom).

Table 2. Nemenyi test (CORR metric).

#methods (<i>k</i>)	GRNN	AVP	WMM
ARIMA	-0.666	1.266 **	1.666 ***
GRNN		1.933 ***	2.333 ***
AVP			0.400

*, **, *** : at a significance level of, respectively, 10%, 5% and 1%

From the table, it can be concluded that the proposed WMM method outperforms ARIMA and GRNN at the levels of 1%. In turn, at a level of 1% and 5% the SVR outperforms, respectively, the GRNN and ARIMA methods. One the other hand, the methods AVP and WMM presents similar results and therefore can not be considered different.

3.3 Experiment 2: tachycardia risk assessment

A set of experiments was carried out particularly applied to patients whose heart rate values were in a critical range (around the threshold of tachycardia). The main goal was to determine whether during the following week the heart rate signal of a patient would evolve towards tachycardia values or, on the contrary, would be maintained or decrease to normal values. Figure 4 illustrates this idea.

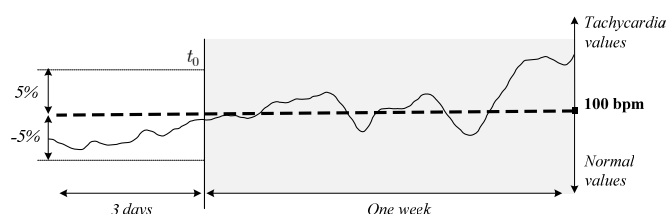


Fig. 4. Assessment of tachycardia risk.

The procedure started by identifying patients that had recently shown heart rate values in a critical range, more specifically, that had presented HR values in the range $[-5\%, +5\%]$ of the limit value of 100 bpm during three consecutive days. Then, for those patients, the HR values of the following week were predicted using the methodology previously described. According to the percentage of values that were above the limit threshold (100 bpm), the risk of the patient was assessed: if the percentage was higher than 75%, the patient was considered to be at risk of developing an tachycardia condition; in the other case (less than 75%), the patient was considered to have no tachycardia risk.

The effectiveness of the proposed strategy was tested by selecting, from a set of 600 random templates, the ones that verified the referred requirement (to be in the critical range). In effect, 58 verified this condition: in 26 cases the patient presented risk of developing a tachycardia condition, and in 32 cases the patient revealed no risk.

Table 3 shows the discrimination capability of the method.

Table 3. Confusion matrix.

	Actual class		
	No risk	In risk	
Predicted class	No risk	28	10
	In risk	4	16

To quantify the validity of the method, the sensitivity (SE) and specificity (SP) were determined, resulting in a SE of 62% and a SP of 87%.

Although it was not possible to compare these results with other works, considering that the prediction involved fully random templates, the obtained SE and SP values were very satisfactory. In effect, these metrics demonstrate the potential of the trend prediction strategy.

4. CONCLUSIONS

This paper presented a strategy based on wavelet decomposition for the forecast of biosignals, which goal was not to perform accurate predictions but to obtain a reasonable estimation of signals' future evolution trend. The capability of the proposed methodology was, in a first phase compared with other common prediction mechanisms. Then, using the predicted values, the scheme was tested in the assessment of the tachycardia risk in patients whose heart rate values were in a critical range (around the threshold of tachycardia). For the effect, real data collected by the tele-monitoring study TEN-HMS, were used. The obtained values of sensitivity and specificity suggested the capacity of our strategy.

Acknowledgements

This work was partially financed by iCIS (CENTRO-07-ST24-FEDER-002003) and CardioRisk (PTDC /EEL-SII/2002/2012).

REFERENCES

- Agrawal, R. et al (1993), *Efficient similarity search in sequence databases*, Proc. Int. Conf. on Foundations of Data Organizations and Algorithms, 69–84.
- Cleland et al, (2005), *Noninvasive home telemonitoring for patients with heart failure at high risk of recurrent admission and death*, J Am Coll Cardiol, 45:1654-1664.
- Fryzlewicz, P. et al (2003), *Forecasting nonstationary time series by wavelet process modeling*, ISM, 55, 4, 737-764.
- Habetha, J. (2006), *MyHeart - A new approach for remote monitoring and management of cardiovascular diseases*, IEEE EMBS, New York, USA.
- Haykin, S, (2008), *Neural networks and learning machines*, Third Edition. Prentice Hall
- Kriegel et al, (2003), *Density-based clustering*, WIREs Data Mining and Knowledge Discovery, 1 (3): 231–2.
- Park, S. et al (2000), *Efficient searches for similar subsequences of different lengths in sequence databases*, Proc. Int. Conference of Data Engineering, 23-32.
- Renaud et al (2003), *Prediction based on a multiscale decomposition*, Int- Journal of Wavelets, Multiresolution and Information Processing, v.1, n.2, 217-232.
- Rocha, T. et al, (2012) *An Efficient Strategy for Evaluating Similarity between Time Series based on Wavelet/Karhunen-Loève Transforms*, IEEE EMBS, San Diego, 6216- 6219.
- Saeed, M. and Mark, R. (2006), *A Novel Method for the efficient retrieval of similar multi-parameter physiologic time series using wavelet-based symbolic representations*, AMIA symp.
- SIGN (2007), *Scottish Intercollegiate Guidelines Network Cardiac arrhythmias in coronary heart disease, national clinical guideline* (www.sign.ac.uk/pdf/sign94.pdf).
- Yang, K. and Shahabi (2004), C., *A PCA-based Similarity Measure for Multivariate Time Series*, Proc. of the 2nd ACM international workshop on Multimedia databases.