J. Health Inform. *2017 Julho-Setembro; 9(3): 73-80* **Artigo Original**

Medical time series classification using global and local feature extraction strategies

Classificação de séries temporais médicas por meio da extração de características globais e locais

Classificación de series temporales médicas mediante la extracción de características globales y locales

André Gustavo Maletzke[1], Carlos Andres Ferrero[2], Chris Mayara Tibes[3], Everton Alvares Cherman[4], Willian Zalewski[5]

Keywords: Artifical Intelligence; Electrocardiography; Electroencephalography

ABSTRACT

Objective: Present a method to improve the accuracy of the time series classification task, as well as to enable the interpretation of its generated model. **Method:** Features were extracted from time series combining two strategies: the global strategy, which uses statistical and complexity descriptors; and the local strategy, which uses the motif representation. In the next step, the data was submitted to three different learning algorithms in order to create classification models. The performances of the models were evaluated in terms of mean error rate using five medical datasets. **Results:** fFr all datasets, the best classification accuracy was obtained combining both local and global strategies. The approach improved the performance of the J48 algorithm, which generates a more interpretative model. The comparison among 1-NN, MLP, and J48 shows no significant statistically difference. **Conclusion:** The method aims at an enhanced descriptive power for time series data and increasing the performance of the models.

Descritores: Inteligência Artificial; Eletrocardiografia; Eletroencefalografia

RESUMO

Objetivo: Apresentar um método para melhorar a precisão de classificação de séries temporais, bem como a interpretabilidade dos modelos. **Método:** Foram extraídas características de séries temporais mediante duas estratégias: a estratégia global, na qual foram utilizados descritores estatísticos e de complexidade; e a estratégia local, que consistiu na identificação de *motifs*. Após, foram utilizados três algoritmos para a indução de modelos preditivos. A eficácia dos modelos foi avaliada mediante a taxa de erro médio usando cinco bases de dados médicas. **Resultados:** Para todas as bases a menor taxa de erro médio foi utilizando as estratégias de maneira combinada. O método proposto melhorou a eficácia do algoritmo J48 e a interpretabilidade dos modelos. Na comparação entre os algoritmos 1-NN, MLP e J48 não foi observada diferença estatística. **Conclusão:** O método contribuiu para a construção de modelos simbólicos interpretáveis tão eficientes quanto os não simbólicos para a classificação de séries temporais médicas.

Descriptores: Inteligencia Artificial; Electrocardiografía; Electroencefalografía

RESUMEN

Objetivo: Presentar un método para aumentar la precisión de clasificación de series temporales y la interpretabilidad de los modelos. **Método:** Fueron determinadas características de las series mediante dos estrategias: la estrategia global: en la cual fueron utilizados descriptores estadísticos y de complejidad; y la estratégia local, mediante la identificación de motifs. Luego, fueron utilizados tres algoritmos para la inducción de modelos de predicción. Para evaluar los models fue utilizada la tasa de error promedio usando cinco base de datos médicas. **Resultados:** Para todas las bases la menor tasa de error fue utilizando las estategias de forma conjunta. El método ha mejorado la precisión del algoritmo J48 y la interpretabilidad de los modelos. No fue observada diferencia estadística entre los algoritmos 1-NN, MLP y J48. **Conclusión:** El método forneció modelos simbólicos interpretábles tan eficientes cuanto los demás métodos utilizados para la clasificación de series temporales médicas.

[1] *Doutorando em Ciências de Computação e Matemática Computacional pelo Instituto de Ciências Matemáticas e de Computação da Universidade de São Paulo – ICMC/USP, São Carlos (SP), Brasil.*

[2] *Doutorando pelo Programa de Ciência da Computação, Departamento de Informática e Estatística, Universidade Federal de Santa Catarina – UFSC, Florianópolis (SC), Brasil.*

[3] *Doutoranda em Enfermagem Fundamental pela Escola de Enfermagem de Ribeirão Preto da Universidade de São Paulo – EERP/USP, Ribeirão Preto (SP), Brasil.*

[4] *Doutor pelo Instituto de Ciências Matemáticas e de Computação da Universidade de São Paulo – ICMC/USP, São Carlos (SP), Brasil.*

[5] *Professor do Departamento de Informática, Universidade Tecnológica Federal do Paraná – UTFPR, Pato Branco (PR), Brasil.*

Autor Correspondente: **André Gustavo Maletzke**
e-mail: **andregustavom@gmail.com**

Artigo recebido: **19/02/2016**
Aprovado: **25/05/2016**

INTRODUCTION

Large amounts of information that reflect the patient's clinical condition are generated and stored daily in medical applications. Medical data such as Electrocardiogram (ECG) and Electroencephalogram (EEG) exams are daily collected around the world. These exams consist in recording the electrical activity of the heart and brain, respectively. The ECG and EEG exams have paramount importance to experts because they can contribute to early diagnosis and then provide an effective treatment for vascular diseases and epileptic disturbances[1-3]. Unfortunately, only a small portion of this information will receive any further analysis to identify patterns and build models to support the decision-making process. Moreover, due to the large volume of ECG and EEG data, analyses unaided by computational tools are extremely complex and can be incomplete due to human limitations when dealing with large amounts of data. To overcome these limitations, one can apply the data mining process, supported by Machine Learning (ML) techniques. This strategy has been promising when applied to many other problems, and so we explore this strategy in medical data, specifically for ECG and EEG exams. However, there are some limitations when traditional machine learning methods are applied directly to the data of ECGs and EEGs. These data represent variations of electric pulses over time and therefore there is a temporal relationship between each observed value, which will be lost if traditional ML techniques are applied. Data that have a temporal relationship between each observed value are called time series or sequential data, and the relationship between subsequent observations should be considered into the process of elaborating a model. A great challenge in machine learning is to integrate temporal and sequential data into the data mining process. A widely used strategy for building attributes from ECG and EEG time series consists in determining features that describe a global behavior[2-4].

In this paper, we present a method for building attributes in medical temporal databases, combining global and local features derived from different visions. Our method improves the descriptive power of the time series data, focusing on increasing the performance of the supervised learning task (classification). The method is an improvement of a previous approach based on statistical feature extraction and motifs discovery[5]. For this, we include new strategies to extract characteristics of time series and provide a guide to mining medical time series databases with a focus on building accurate symbolic models.

The remaining of this paper is organized as follows: section 2 describes our method and the experimental evaluation performed. Section 3 presents the results and the discussion of each experiment, and Section 4, subsequently, presents the conclusions.

METHOD

The method aims to explore three different approaches that provide different views of a time series, so that after its application to building attributes, we achieve a structured representation of a time series or a set of time series, which we call the attribute-value representation. The method is described below.

Statistical Metrics

In this strategy, we construct attributes through measures of descriptive statistics in order to capture the overall behavior of the time series. Statistical measures, such as the average, the maximum, and the minimum, are descriptive information that have good interpretability and consequently can contribute to maintaining the readability of the models built. Moreover, determining descriptive statistical measures is a simple task with low computational cost.

After the attribute construction using this strategy, we obtain an attribute-value table, in which the attributes represent the calculated metrics. Figure 1 (a) shows a schematic representation of this process, where SM_1, SM_2, ..., SM_k are the different statistical metrics calculated from the time series T_1 and T_2.

Statistical measure extraction enables the traditional machine learning algorithms to be applied in an adequate way, because the temporal relationship among the

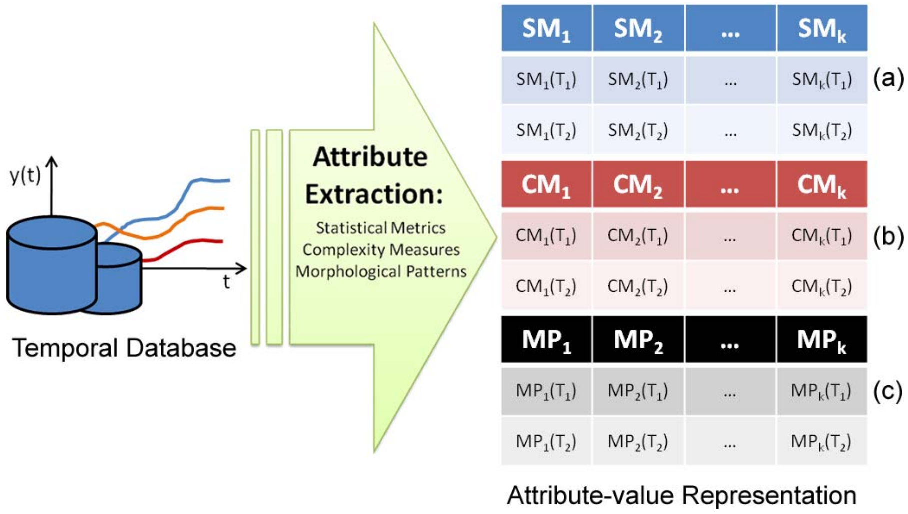

Figure 1 - Schematic representation of the proposed method.

observations no longer exists. Although the determination of statistical measures has been widely used in the literature in different domains, this approach cannot provide satisfactory results in situations where the statistical summarization is limited by the type and amount of data available, i.e., most of ECG our EEG signals have nonlinear or non-stationary behavior. Therefore the following strategy aims to provide a summarization of the time series with an alternative view to the statistical description.

Complexity Measures

Complexity measures are intended to provide a degree of disorder related to the time series. This information can differentiate or congregate time series data, based on the degree of disorder of the observations, providing an alternative to the statistical measures. There are some desirable characteristics in a complexity measure, such as low space and time complexities[6]. There are several complexity measures that may be determined from time series and subsequently represented in an attribute-value table; examples of these measures are the complexity estimate, the entropy, and the fractal dimension. Figure 1 (b) shows the time series representation using complexity measures where CM_1, CM_2, ..., CM_k are different complexity metrics extracted from the time series T_1 and T_2.

However, global characteristics have some drawbacks when used to describe time series. One of them is the low capacity to describe local aspects of the problem at hand. Therefore, our method explores the potential of a third approach to represent time series, which aims to identify some of the local features in the time series. This approach is presented in the following section.

Morphological Patterns

Discovering morphological patterns, called motifs, in time series is an important data mining task, to which increasing attention has been paid[7-9]. Figure 2 show an example of motif discovery.

Motif definition consists in identifying similar subsequences in a time series in significantly different positions[9]. Motif identification can be used to overcome the aforementioned difficulties related to the global strategies. In this context, we want to promote the motif transformation in an efficient way to construct a consistent attribute-value representation. This process is illustrated in Figure 1 (c), where MP_1, MP_2, ..., MP_k representing different motifs found in time series T_1 and T_2, whose values in the attribute-value table represent the frequency, the presence in binary value, or the position in the time series. The motif identification problem in time series requires a high computational effort: it has quadratic complexity in the size of the time series. In this paper, a probabilistic approach was used[9].

As mentioned, different strategies can be used to find motifs in time series data, but we have selected a simple probabilistic strategy that requires only a few parameters. In this paper, the focus is on evaluating the effectiveness of the joint use of statistical metrics, complexity measures, and motifs, in constructing a structured representation. Figure 1 illustrates the goal of the proposed method. We present several experiments that evaluate its real contribution.

Experiments

In this work, we report several experiments on different medical time series related to ECG and EEG exams. The intended contribution of our method in medical time series problems focuses on symbolic classification models, which we argue can give real support to the decision making process. For that, we have selected the J48 machine learning algorithm, which is a traditional and often employed algorithm for inducing decision trees.

Although we concentrate on symbolic models, we would also like to analyze the behavior of the decision tree models compared to two of the most used machine learning algorithms: 1-Nearest Neighbor (1-NN) and Multilayer Perceptron (MLP). Finally, in order to evaluate the performance of the proposed method, we conducted three different analyses:

a) Contribution of each strategy: we evaluated the real contribution of each featured extraction strategy separately – statistical metrics, complexity metrics, morphological patterns and all of them combined for each machine learning algorithm;

b) Contribution of the method in terms of symbolic models: we also conducted a statistical analysis considering only the symbolic models induced by the J48 algorithm for each dataset;

c) Comparison among machine learning algorithms using our method: in this analysis, we performed a statistical comparison between the results obtained by the three different machine learning algorithms.

The experimental evaluation was conducted using the

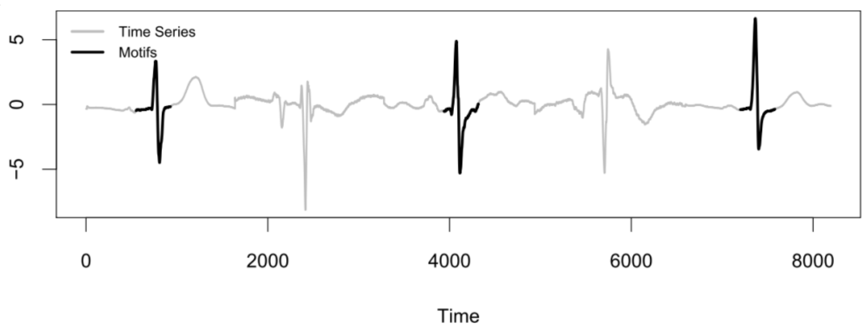

Figure 2 - An example of motifs occurring in a ECG dataset.

10-fold cross validation sampling method and the predictive models were induced using the WEKA tool by rWeka library for R, applying the default parameter values. In the first analysis, we also applied the Feature Selection (FS) task using the Correlation Feature Selection (CFS) algorithm[10]. The statistical evaluation was conducted in the R environment and the ANOVA statistical test with Tukey post-hoc was performed at the 95% confidence level. The experiments were conducted on five datasets widely known in the community. In the next subsection, we describe the used datasets.

Datasets

The experimental evaluation was performed on five medical time series databases obtained from two dataset repositories: the UCR Time Series Classification/ Clustering[11], and the EEG Database from the University of Bonn[3]. Table 1 presents a summary description of the selected datasets.

The datasets description is presented below.

· ECG200: time series from different subjects related to supraventricular and non-supraventricular tachycardia;

· ECGFiveDays: data of the same subject recorded at intervals of five days;

· TwoLeadECG: time series from MIT:BIH Long-Term ECG Database about two-lead ECG records;

· EEG: electroencephalogram data for five healthy patients and five epileptic patients were obtained from the EEG Database at the University of Bonn. The database has five datasets, A, B, C, D, and E. Sets A and B represents data segments about healthy patients. The data segments in C, D, and E come from epileptic patients. We have joined the subsets into two classes representing healthy patients (A and B) and epileptic patients (C, D, and E);

· CinC_ECG_torso: time series obtained from ECG data from multiple torso-surface sites.

Experimental Setup

Our method was applied using the follow parameters for each strategy:

Statistical Metrics (SM): to construct the attributes using statistical measures, we used the mean, maximum, minimum, and coefficient of variation calculated over the whole time series;

Complexity Measures (CM): to extract the degree of disorder of the time series the following methods were used: box counting[12], Katz[13], complexity estimate[6] and empirical entropy[14].

Morphological Patterns (MP): the patterns were extracted in different sizes, considering the intervals of 1% to 25% of the whole time series with increments of 1%. The Euclidean distance with z normalization was used to calculate the similarity between two subsequences that are possible motifs. However, only subsequences that differ at least by 5% will be considered a motif. The alphabet size used is composed of six characters, no dimensionality reduction method was applied, and only 20% of the search space was explored.

RESULTS AND DISCUSSION

In this section, we present the results organized by the objective of each analysis described in the previous section.

Contribution of the method strategies

Initially, the contribution provided by each strategy applied in the proposed method was evaluated separately. We start by presenting, in Figure 3, the mean errors classification rate with its corresponding standard deviation for each dataset, first when only one strategy is used to induce the classifiers, and then when we combine all of them. For the 1-NN algorithm, our method provides better results in 60% of the datasets (Figure 3 (a)) and 80% for the MLP algorithm (Figure 3 (b)). The performance is most evident for symbolic algorithm (J48), which has a lower mean error rate for all datasets (100%) (Figure 3 (c)). These results accord with our objective to provide a generic method to construct symbolic and accurate models for medical time series datasets. We can observe, especially for 1-NN, datasets ECG200 and ECGFiveDays, and MLP, dataset ECG200, that the joint use of the strategy produced lower classification accuracy than when only one strategy was used. This result was expected due to the fact that our method may produce some morphological attributes with low discrimination capacity. This fact is more critical with the 1-NN algorithm, because it imputes the same weight to all attributes; the same happens with the MLP algorithm even considering that this algorithm can assign greater importance to certain attributes during the neural network training process.

In the other hand, for the J48 algorithm, all the results were better when all the strategies were used in combination, which can be explained by the fact that J48

Table 1 - Datasets description.

Dataset	# Ex.	Time Series Length	Class	%Class
ECG200	200	96	1	65.5%
			2	33.5%
ECGFiveDays	884	136	1	50.0%
			2	50.0%
TwoLeadECG	1162	82	1	50.0%
			2	50.0%
EEG	500	4097	1	40.0%
			2	60.0%
CinC_ECG_torso	1420	1639	1	50.0%
			2	50.0%

promotes an embedded attribute selection. Thus, the CFS algorithm was applied as a previous step for building the classifiers. These results are included in the last bar of Figure 3. Using CFS, an important improvement in the classification error rate was obtained for the 1-NN algorithm in ECG200, ECGFiveDays, and TwoLeadECG datasets. For the MLP algorithm, the feature selection task showed no contribution in most datasets. The feature selection task, considering the J48 algorithm, has promoted a slight improvement except with EEG and CinC_ECG_torso. Table 2 presents, for each dataset, the best combination of feature extraction strategy and supervised learning algorithm, *i.e.*, the combination that achieved the best mean classification error.

It can be observed that, independently of the classifier used, all the best results were obtained by using the combination of both local and global feature extraction strategies.

Contribution of the method in terms of symbolic models

In order to get an enhanced analysis, we performed a statistical test to detect significant differences between the strategies when the J48 algorithm is used. Figure 4 shows the statistical comparison between the symbolic classifiers constructed using only statistical measures, only complexity measures, only motifs, and when all are combined. This analysis allows us to observe statistical differences when all strategies are employed in a combined way compared with each strategy separately. The pairwise comparisons between strategies, CM×MP, CM×SM, and MP×SM, have the following p-values: 0.9996, 0.9884, and 0.9965, respectively. However, the comparisons between each strategy separately, CM, MP, and SM, with all strategies combined (SM+CM+MP) have the following p-values: 0.0081, 0.0112 and 0.0200, respectively, indicating a significant statistical difference. Another aspect that is as important as having a good average performance on several datasets, and which has motived us, consists in providing a simple model with low complexity. For this, the models' complexity was analyzed using as a metric the number of leafs, which defines the amount of rules extracted by each model, i.e, this metric represents the complexity of the knowledge that will be shown to the experts. Figure 5 shows the complexities for each dataset without feature selection. At this point, our method can generate a simple model at least in four of five datasets,

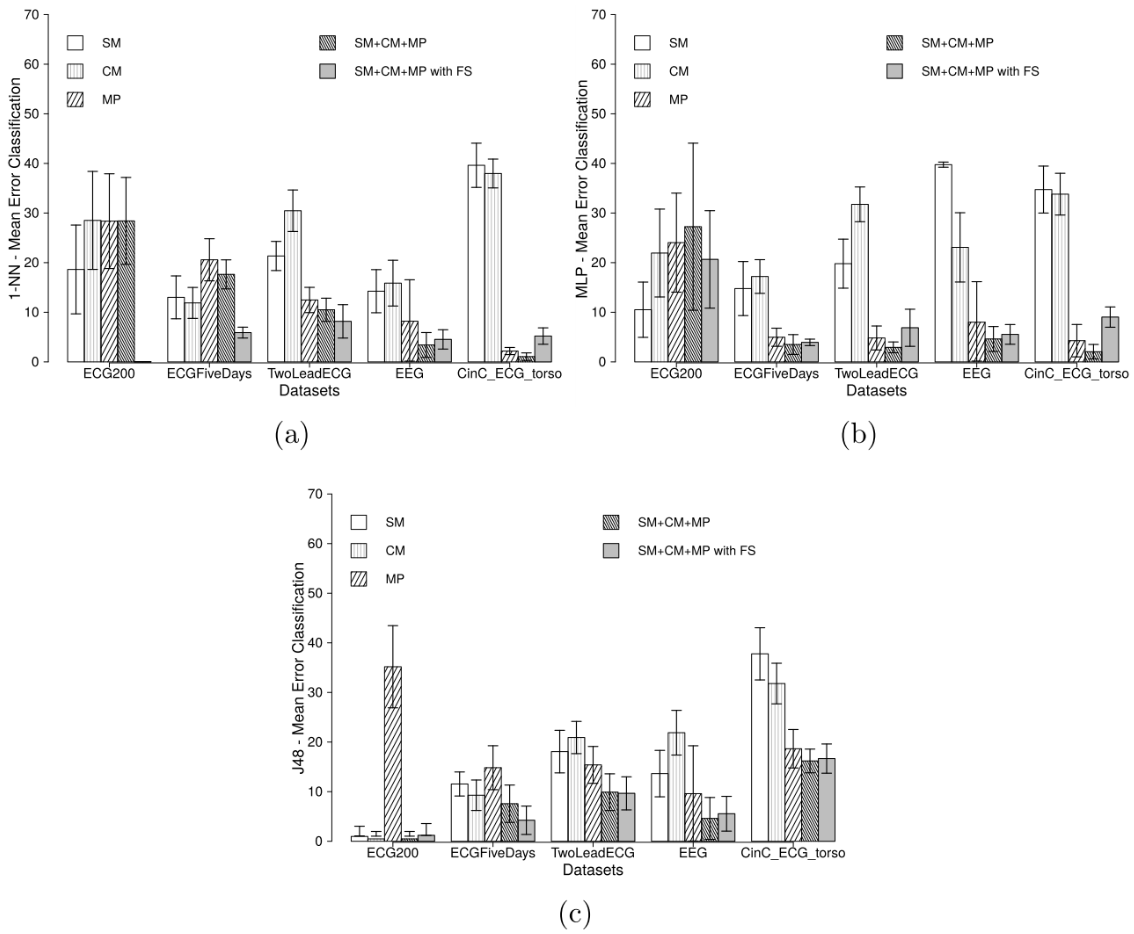

Figure 3 - Error rates when we apply each strategy separately considering (a) 1-NN, (b) MLP and (c) J48, estimated using 10-fold cross-validation.

Table 2 - The best combination of feature extraction strategy for each dataset.

Dataset	Classifier	Feature Extraction
ECG200	1-NN	SM+CM+MP with FS
ECGFiveDays	MLP	SM+CM+MP
TwoLeadECG	MLP	SM+CM+MP
EEG	1-NN	SM+CM+MP
CinC_ECG_torso	1-NN	SM+CM+MP

and only in one dataset the number of rules was higher. The complexity results are related with the mean error observed in Figure 3, in which the mean error is superior for the CinC_ECG_torso dataset.

Indeed, our method can produce a low number of rules for most datasets, and thus providing a remarkable approach for mining medical time series with a focus on symbolic models.

Comparison among machine learning algorithms using our method

Table 3 summarizes the efficiency of the method for each algorithm in terms of the mean error classification rate and its standard deviation. However, in order to illustrate its potential, especially on the J48 algorithm, we performed a statistical test among the algorithms using the proposed method. The statistical results are shown in Figure 6. In the statistical comparison between 1-NN and MLP, we observe a p-value of 0.7034, and between 1-NN and J48 the p-value was 0.1867. For the comparasion between MLP and J48, the p-value was 0.6058. Therefore, we observe that there are no significant differences between the three algorithms. Our method has been an appropriate alternative for time series mining using symbolic models.

All the statistical analyses are according with the central assumption of this research that is: combining global measures (statistics and complexity) with local features (motifs) can provide more accurate symbolic models. Hence, the results can help to elucidate the importance of our method in enabling the use of the traditional supervised learning algorithms, especially the symbolic ones, in a simple way.

These results are encouraging, even in comparison to related work that has explored the same datasets, applying distinct approaches[4-8]. Others strategies work well reporting a very low error rates for EEG dataset, even when only global features are used to induce non-symbolic models[4].

Another approach extracts subsequences of time series that could be representative of a class, named shapelets, to induce symbolic models[7]. The results are interesting for the ECGFiveDays dataset (1.05% error rate), but for the ECG200 and TwoLeadECG datasets the results were weak (79.00% and 13.61%, respectively) and the CinC_ECG_torso dataset was not explored by the authors. Shapelets were also used to extract the best shapelets and symbolic models were built on many datasets including ECGFiveDays and TwoLeadECG with error rates 3.38% and 14.75%., respectively[7]. The experimental evaluation of the mentioned works was performed based on the separation of the dataset into training and testing subsets[11], except for the EEG dataset. Thus, the comparison of the results presented in this paper with the above ones should be made carefully. Additionally, results show strong evidences that the use of both global and local strategies combined with symbolic supervised algorithms allow us to generate interpretable models with competitive performance compared to non-symbolic algorithms, at least for those domains represented by the evaluated medical datasets.

It is worth mentioning that the ease to interpreting a generated model also depends on the ease to interpreting the concepts of the used attributes. Overcoming this challenge, symbolic models enable human interpretation

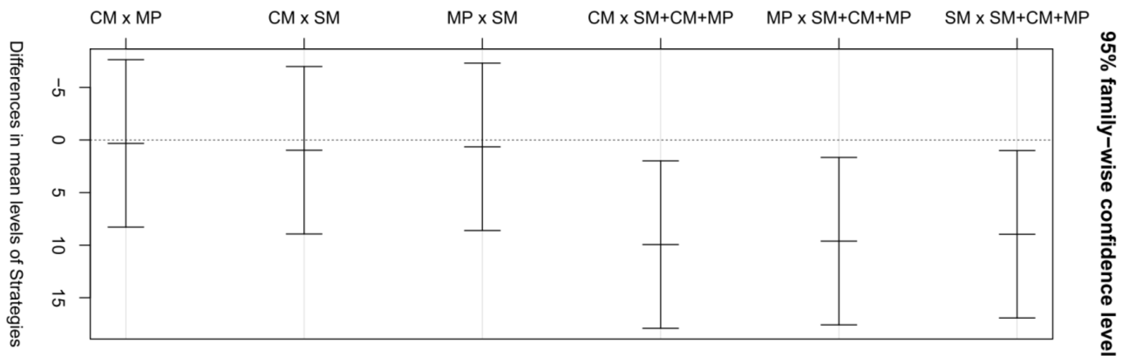

Figure 4 - Tukey post-hoc test comparing each strategy of the proposed method using the J48 algorithm.

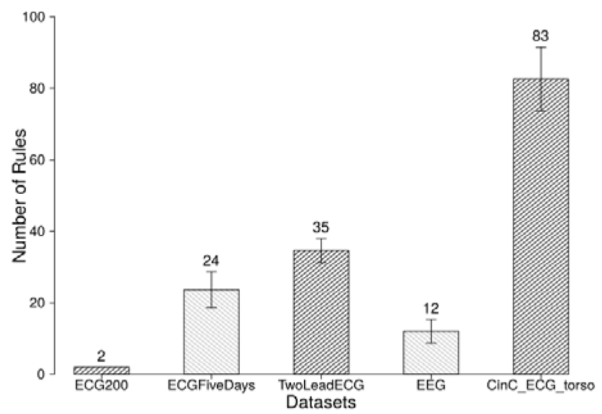

Figure 5 - The mean complexity of the models in terms number of rules and the respective standard deviation, estimated using 10-fold cross-validation.

Table 3 - The classification results in terms of mean error classification with the corresponding standard deviation, estimated using 10-folds cross-validation.

Dataset	SM+CM+MP		
	1-NN	J48	MLP
ECG200	28.42 (8.76)	0.48 (1.51)	27.25 (16.84)
ECGFiveDays	17.64 (2.95)	7.57 (3.76)	3.50 (2.00)
TwoLeadECG	10.50 (2.34)	9.90 (3.69)	2.93 (1.09)
EEG	3.41 (2.51)	4.61 (4.22)	4.62 (2.51)
CinC_ECG_torso	1.06 (0.76)	16.19 (2.39)	2.05 (1.46)

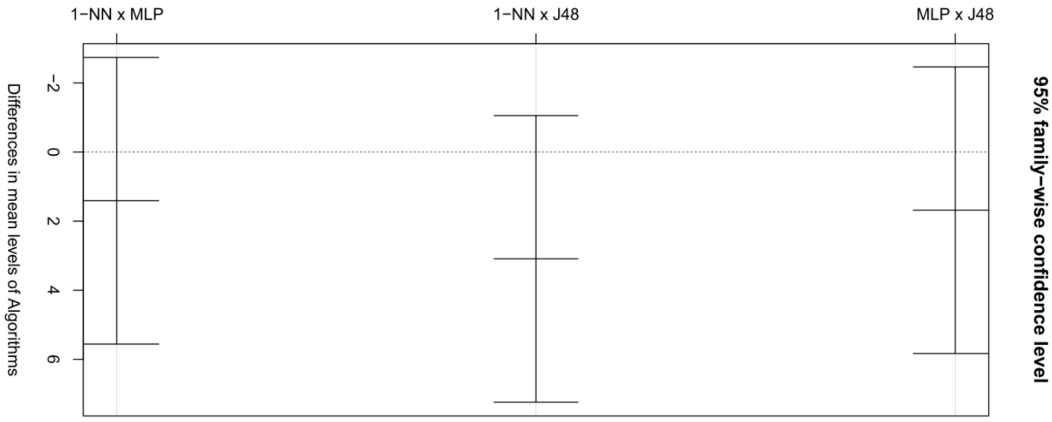

Figure 6 - Tukey post-hoc test appplied to the 1-NN, J48 and MLP algorithms.

in a easier manner if compared to other paradigms like the neural network.

CONCLUSIONS

In this paper, a generic method to mine time series with a focus on symbolic models has been presented. The proposed approach employs different strategies to enable the use of traditional machine learning algorithms based on the assumption that models' performance can be improved by combining global and local features. Different analyses of medical datasets showed the effectiveness of the method, chiefly for symbolic models. The most significant contribution of this paper consists in presenting a method to mine medical time series that improves the performance of classifiers, mainly for the symbolic ones, which also contributes to generating more interpretative and efficient models. In future research, we intend to explore methods to convert time series to another representation, such as an image, using the concept of a recurrence plot, which could be included as a new visualization of the data, and so contribute to the method.

ACKNOWLEDGEMENTS

This work was supported by the Laboratory of High Performance Computing (LCAD) at the Federal University of Latin-American Integration (UNILA) and by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).

REFERENCES

1. Acharya UR, Fujita H, Sudarshan VK, Sree VS, Eugene LWJ, Ghista DN, et al. An integrated index for detection of sudden cardiac death using discrete wavelet transform and nonlinear features. Knowl Based Syst. 2015; 83:149-58.
2. Passos URC, Matias IO, Passos UR, Perreira MA, Carvalho VM. A based on evolutionary computation tool to support the diagnosis of heart disease. J. Health Inform. 2014;6(4):153-60.
3. Andrzejak R, Lehnertz K, Mormann F, Rieke C, David P, Elger C. Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: dependence on recording region and brain state. Phys. Rev. 2001; 64(6):061907.
4. Xie S, Krishnan S. Wavelet-based sparse functional linear model with applications to EEGs seizure detection and epilepsy diagnosis. Med Biolog Engin Comput. 2013; 51(1):49-60.
5. Maletzke AG, Lee HD, Batista GE, Coy CSR, Fagundes JJ, Wu FC. Time series classification with motifs and characteristics. In: Soft Computing for Business Intelligence. Berlin, Heidelberg: Springer Berlin Heidelberg; 2014. p.125-38.
6. Batista G, Keogh E, Tataw O, Souza VMA. CID: an efficient complexity-invariant distance for time series. Data Min. Knowl. Discov. 2014; 28(3):634-69.
7. Yuan JD, Wang ZH, Han M. A discriminative shapelets transformation for time series classification. Int J Pattern Recogn Artif Intell. 2014;28(6):1450014 [28pages].
8. Lines J, Davis LM, Hills J, Bagnall A. A shapelet transform for time series classification. In: Proceedings of the the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD'12, ACM; 2012. Aug. 12 – 16; Beijing, China [Internet] [citado 2015 jan 11]. Disponível em: https://doi.org/10.1145/2339530.2339579
9. Chiu B, Keogh E, Lonardi S. Probabilistic discovery of time series motifs. In: The Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD'03, ACM; 2003. Aug. 24 - 27; Washington, USA [Internet] [citado 2015 jan 11]. Disponível em: https://doi.org/10.1145/956750.956808
10. Hall MA. Correlation-based feature selection for discrete and

numeric class machine learning. In: Proceedings of the Seventeenth International Conference on Machine Learning, ICML'00; 2000. Jun. 29 – Jul. 02; San Francisco, USA [Internet] [citado 2015 jan 11]. Disponível em: http://dl.acm.org/citation.cfm?id=657793

11. Chen Y, Keogh E, Hu B, Begum N, Bagnall A, Mueen A, Batista G. The UCR Time series classification archive. 2016 Fev [citado 2016 fev 11]. Disponível em: http://www.cs.ucr.edu/~eamonn/time_series_data/
12. Peitgen HO, Jrgens H, Saupe D. Fractals for the classroom: introduction to fractals and chaos. New York: Springer-Verlag; 1992.
13. Esteller R, Vachtsevanos G, Echauz J, Litt B. A comparison of waveform fractal dimension algorithms. IEEE Trans. Circuits Syst. I, Fundam. Theory Appl. 2001;48(2):177-83.
14. Chao A, Shen TJ. Nonparametric estimation of shannons index of diversity when there are unseen species in sample. Environ. Ecol. Stat. 2003;10:429-43.