31st Annual International Conference of the IEEE EMBS
Minneapolis, Minnesota, USA, September 2-6, 2009

Comparison of Several Classifiers to Evaluate Endocardial Electrograms Fractionation in Human

V. Křemen, P. Kordík, L. Lhotská

***Abstract*— Complex fractionated atrial electrograms (CFAEs) may represent the electrophysiological substrate for atrial fibrillation (AF). Progress in signal processing algorithms to identify CFAEs sites is crucial for the development of AF ablation strategies. A novel algorithm for automated description of atrial electrograms (A-EGMs) fractionation based on wavelet transform and several statistical pattern recognition methods was proposed and new methodology of A-EGM processing was designed and tested. The algorithms for A-EGM classification were developed using normal density based classifiers, linear and high degree polynomial classifiers, nearest mean scaled classifiers, nonlinear classifiers, neural networks and j48. All classifiers were compared and tested using a representative set of 1.5 s A-EGMs (n = 68) ranked by 3 independent experts 100% coincidentialy into 4 classes of fractionation: 1 – organized atrial activity; 2 – mild; 3 – intermediate; 4 – high degree of fractionation. Feature extraction and well performing classification algorithms tested here showed maximal error of 15% and mean classifiacation error across all implemented classifiers 9%, and the best mean classification error 5.9% (nearest mean classifier), and classification error of highly fractionated A-EGMs of ~ 9%.**

I. INTRODUCTION

Curative ablation is one possible tretatment of atrial fibrillation (AF). Significant progress has been achieved in the field of curative ablation of AF in recent years. While empirical isolation of pulmonary veins is usually an effective strategy in paroxysmal AF, targeting extrapulmonary substrate within left (right) atrium is often necessary in the case of persistent/permanent AF [1]. Both areas with high dominant frequency of atrial electrograms (A-EGMs) [2] and areas with complex fractionated atrial electrograms (CFAEs) [3] were shown to play a role in the maintenance of the arrhythmia. In order to identify those sites, great effort has been made to describe the patterns of activation in AF [4] and to quantify general characteristics of A-EGMs either in time- or frequency-domain ([5], [6]). Recently, two software algorithms for time-domain analysis of CFAEs were implemented in commercially available mapping systems - CARTO (Biosense-Webster) and EnSite NavX system (St. Jude Medical). Both methods require initial setting of specific input parameters making them, at least to some extent, operator-dependent. We used algorithms for automatic classification (pattern recognition), based on description of signal, using features extracted from recorded and preprocessed signals. This approach is based on the idea that there are signal complexes [7] in every A-EGM signal, which are related to electrical activation of electropathological substrate during AF. These signal complexes – fractionated segments of A-EGM (FSs) can be found automatically and then used for several features extraction (degrees of freedom of the signal), which could be used for automatic evaluation of electrogram complexity (or level of fractionation) in next stages. In this paper we focus on evaluation of A-EGM signal complexity of A-EGMs recorded during AF. We use the results of a novel robust method for A-EGM processing [8], based on the wavelet transform signal analysis, several feature extraction followed by classifiers to compare results of their classification.

Manuscript received April 19, 2009. This work was supported by Ministry of Education, Youth and Sport of the Czech Republic with the grant No. MSM6840770012 entitled “Transdisciplinary Research in Biomedical Engineering II” and by the Grant Agency of the Academy of Science of the Czech Republic, grant Automated Knowledge Extraction (KJB201210701).

V. Křemen and L. Lhotská are with Department of Cybernetics, and P. Kordík with Department of Computer Science and Engineering, Czech Technical University in Prague, Czech Republic, xkremen@fel.cvut.cz.

II. METHODS

A. Experimental Dataset of A-EGMs

Atrial bipolar electrograms were collected during left-atrial endocardial mapping using 4-mm irrigated-tip ablation catheter (NaviStar, Biosense-Webster) in 12 patients (9 males, aged 56±8 years) with persistent AF. The A-EGMs acquired before the ablation procedure were band-pass filtered (30-400Hz) and sampled at frequency of 977Hz by CardioLab 7000 (Prucka Inc.). Discontinuous recordings from distal catheter bipole during left-atrial mapping outside the pulmonary veins and their tubular ostia (in order to treat only the signals from sites that are usually targeted during extrapulmonary substrate modification) were exported in digital format and reviewed by independent expert. The fragments with inadequate endocardial contact, relatively high signal-to-noise ratio or artifacts were excluded. Remaining parts of recordings were split into not necessarily contiguous 1500 ms segments with stable signal pattern. The A-EGMs very close to the mitral annulus were discarded to prevent the interference of relatively sharp ventricular signals with atrial signal analysis. This yielded approximately 250 segments with high-quality A-EGMs. This set of A-EGMs was further scrutinized. Finally, selection of 113 such segments represented wide spectrum of A-EGMs including those very organized, extremely fractionated, and all intermediate forms.

978-1-4244-3296-7/09/$25.00 ©2009 IEEE

Although the degree of fractionation of the A-EGM signals in the experimental dataset was a continuous variable by nature, expert classification into categories was chosen for the purpose of our study. Three experts, who perform AF ablation procedures on regular basis, independently ranked raw A-EGMs into those 4 classes of fractionation (1 - organized activity; 2 - mild degree of fractionation; 3 - intermediate degree of fractionation; 4 - high degree of fractionation) according to the subjective perception of signals (Fig. 1). This procedure was facilitated by the purpose-written software for displaying A-EGMs in the same aspect ratio as on real-time screen during the left atrium mapping. This software also allowed experts scrolling through all A-EGMs with the possibility to reorder them repetitively according to assigned classification until the final ranking was reached. No specific criteria for signal assessment (e.g. dominant frequency or percentage of continuous electrical activity) were given. The experts were asked to classify the A-EGMs by their subjective judgment according to how the ablation at particular site would be valuable for atrial debulking. They were only instructed to keep approximately equal percent occurrence for each A-EGM category. According to this procedure these 68 A-EGMs, where were 100% coincident in ranking, were selected with this distribution to classes I to IV – 19, 17, 21, and 11.

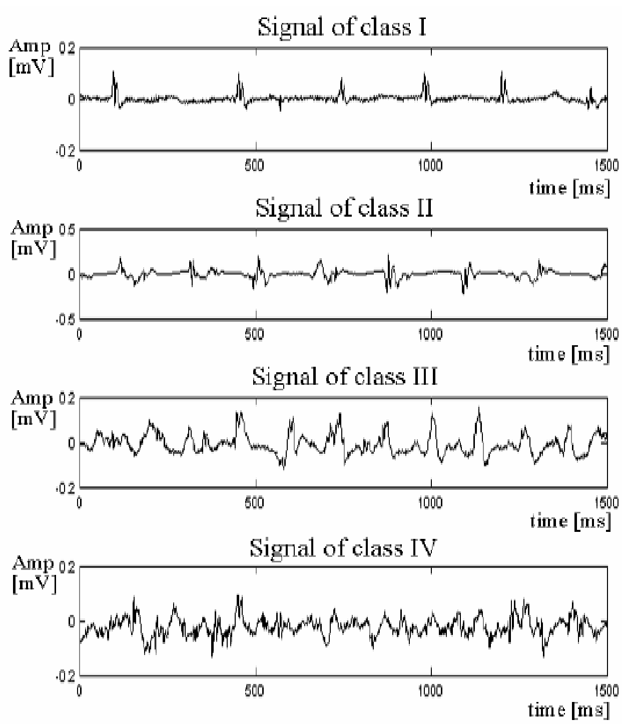

Fig. 1. Four complex fractionated electrograms are shown. These are representatives of each ranking class of degree of fractionation ranked by three independent experts. From the top to bottom: 1 - organized atrial activity; 2 - mild, 3 - intermediate; 4 - high degree of fractionation.

B. A-EGMs processing and feature extraction

A-EGM preprocessing algorithm described in [8] were used to filter and prepare signals for the phase of feature extraction. The algorithms for A-EGM feature extraction [9] were used to describe A-EGM complexity in a new way. Based on the Automated Fractionated segments Search (AFS) preprocessing algorithm [8], the algorithms automatically search for areas of the A-EGM signal, where local electrical activity is found (FSs), also described by Faes et al. as local activation waves (LAWs) [6]. Several features of A-EGM are then defined based on FSs description. Following seven features are therefore derived from the characteristics of the automatically observed FSs or LAWs:

1) A number of fractionated segments found by AFS in particular A-EGM signal in dataset.
2) Minimum of Inflection Points in found FSs in a particular A-EGM signal.
3) Maximum of Inflection Points in found FSs in a particular A-EGM signal. Arithmetical Mean value of Inflection Points added together in automatically found FSs.
4) Sum of width of all FSs found in particular A-EGM.
5) Minimal width of found FSs in particular A-EGM signal.
6) Maximal width of found FSs in particular A-EGM signal.
7) Arithmetical Mean value of inter-segment distance of automatically found FSs.

C. Used Classifiers

We used the extracted features to construct several automated classifiers that enable to give an operator independent look on A-EGM signal and classify its degree of fractionation. Used methods allow to extract useful knowledge from data. In this contribution we studied the results of individual classifiers to compare their performance. All classifiers were well established methods available in the WEKA data mining environment [10]: j48 (j48 decision tree), mlp (multilayer perceptron), rbfn (radial bases function network); Matlab environment [11]: Normal density based classifiers – ldc (linear density based classifier), qdc and ndc (quadratic and normal density based classifier); Linear and high degree polynomial classifiers – loglc (logistic linear), nmc (nearest mean), nmsc (nearest mean scaled); Nonlinear – parzenc (Parzen), knnc (k-nearest neighbor), naivebc (naive Bayes), and bpxnc (neural network classifier trained by back-propagation). Then we used also our GAME classifier implemented in Java within the FAKE GAME environment [12].

D. GAME classifier

The Group of Adaptive Models Evolution (GAME) algorithm combines neurons of several different transfer function within one neural network (see Fig. 2). The structure, connections and type of neurons are evolved by menas of special niching genetic algorithm. The parameters of neurons are adjusted by gradient based optimization methods (e.g. Quasi-Newton) on training data subset, whereas from validation subset, the fitness of neurons is computed. Evolved structure reflects on the complexity of data set and relationsships of system variables.

For detailed description of the GAME method, please refer to [12].

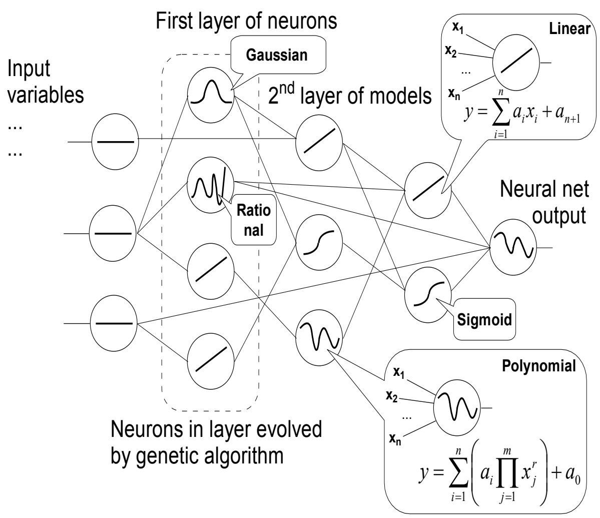

Fig. 2. An example of GAME neural network generated layer by layer using the niching genetic algorithm consists of neurons with different transfer functions. The complexity of the network increases together with the classification accuracy on the validation dataset.

E. Methodology

For the comparison, we repeated ten-fold cross validation 10 times to get reliable results. Results are presented in form of statistically well transparent box plot charts.). All methods from Matlab and Weka environment were applied using their default settings. The configuration of the GAME neural network was the following. We enabled linear transfer function neuron; polynomial neurons with random inicialization, genetic evolution and combinatorial selection of the transfer function; three variants of gaussian transfer function neuron; sine, sigmoid, rational and exponential neurons. The genetic algorithm runed for 30 generation with population of 15 neurons in every layer. The distance of neurons was computed as a combination of their genotypic distance and inverse correlation of their errors. Training and validation data were split 50%/50%. Only Quasi-Newton method was enabled to optimize parameters of neurons.

III. Results

In Weka data mining environment, the worst performing classifier was the decision tree (j48), while the multilayered perceptron trained by backpropagation (MLP) was the best performing algorithm. The GAME classifier performed slightly (the difference is not statistically significant) worse than MLP. In Matlab environment, implemented classifiers had very good results, aside from qdc and loglc. The maximal error of 15% and mean error across all well performing classifiers 9%, and with the best mean classification error 5.9% (nmc), see Fig. 4, and classification error of highly fractionated A-EGMs of ~ 9%.

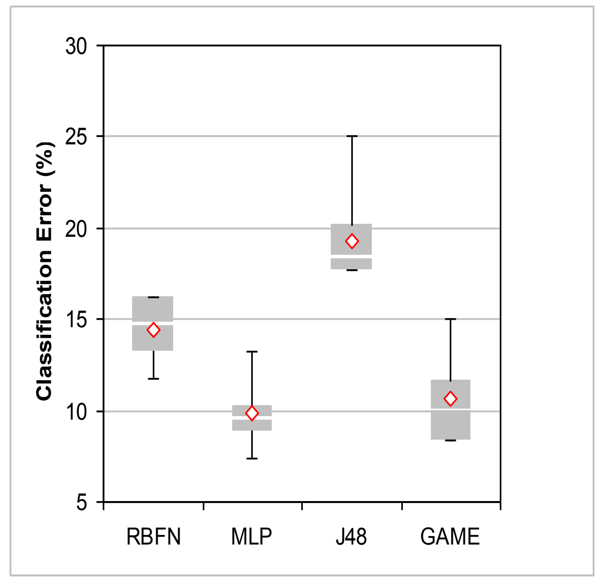

Fig. 3. Classification error in percent and comparison Weka classifiers and GAME.

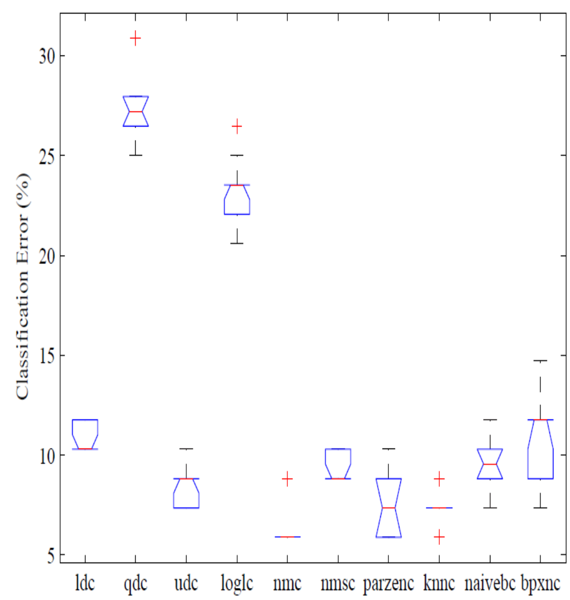

Fig. 4. Comparison of results of A-EGM classifiers used in Matlab environment.

IV. Conclusion

Our results show, that extracted features for CFAE bear significant information allowing us to classify A-EGM signals into all classes with high accuracy. That is good result with respect to 60% of consistent assignments into four classes performed by experts [8], [9] and [13]. For this data set, simple algorithms (nearest neighbor and nearest mean) implemented in Matlab significantly outperformed well established data mining methods and also our GAME algorithm.

Our future work is to incorporate such simple processing elements as neurons into the GAME classifier. Then the genetic algorithm should automatically choose such neurons when classifying a data set with similar character as the A-EGM dataset. Such algorithm resembles several metalearning strategies such as stacking classifiers [14].

A strong selection of the signals to be included into the database was done (given the initial acquisitions, just 68 "high-quality" segments with a length of 1.5 sec each were selected) in order to exclude all the noisy or, in general, not-optimal signals: this fact could have increased the amount of right recognitions of the algorithms but, if the algorithm is intended for real-time application the studies on the capabilities of the system in recognizing artifacts should be performed in future. We plan to do such work in the future studies.

An analysis, how the performances of the classifiers vary by selectively reducing the sets of inputs (in order to understand if all the features are mutually independent and really useful to obtain an high amount of correct classifications) is another aspect of the classification task. Though given the low computational complexity needed in order to evaluate the described features, the reduction of the number of input of the classifier should not be an essential issue, but this facet of the problem is anyway interesting from a signal processing and classification point of view. This study is partially introduced in [9] but is mainly planned for the next work on this theme of CFAEs classification.

In conclusion, we proposed novel algorithms for A-EGM processing and feature extraction based on wavelet transform. And far more, we tested and showed that these algorithms are suitable for the next steps of automatic A-EGMs evaluation to classify level of complexity of A-EGMs. These algorithms then can be used for automated and operator-independent assessment of A-EGMs fractionation to facilitate CFAEs identification and to guide AF substrate ablation. Because of the low computational costs it can be easily incorporated into real-time mapping systems provided it will be first validated off-line in larger and independent A-EGMs sample and compared with currently available algorithms. By now, its clinical value is unknown and warrants further investigation.

REFERENCES

[1] Haïssaguerre M, Wright M, Hocini M and Jaïs P 2008 The substrate maintaining persistent atrial fibrillation *Circ. Arrhythmia Electrophysiol.* **1** 2-5

[2] Sanders P, Berenfeld O, Hocini M, Jais P, Vaidyanathan R, Hsu L F, Garrigue S, Takahashi Y, Rotter M, Sacher F, Scavee C, Ploutz-Synder R, Jalife J and Haissaguerre M 2005 Spectral analysis identifies sites of high-frequency activity maintaining atrial fibrillation in humans *Circulation* **112** 789–797

[3] Nademanee K, McKenzie J, Kosar E, Schwab M, Sunsaneewitayakul B, Vasavakul T, Khunnawat C and Ngarmukos T 2004 A new approach for catheter ablation of atrial fibrillation: mapping of the electrophysiologic substrate *J. Am. Coll. Cardiol.* **43** 2044–2053

[4] Kumagai K 2007 Patterns of activation in human atrial fibrillation *Heart Rhythm* **4** (3)Suppl. S7–S12

[5] Mainardi L T, Corino V D, Lombardi L, Tondo C, Mantica M, Lombardi F and Cerutti S 2004 Assessment of the dynamics of atrial signals and local atrial period series during atrial fibrillation: effects of isoproterenol administration *Biomed. Eng. Online* **3** 37

[6] Ravelli F, Faes L, Sandrini L, Gaita F, Antolini R, Scaglione M and Nollo G 2005 Wave similarity mapping shows the spatiotemporal distribution of fibrillatory wave complexity in the human right atrium during paroxysmal and chronic atrial fibrillation *J. Cardiovasc. Electrophysiol.* **16** 1071–1076.

[7] V. Křemen and L. Lhotská. Novel approach to search for individual signal complexes in Complex Fractionated Atrial Electrograms using wavelet transform. *In 6th International Special Topic Conference on ITAB - Proceedings*, pages 83-86, Piscataway, 2007. IEEE.

[8] Křemen, V., Lhotská, L., Macaš, M., Čihák, R., Kautzner, J., et al.: A New Approach to Automated Assessment of Fractionation of Endocardial Electrograms During Atrial Fibrillation. *Physiological Measurement*. 2008, vol. 29, no. 12, p. 1371-1381. ISSN 0967-3334.

[9] Křemen, V. Automated Assessment of Endocardial Electrograms Fractionation in Human. PhD Thesis. *Department of Cybernetics, Faculty of Electrical Engineering, Czech Technical University in Prague*. 2008.

[10] Weka open source data mining software. Available online at http://www.cs.waikato.ac.nz/ml/weka/, September 2006.

[11] Matlab, the language of technical computing. ver. 7.0. http://www.mathworks.com.

[12] P. Kordík. *Fully Automated Knowledge Extraction using Group of Adaptive Models Evolution*. PhD thesis, Czech Technical University in Prague, FEE, Dep. of Comp. Sci. and Computers, FEE, CTU Prague, Czech Republic, September 2006.

[13] P. Kordík, V. Křemen and L. Lhotská. The GAME Algorithm Applied to Complex Fractionated Atrial Electrograms Data Set. In Proceedings of 18th International Conference on Artificial Neural Networks, Prague, 3-6 September, 2008.

[14] Brazdil et al.: Metalearning, Springer-Verlag Berlin Heidelberg 2009, ISBN: 978-3-540-73262-4