



Redefining Performance Evaluation Tools for Real-Time QRS Complex Classification Systems

Philippe Ravier, F. Leclerc, C. Dumez-Viou, G. Lamarque

► To cite this version:

Philippe Ravier, F. Leclerc, C. Dumez-Viou, G. Lamarque. Redefining Performance Evaluation Tools for Real-Time QRS Complex Classification Systems. IEEE Transactions on Biomedical Engineering, 2007, 54 (9), pp.1706–1710. 10.1109/TBME.2007.902594 . hal-00600046

HAL Id: hal-00600046

<https://hal.science/hal-00600046>

Submitted on 13 Jun 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Redefining performance evaluation tools for real-time QRS complex classification systems

Philippe Ravier*, Frédéric Leclerc, Cedric Dumez-Viou, and Guy Lamarque

Abstract

In a heartbeat classification procedure, the detection of QRS complex waveforms is necessary. In many studies, this heartbeat extraction function is not considered: the inputs of the classifier are assumed to be correctly identified. This paper aims to redefine classical performance evaluation tools in entire QRS complex classification systems and to evaluate the effects induced by QRS detection errors on the performance of a heartbeat classification processing (normal vs abnormal). Performance statistics are given and discussed considering the MIT/BIH database records that are replayed on a real-time classification system composed of the classical detector proposed by Hamilton & Tompkins, followed by a neural network classifier. This study shows that a classification accuracy of 96.72% falls to 94.90% when a drop of 1.78% error rate is introduced in the detector quality. This corresponds an increase of about 50% bad classifications.

Index Terms

QRS complex detection, heartbeat recognition, neural network, classification, hardware implementation.

I. INTRODUCTION

The problem of heartbeat classification has been widely explored in the literature [1]-[8]. However, real-time classification systems often necessitate the detection of QRS complex shapes before any classification procedure. In such classification systems, classical performance evaluation tools become irrelevant since the detection errors are not taken into account in the classification statistics. In this paper, we therefore propose to:

- redefine the classical performance evaluation tools taking the entire classification system into consideration;
- experimentally discuss the influence of the detection stage on classification results;
- give material for finding the best compromise between the quality of the detector, the quality of the classifier and the computational time in real-time classification systems.

*P. Ravier, C. Dumez-Viou and G. Lamarque are with the Laboratory of Electronics, Signals and Images, University of Orleans, BP 6744, 45067 Orléans Cedex, France (e-mail: philippe.ravier@univ-orleans.fr; cedric.dumez-viou@obs-nancay.fr; guy.lamarque@univ-orleans.fr).

F. Leclerc is with the Laboratory of Electronics, Signals and Images, University of Orleans, BP 6744, 45067 Orléans Cedex, France (e-mail: frederic.leclerc@univ-orleans.fr) and the Department of Aerospace Physiology, Institute of Aerospace Medicine of the Army Health Department, BP 73, 91223 Brétigny-Sur-Orge Cedex, France.

In order to achieve these goals, we implemented a simple real-time classification system on an electronic board. The proposed system is composed of a QRS complex extractor (detection part) followed by a normal or abnormal peak recognition step (classification part). Abnormal beats are opposed to normal beats according to their QRS waveform shapes.

For instance, a normal/abnormal classification system may be useful for Holter monitoring where only the critical parts as well as the cardiac history (*i.e.* peak to peak intervals) are of medical interest. It is not worth recording the healthy electrocardiographic (ECG) samples. The amount of data can therefore be drastically reduced and the entire breakdown will be shorter since the physician will devote his expertise to the critical parts of the ECG signal, while also taking into account the peak to peak intervals of the entire monitoring.

Several algorithms have been proposed in the literature for the detection and classification of ECG beats. Since our system has to be embedded with real-time processing constraints, the computation cost must be low. Various approaches (wavelets [9], [10], filter banks [11], neural networks [12]) have been investigated to improve the quality of the detectors. However, the gain obtained is generally offset by the greater complexity of the algorithms, involving higher computational costs. It is the reason why the well known Hamilton & Tompkins [13], [14] detector has been chosen here. Similarly, many methods have been investigated for the classification part (neural networks [1]-[3], fuzzy theory [3], [5], support vector machine [4], higher order spectral techniques [6], hidden Markov models [7], ...). A neural network approach has been adopted in this study because its parallel processing implementation is well adapted to real-time constraints. Finally, we tested the system on the MIT-BIH database [15] since this database is commonly used for performance evaluations of detection or classification algorithms.

The paper is organized as follows. The classification system is introduced in the second part. Such a system leads to the redefinition of the performance evaluation tools in the third part. These new tools are tested on real data in the fourth part, considering real-time classification of the MIT/BIH recordings. Discussion and conclusion finish the article.

II. NORMAL/ABNORMAL REAL-TIME BEAT CLASSIFICATION SYSTEM

After digitizing the analog ECG signal through a 16 bits Analog to Digital Converter (ADC) using a 360 Hz sampling rate, the algorithmic part is divided into two main functions (Fig. 1):

- the ECG beat extractor that detects QRS complex waveforms and extracts ECG waveform patterns after they have been normalized and centered about the fiducial point. Two versions of the Hamilton and Tompkins detector have been real-time implemented [13], [14]: the first one includes the search back procedure (SB+) and the second one does not (SB-).
- the neural network classifier that is fed by Fourier coefficients of the ECG waveform patterns provided by the previous stage. The classifier is a multilayer perceptron (MLP) [16] with 16 input nodes (16 first absolute values of the 128 Fourier samples excluding the DC component), 4 neurons in the hidden layer and one output neuron. Each neuron labeled 1 to 5 has four inputs and one output with the classical associated sigmoidal activation function $y_j(n) = \frac{1}{1+e^{-v_j(n)+\theta_j}}$, where n is the number of QRS pattern to be classified

and j the neuron number. The value $v_j(n)$ is the net internal activity level of neuron j and equals $v_j(n) = \sum_{i=1}^4 w_{ji}x_{ji}(n)$, for $j = 1$ to 4 and $v_5(n) = \sum_{i=1}^4 w_{5i}y_i(n)$. The output $S(n) = y_5(n)$ is compared to a threshold $\eta = 0.5$ to make the decision. The weights w_{ji} and bias θ_j are computed with the software SNNS [17]: $w_{1i} = \{7.62, 2.98, -11.06, 1.57\}$, $w_{2i} = \{8.08, 1.71, 3.31, -0.5\}$, $w_{3i} = \{-4.4, 24.86, 4.74, 8.58\}$, $w_{4i} = \{4.56, 4.6, 19.86, 4.89\}$, $w_{5i} = \{-8.85, -29.81, 11.14, 10.56\}$ and $\theta_j = \{2.15, -5.45, -3.16, -1.61, 7.21\}$. The two output classes separate the normal beats (expert annotations 'N' and 'A') from the abnormal beats (expert annotations 'L', 'R', 'B', 'a', 'J', 'S', 'V', 'r', 'F', 'e', 'j', 'n', 'E', '/', 'f', 'Q' and '?').

The detector is implemented on a Digital Signal Processor (DSP TMS320VC5402, Texas Instruments). The pre-processing of the input delivered to the neural network is also computed on the DSP. The neural network is parallel processed on a Field Programmable Gate Array architecture (FPGA XC2S200, Xilinx).

III. REDEFINING PERFORMANCE EVALUATION TOOLS

Heartbeat detection consists in discriminating at any time a QRS complex from the noise. In fact, this detection problem may be formulated as a binary classification problem, in which the two classes are the noise and the QRS complex. The same statistical tools will therefore be used to evaluate the performances of both classifiers: the QRS complex/noise classifier and the normal/abnormal beat classifier. However, the statistics for the noise class will not be defined since they provide no useful information.

A. Sensitivity S and positive predictivity P^+

Making the decision within the i th class may be true or false. Each right decision increments true numbers whereas each erroneous decision increments false numbers in the following statistics: the True Positive number (TP_i) stands for the right classification number among the patterns belonging to the i th class; the False Positive number (FP_i) is the number of patterns that were found to be in the i th class while they did not belong to this class; the False Negative number (FN_i) is the number of patterns that were not found to be in the i th class while they belonged to this class. Note that, with only the two classes {Normal, Abnormal}, the following relations hold: $FP_{Normal} = FN_{Abnormal}$ and $FP_{Abnormal} = FN_{Normal}$.

The comparison of classification performance is carried out from the values given by the sensitivity S and the positive predictivity P^+ . Classically, sensitivity is the percentage of patterns truly assigned to the class i among all the patterns really belonging to this class (expert decision). Positive predictivity (also named specificity) is the percentage of patterns truly assigned to the class i among all the patterns assigned to this class by the classifier (machine decision). These statistics quantify each kind of error in the decision making since S involves FN and P^+ involves FP . Considering the single detector or the single classifier, these definitions clearly hold. However, they are no longer valid when considering the global classifier composed of both the classifier and the detector where detection errors must be taken into account.

In particular, the detection errors lead to a number of classified beats that may be different from the number of beats presented at the input of the global classifier. This difference is due to the erroneous FN_{QRS} and FP_{QRS}

decisions. The FN_{QRS} are the number of beats which have not been detected at the first stage. These beats will thus never be classified. However, these beats are known to be either normal or abnormal beats so the FN_{QRS} number may be broken down as $FN_{QRS} = FN_{QRS_{Normal}} + FN_{QRS_{Abnormal}}$ (medical expertise separation). On the other hand, the FP_{QRS} number introduces new false QRS beats. Since these waveforms are detected, they will be classified as normal ($FP_{QRS_{Normal}}$) or abnormal ($FP_{QRS_{Abnormal}}$) while they do not belong to either of these classes. So the false alarm detection rate may be broken down as $FP_{QRS} = FP_{QRS_{Normal}} + FP_{QRS_{Abnormal}}$ (machine separation). The sensitivity S and the positive predictivity P^+ are then redefined by introducing the detection errors as:

$$S_i = \frac{TP_i}{TP_i + FN_i + FN_{QRS_i}}; P_i^+ = \frac{TP_i}{TP_i + FP_i + FP_{QRS_i}} \quad (1)$$

for the classes $i = Normal, Abnormal$.

Regarding S_i , all the beats classified by the experts as belonging to the class i may become correctly classified (incrementing the number TP_i), or become classified in the other class (incrementing the number FN_i) or even lost forever (incrementing the number FN_{QRS_i}). Regarding P_i^+ , all the beats classified by the machine to the class i may come from correctly classified beats (incrementing the number TP_i), from misclassified beats (incrementing the number FP_i) or from the false alarm beats generated by the detector (incrementing the number FP_{QRS_i}).

B. Total classification accuracy TCA

The quality of the beat classification machine is classically evaluated through the Total Classification Accuracy (TCA) which corresponds to the fraction of good classifications among the number of beats to be classified. Regarding the global classification machine, the TCA is not sufficiently pertinent since it does not take into account the false alarms generated by the detector. For example, if the classification machine is characterized by a 100% TCA rate (estimated with no detection errors), the physician will have total confidence in the classification task even if the FP_{QRS} attains a high level. In that case, when these non existing beats are classified as abnormal beats, the machine will alert the physician without any reason. Similarly, a 100% TCA rate combined with a high level FN_{QRS} could lead to a dramatic situation because of a lack of complexes analyses. For all these reasons, the TCA must be redefined in order to take into account not only the quality of the classifier but also the quality of the detector.

For the physician, the classification system is a black box characterized by an estimated TCA which needs to reflect the observations. In order to retrieve confidence in the TCA , we propose a novel definition of this performance evaluation tool. Let us call $TCE = 1 - TCA$ the Total Classification Error rate expressed as:

$$TCE = \frac{FP_{Normal} + FP_{Abnormal} + FP_{QRS} + FN_{QRS}}{N_{Normal} + N_{Abnormal}} \quad (2)$$

The false alarm number FP_{QRS} really stands for non-classifications. Since the classification is binary, all the FP_{QRS} beats could be considered as misclassifications. Let us now consider the missed detections FN_{QRS} . If all these beats could be correctly detected, the worst case that could occur would be a misclassification of all these

beats. So the $FP_{QRS} + FN_{QRS}$ beats are included in the TCE definition that describes the worst case. Finally, the TCA is computed as the complementary value of the TCE .

These redefined statistics have been used to characterize the real-time classification process of 23 free records of the MIT/BIH database. In order to evaluate the influence of the detection stage on the classification performance, the results for three detectors are presented (table I): the detector is perfect (the beats are actually extracted by the cardiologist's expertise); the detector makes erroneous decisions at the mean rate of 0.85% (result for the single detector when the search back rule is implemented, label SB+); the detector makes twice as many erroneous decisions at the mean rate of 1.78% (result for the single detector when no search back rule is implemented, label SB-).

IV. DISCUSSION

From the results presented in table I, it can be seen that our system shows coherent performance values in comparison with other single detectors or classifiers [8], [9], [11]. In fact, the only objective performance comparison that can be made is with the detector proposed by Hamilton and Tompkins. Our results (0.85% detection error rate) slightly differ from those presented in [14] (0.52%) because not all the refinements suggested in [14] have been implemented, in order to optimize the computation time.

A. General comments on the results

For those records where the proportion of a given class with respect to the other class is highly unbalanced, the results in percentage of S and P^+ are clearly not significant. This explains that specificity and positive predictivity may vary from the extreme percentages 0% to 100% or even NaN (Not a Number). This is the case for the files 100, 101, 103, 107, 209, 214 (see tables I and II where the number of expert beats in the normal or abnormal class is less than or equal to 2). Our system presents high P^+ values with around 96% of the classified beats that are correctly assigned. Concerning S , from a medical point of view, a high $S_{Abnormal}$ rate is preferable to a high S_{Normal} rate since the alarms are triggered on abnormal events. Our classifier shows the opposite tendency (94.02% vs 98.14%). However, at a $S_{Abnormal} = 94.02\%$ rate, the clinician will have at his disposal a sufficient number of abnormal beats for his expertise.

The Search Back method is a rule of reduced threshold backward detection that is triggered whenever no beat is detected during more than 150% of the previous 8 beat-to-beat intervals median elapsed time [14]. The analysis of the evolution of the TCA according to the quality of the detector (Table II) shows the efficiency of the search back rule for a few records: 106, 201, 210, 213, 214 and more particularly 208 (where FN_{QRS} decreases from 462 to 46). Overall, the Search Back method strongly decreases the FN_{QRS} number (821 to 182 counts) while increasing the FP_{QRS} number (223 to 314 counts) at the same time. The influence of these detection errors on the quality of classification is discussed in the next section.

B. Influence of the detection errors on the S and P^+ statistics

The quality of detection globally seems to have no influence concerning normal beats: the 1.78% detection error rate induced by the SB- detector - with respect to a perfect detector - does not really change S_{Normal} and P_{Normal}^+ (Table I).

On the contrary, concerning abnormal beats, the drop affecting the sensitivity becomes serious when working with imperfect detectors. The observed difference of 4.12% between S_{Normal} and $S_{Abnormal}$ with a perfect detector now drops to 8.74% when the SB- detector is used. At a reduced scale, the observed difference of 0.49% between P_{Normal}^+ and $P_{Abnormal}^+$ becomes 0.98% with the SB- detector. This bad news concerning the statistics for abnormal beats may be significant and suggests that the quality of detection should not be neglected within a classification procedure.

C. Influence of the detection errors on the TCA statistics

A coherent behavior of the TCA with respect to the quality of the detector is observed on average in table I: the TCA decreases with the deterioration of the detection performances. An error rate of 1.78% at the first detection stage produces a loss of 1.82% in the TCA ratio (Table I). However, this ratio may individually lose about 16% (record 208) or even gain about 0.7% (record 200).

In particular, the biggest deterioration is shown in record 208 with an increase from 46 (SB+) to 462 (SB-) in the FN_{QRS} number (table II). The result is a marked decrease in the classification ratio from $TCA_{(SB+)} = 93.13\%$ to $TCA_{(SB-)} = 79\%$. Note that the FN_{QRS} number is mainly due to the contribution of abnormal beats. Indeed, the $FN_{QRS_{Abnormal}}$ number at the denominator of $S_{Abnormal}$ makes sensitivity dramatically decrease from 91.96% (ideal detector) to 58.00% (detector with SB- procedure) whereas minor changes are observed in the other cases for S and P^+ . Thus the importance of having a good detection stage is once again demonstrated.

In average, this study shows that a classification accuracy of 96.72% falls to 94.90% (table I) when a drop of 1.78% error rate is introduced in the detector quality. Thus, for quite good classifiers, particular care must be taken with the quality of the first detection stage since this increases the number of bad classifications by about 50%. In this case, it is worth trying to improve the detection algorithm, provided the real-time constraint is fulfilled.

V. CONCLUSION

In this paper, we have redefined the performance evaluation tools for characterizing a complete classifier, which consists of a detector followed by a single classifier. The detection errors are introduced in these new definitions, providing more coherent statistics for the clinician. These new statistics increase safety since they now provide reliable lower limits for the performance values. We experimentally tested these new tools on the MIT/BIH database with a dedicated real-time hardware and software architecture.

The proposed tools can be applied considering other kinds of algorithms and architectures. The comparison between various implementation configurations therefore becomes objective.

Even if an exhaustive comparative study is not presented in this paper, the results obtained are quite significant. Indeed, the MIT/BIH database presents a great diversity of pathologies, exhibiting a good panel of QRS waveforms morphologies. The results show that the detection stage has a reduced influence concerning the sensitivity for normal beats and the positive predictivity for normal/abnormal beats. This suggests that the classification system may be improved by investigating the classification stage. Nevertheless, care must be taken with the detector stage that may deteriorate the total classification error rate by about 50%, considering a drop of 1.78% error rate in the detection quality. In conclusion, the higher the performance of the classifier, the greater the impact of the detector.

ACKNOWLEDGMENT

The authors would like to thank CRESITT Industrie for the design and realization of the DSP-FPGA hybrid electronic board (<http://www.cresitt.com>).

REFERENCES

- [1] K. Minami, H. Nakajima and T. Toyoshima, "Real-time discrimination of ventricular tachyarrhythmia with Fourier-transform neural network," *IEEE Trans. Biomed. Eng.*, vol. 46, no 2, pp. 179-185, Feb. 1999.
- [2] Z. Dokur and T. Ölmez, "ECG beat classification by a novel hybrid neural network," *Comput. Methods and Prog. an Biomed.*, vol. 66, pp. 167-181, 2000.
- [3] L.-Y. Shyu, Y.-H. Wu and W. Hu, "Using wavelet transform and fuzzy neural network for VPC detection from the Holter ECG," *IEEE Trans. Biomed. Eng.*, vol. 51, no 7, pp. 1269-1273, July 2004.
- [4] S. Osowski, L. T. Hoai and T. Markiewicz, "Support vector machine-based expert system for reliable heartbeat recognition," *IEEE Trans. Biomed. Eng.*, vol. 51, no 4, pp. 582-589, April 2004.
- [5] S. Osowski and L. T. Hoai, "ECG beat recognition using fuzzy hybrid neural network," *IEEE Trans. Biomed. Eng.*, vol. 48, no 11, pp. 1265-1271, Nov. 2001.
- [6] L. Khadra, Y.S. Al-Fahoum and S. Binajjaj, "A quantitative analysis approach for cardiac arrhythmia classification using higher order spectral techniques," *IEEE Trans. Biomed. Eng.*, vol. 52, no 11, pp. 1840-1845, Nov. 2005.
- [7] R.V. Andreão, B. Dorizzi and J. Boudy, "ECG analysis through hidden Markov models," *IEEE Trans. Biomed. Eng.*, vol. 53, no 8, pp. 1541-1549, August 2006.
- [8] Y.H. Hu, W. J. Tompkins, J.L. Urrusti and V.X. Afonso, "Applications of artificial neural networks for ECG signal detection and classification," *J. Electrocardiology*, vol. 26 Supplement, pp. 66-73, 1994.
- [9] C. Li, C. Zheng and C. Tai, "Detection of ECG characteristic points using wavelet transforms," *IEEE Trans. Biomed. Eng.*, vol. 42, no 1, pp. 21-28, Jan. 1995.
- [10] S. Kadambe, R. Murray and G. F. Boudreaux-Bartels, "Wavelet transform-based QRS complex detector," *IEEE Trans. Biomed. Eng.*, vol. 46, no 7, pp. 838-848, Jul. 1999.
- [11] V. X. Afonso, W. J. Tompkins, T. Q. Nguyen and S. Luo, "ECG beat detection using filter banks," *IEEE Trans. Biomed. Eng.*, vol. 46, no 2, pp. 192-201, Feb. 1999.
- [12] Q. Xue, Y. Hen Hu and W. J. Tompkins, "Neural-network-based adaptive matched filtering for QRS detection," *IEEE Trans. Biomed. Eng.*, vol. 38, no 4, pp. 317-329, Apr. 1992.
- [13] J. Pan and W. J. Tompkins, "A real-time QRS detection algorithm," *IEEE Trans. Biomed. Eng.*, vol. 32, no 3, pp. 230-236, Mar. 1985.
- [14] P. S. Hamilton and W. J. Tompkins, "Quantitative investigation of QRS detection rules using the MIT/BIH arrhythmia database," *IEEE Trans. Biomed. Eng.*, vol. 33, no 12, pp. 1157-1165, Dec. 1986.
- [15] R. Mark and G. Moody, *MIT-BIH Arrhythmia Database Directory*, Cambridge, MA:MIT, 1988.
- [16] S. Haykin, *Neural networks: a comprehensive foundation*. New-York: Macmillan College Publishing, 1994.
- [17] A. Zell and al., *SNNS, Stuttgart Neural Network Simulator*, User Manual, Version 4.2, University of Stuttgart, www-ra.informatik.uni-tuebingen.de/SNNS/, 1998.

LIST OF FIGURES

1 Normal/abnormal beat classification: the ECG beat extraction function is compounded in the first two blocks; the neural based classification function is compounded in the last two blocks. 9

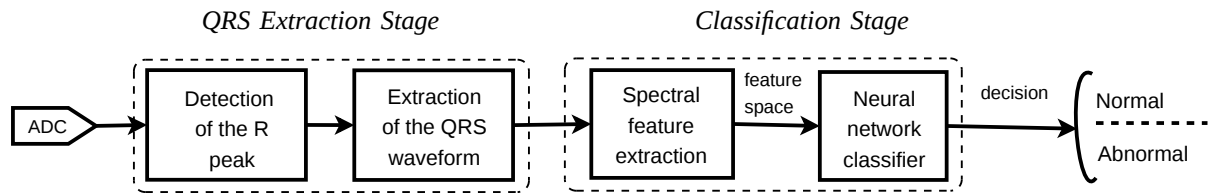


Fig. 1. Normal/abnormal beat classification: the ECG beat extraction function is compounded in the first two blocks; the neural based classification function is compounded in the last two blocks.

LIST OF TABLES

I	Evolution of the classifier performance (S_i , P_i^+ and TCA) with the quality of the detector at the input stage: 0%, 0.85% (labeled SB+), 1.78% (labeled SB-) detection errors. For those records where $TP_i = FN_i = 0$ (resp. $TP_i = FP_i = 0$), S_i (resp. P_i^+) is assigned to NaN (Not a Number) because its denominator is zero.	11
II	Beat-by-beat, record-by-record testing results of the experiment. Results for the real-time classification machine: the detection stage (labels SB+ and SB-), the single classification stage and the global classification machine (labels SB+ and SB-)	12

TABLE I

EVOLUTION OF THE CLASSIFIER PERFORMANCE (S_i , P_i^+ AND TCA) WITH THE QUALITY OF THE DETECTOR AT THE INPUT STAGE: 0%, 0.85% (LABELED SB+), 1.78% (LABELED SB-) DETECTION ERRORS. FOR THOSE RECORDS WHERE $TP_i = FN_i = 0$ (RESP. $TP_i = FP_i = 0$), S_i (RESP. P_i^+) IS ASSIGNED TO NaN (NOT A NUMBER) BECAUSE ITS DENOMINATOR IS ZERO.

File #	Sensitivity (%)						Positive Predictivity (%)						TCA (%)		
	Normal			Abnormal			Normal			Abnormal			T	$T_{(SB+)}$	$T_{(SB-)}$
	S	$S_{(SB+)}$	$S_{(SB-)}$	S	$S_{(SB+)}$	$S_{(SB-)}$	P^+	$P_{(SB+)}^+$	$P_{(SB-)}^+$	P^+	$P_{(SB+)}^+$	$P_{(SB-)}^+$			
100	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
101	99.36	99.19	99.36	100.0	33.33	33.33	100.0	100.0	100.0	14.29	4.55	5.26	99.36	98.77	98.93
102	7.07	5.05	5.05	99.95	99.95	99.95	87.50	83.33	83.33	95.78	95.69	95.69	95.75	95.66	95.66
103	100.0	100.0	100.0	NaN	0.00	0.00	100.0	100.0	100.0	NaN	NaN	NaN	100.0	99.95	99.81
104	82.82	86.42	87.04	94.19	92.69	92.60	52.94	48.95	50.36	98.58	93.33	94.99	93.36	86.67	88.38
105	90.58	90.96	90.44	50.00	40.32	33.87	99.00	98.83	98.65	8.81	7.76	6.65	89.85	86.74	86.70
106	99.93	99.87	99.87	79.42	76.92	63.08	93.37	92.79	93.02	99.76	99.50	99.39	94.67	93.98	90.43
107	NaN	NaN	NaN	99.86	99.86	99.77	0.00	0.00	0.00	100.0	100.0	100.0	99.86	99.81	99.72
118	1.04	1.04	1.04	99.31	99.22	99.22	6.25	5.26	5.26	95.80	95.67	95.67	95.17	94.91	94.91
119	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
200	96.84	97.29	97.52	83.70	84.20	85.40	92.71	92.89	93.56	92.52	92.33	93.40	92.66	92.54	93.35
201	99.40	100.0	100.0	84.42	71.84	60.19	97.16	96.61	96.84	96.30	100.0	100.0	97.04	95.57	93.73
202	99.48	99.57	99.57	66.67	60.00	35.00	99.38	99.48	99.48	70.27	72.73	60.87	98.88	98.83	98.36
205	99.92	100.0	100.0	81.71	78.57	77.38	99.42	99.42	99.46	97.10	100.0	100.0	99.36	99.32	99.28
208	97.54	98.10	98.09	91.96	88.24	58.00	93.36	92.91	92.78	97.00	96.89	95.47	94.95	93.13	79.00
209	99.73	99.67	99.63	100.0	100.0	100.0	100.0	100.0	100.0	11.11	8.33	7.69	99.73	99.63	99.60
210	97.48	97.89	97.93	68.72	63.64	50.87	97.08	97.09	97.21	71.89	71.36	69.23	95.02	94.45	93.62
212	98.16	97.72	98.16	99.07	99.07	99.12	98.16	98.26	98.37	99.07	98.85	99.07	98.76	98.62	98.80
213	99.96	100.0	100.0	57.44	57.00	53.41	91.46	91.42	91.36	99.70	100.0	100.0	92.31	92.25	91.60
214	NaN	NaN	NaN	93.68	90.58	89.78	0.00	0.00	0.00	100.0	99.85	99.90	93.67	90.44	89.60
215	99.94	99.94	99.94	81.21	77.58	76.97	99.04	98.89	98.89	98.53	97.71	97.69	99.02	98.81	98.78
217	100.0	99.59	99.59	98.07	98.01	97.81	86.52	87.10	86.79	100.0	99.84	99.90	98.28	98.10	97.92
219	99.86	99.86	99.95	84.62	84.62	74.63	99.52	99.52	99.52	94.83	94.83	98.04	99.40	99.40	99.16
all	98.14	98.27	98.27	94.02	92.73	89.53	96.88	96.56	96.58	96.39	95.41	95.60	96.72	95.89	94.90

TABLE II
BEAT-BY-BEAT, RECORD-BY-RECORD TESTING RESULTS OF THE EXPERIMENT. RESULTS FOR THE REAL-TIME CLASSIFICATION MACHINE: THE DETECTION STAGE (LABELS SB+ AND SB-), THE
SINGLE CLASSIFICATION STAGE AND THE GLOBAL CLASSIFICATION MACHINE (LABELS SB+ AND SB-)

File #	<i>Expert counts</i> (# beats)		<i>QRS Detector (single)</i>				<i>Classifier (single)</i>				<i>Classifier SB+ (global)</i>				<i>Classifier SB- (global)</i>			
			SB+		SB-		Normal		Abnormal		Normal		Abnormal		Normal		Abnormal	
	Normal	Abnorm	FP	FN	FP	FN	TP	FP	TP	FP	TP	FP	TP	FP	TP	FP	TP	FP
100	2271	1	0	0	0	0	2271	0	1	0	2271	0	1	0	2271	0	1	0
101	1863	2	6	2	6	2	1851	0	2	12	1847	0	1	15	1850	0	1	12
102	99	2088	0	0	0	0	7	1	2087	92	5	1	2087	94	5	1	2087	94
103	2084	0	0	1	0	4	2084	0	0	0	2083	0	0	0	2080	0	0	0
104	163	2065	124	13	85	18	135	120	1946	28	140	138	1916	22	141	135	1914	21
105	2526	46	77	17	61	16	2288	23	23	238	2283	20	25	227	2270	25	21	240
106	1507	520	0	3	0	79	1506	107	413	1	1505	117	400	2	1505	113	328	2
107	0	2136	1	1	1	3	0	3	2134	0	0	2	2134	0	0	2	2132	0
118	96	2182	4	0	4	0	1	15	2167	95	1	17	2165	95	1	17	2165	95
119	1543	444	0	0	0	0	1543	0	444	0	1543	0	444	0	1543	0	444	0
200	1773	828	15	4	8	4	1717	135	693	56	1724	127	698	48	1728	117	708	44
201	1654	308	0	29	0	69	1644	48	260	10	1654	58	222	0	1654	54	186	0
202	2097	39	0	5	0	15	2086	13	26	11	2087	11	24	9	2087	11	14	9
203	2529	451	57	30	34	48	935	70	381	1594	906	75	358	1611	895	75	343	1619
205	2574	82	0	3	0	5	2572	15	67	2	2572	15	66	0	2572	14	65	0
208	1585	1368	11	46	10	462	1547	110	1259	39	1546	116	1215	30	1542	118	801	30
209	3004	1	1	0	1	0	2996	0	1	8	2994	0	1	10	2993	0	1	11
210	2422	227	12	17	6	49	2362	71	156	61	2367	67	147	51	2369	64	117	50
212	923	1825	0	1	0	1	906	17	1808	17	901	16	1809	21	905	15	1810	17
213	2666	585	0	2	0	21	2665	249	336	1	2665	250	334	0	2665	252	313	0
214	0	2260	3	4	4	9	0	143	2118	0	0	209	2047	0	0	222	2030	0
215	3198	165	1	1	1	2	3196	31	134	2	3196	36	128	2	3196	36	127	2
217	244	1964	2	3	2	7	244	38	1926	0	243	36	1925	1	243	36	1921	1
219	2089	65	0	0	0	7	2086	10	55	3	2086	10	55	3	2086	10	50	1
tot	38910	19652	314	182	223	821	36642	1219	18437	2270	36619	1321	18202	2241	36601	1317	17579	2248