Autonomous Medical Monitoring and Diagnostics

AMIGO

D8 - Executive Summary

DOCUMENT N° : D8 - Executive Summary

ISSUE : 1.0

DATE : 01.09.2016

-

CSEM PROJECT N° : 221-ES.1577

CONTRACT N° : 4000113764 /15/F/MOS

	FUNCTION	NAME	SIGNATURE	DATE
Author	Expert	Ph. Renevey		01.09.2016
Author	Expert	A. Lemkaddem		01.09.2016
Reviewer	Project Manager	M. Lemay		01.09.2016

MODIFICATION LIST

ISSUE	DATE	PAGES	MODIFICATIONS
0.1	25.08.2016	All	First draft
1.0	01.09.2016	All	First complete version

DISTRIBUTION LIST

COMPANY	NAME	MEANS
ESA	DONATI A. MARTINEZ J. DAMANN V.	Email
CSEM	BERTSCHI M. LEMAY M. LEMKADEM A. RENEVEY Ph. TEODORISIS V.	Email

TABLE OF CONTENTS

1 SCOPE 7
2 ABBREVIATIONS 8
3 REFERENCES 9
INTRODUCTION 10
4 MEDICAL CASE 10
5 DEVELOPMENT 10
5.1 STRUCTURE 10
5.2 FEATURE EXTRACTION 11
5.2.1 Nonspecific 11
5.2.2 ECG-specific 11
5.2.3 Respiration-specific 12
5.3 ANOMALIES VERSUS EXTRACTED FEATURES 13
5.4 DATA-MINING TECHNIQUES 13
5.5 MODULE VALIDATION 14
6 PERFORMANCE 14
6.1 NOVELTY DETECTION 14
6.2 *K*-NEAREST NEIGHBOR 14
6.3 SUPPORT VECTOR MACHINE 15
6.4 GAUSSIAN MIXTURE MODEL 16
7 DISCUSSION 17
7.1 DATA-MINING APPLICABILITY 17
7.2 ESA TECHNIQUE APPLICABILITY 18
7.3 TECHNICAL AND MEDICAL VALIDATION 18
7.4 DATA-MINING PERFORMANCE 18
8 CONCLUSION 19

TABLE OF TABLES

Figure 1 - Overview of the proposed evaluation platform. 11
Figure 2 - Fiducial points defined on a two second RMS signal during atrial fibrillation. 12
Figure 3 - Raw 10-second real V1 lead signal during AF (a). V1 signal after preprocessing (b). Atrial activity (AA) estimated on lead V1 (c). AA in (c) amplified by a factor of 10 (d). 12

TABLE OF TABLES

Table 1 - Selected Medical Conditions. 10
Table 2 – Anomalies versus extracted features. 13
Table 3 - Summary of the investigated data mining schemes 13
Table 4: Performance of Novelty Detection. 14
Table 5: Two class *k*NN outcomes with the accuracy of cardiac beat and respiration event classification. 14
Table 6: The obtained confusion matrix of De Novo cardiac arrhythmias using the *k*NN algorithm. 15
Table 7: The obtained confusion matrix of sleep apnea syndrome using *k*NN algorithm. 15
Table 8: The accuracy of cardiac beat and respiration event classification with linear SVM method 15
Table 9: The obtained confusion matrix of De Novo cardiac arrhythmias using the SVM algorithm with a nonlinear method. 16
Table 10: The obtained confusion matrix of sleep apnea syndrome using the SVM Nonlinear algorithm. 16
Table 11: The accuracy of cardiac beat classification with GMM on MIT-BIH Arrhythmia Database. 16
Table 12: The accuracy of cardiac beat classification with GMM on four databases. 16
Table 13: The obtained confusion matrix of De Novo cardiac arrhythmias using the GMM when data from MIT-BIH Arrhythmia Database is used. 17
Table 14: The obtained confusion matrix of De Novo cardiac arrhythmias using the GMM when data from all four the databases are used. 17

1 SCOPE

The objective of this document is to provide a summarized description of the findings during the study untitled “Autonomous Medical Monitoring and Diagnostics - AMIGO”. After the introduction, the selected medical use case are stated in section 4. Thereafter, section 5 begins with a brief description of the structure, the features extracted and the chosen data-mining techniques selected. The results of the performances obtained with the different selected data-mining techniques are detailed in section 6. Section 7 includes a discussion of the benefits and limitation of the different data-mining techniques used, their performance including as well the ESA’s novelty detection approach. This document is finalized with a conclusion in section 8.

2 ABBREVIATIONS

AF	Atrial Fibrillation
AMIGO	Autonomous medical monitoring and diagnostics
ANN	Artificial neural network
BP	Blood pressure
EEG	Electroencephalogram
ECG	Electrocardiogram
EM	Expectation Maximization
ESA	European Space Agency
GMM	Gaussian Mixture Model
HMM	Hidden Markov Model
k-NN	k-nearest neighbors
NASA	National Aeronautics and Space Administration
RNN	Replicator neural network
SOW	State of Work
SVM	Support vector machine

3 REFERENCES

[1] International Space Exploration Coordination Group, "The Global Exploration Roadmap," 2013.

[2] M. Lemay, Data processing techniques for the characterization of atrial fibrillation, Lausanne: EPFL, 2007.

[3] P. Renevey, M. Proença, O. Grossenbacher and J.-A. F. D. Porchet, "Breath rate and SpO2 update," LTMS-S ESA project - D15, 2015.

[4] European Parliament, "Directive 95/46/EC of the European Aprliament and the Council," 1995.

4 INTRODUCTION

Global plans for exploratory missions aim at extending the distances travelled by humans well beyond low Earth orbit and establishing permanent bases on the surface of Moon and Mars [1]. This will inevitably lead to increasing mission duration, radiation intensity, gravity levels and degree of confinement and isolation to which the crews will be exposed. In this extended space mission context, the astronauts should have the means to collect medical/physiological data in order to understand if their physiological conditions are in nominal levels. The astronauts should be informed about possible diagnoses and get practical recommendations about treatment options in order to deal with medical issues with limited or no ground interactions. The mentioned scenario is possible if and only if the astronauts and medical crewmembers have access to an autonomous medical monitoring system with embedded diagnostic algorithms. AMIGO proposes to evaluate if and how data mining can be of benefit for an autonomous medical monitoring/diagnostic system. The following sections will present the target objectives, the definition of medical use cases, the development and validation of the proposed approach a short discussion will conclude this report.

5 MEDICAL CASE

Based on the high probability of occurrence, coherence with actual and realistic medical capabilities, the two following medical conditions have been chosen to be studied:

(1) **De Novo cardiac arrhythmia** and more precisely cardiac AF access; and
(2) **Sleep apnea syndrome** (SAS) which causes urinary disorder during the first three days of the flight.

Table 1 - Selected Medical Conditions.

Medical condition inducing care needs (medical and/or surgical)	**Best estimated occurrence rate / (person × year)**	
	During transfer	During planetary/asteroid surface or orbit activities
De Novo cardiac arrhythmia	3,000000	0,150000 or 3,000 if in orbit
Urinary disorder during sleep apnea syndrome(SAS) (incl. side effect SAS medication)	10% of astronaut during 1 – 3 days at gravity changes	10% of astronaut during 1 to 3 days at gravity changes

6 DEVELOPMENT

The objective of this section is to describe AMIGO's solution which has been developed and used during the lifetime of the project.

6.1 Structure

AMIGO has implemented the evaluation platform shown in Figure 1. The main output of the pipeline is in underlined bold text.

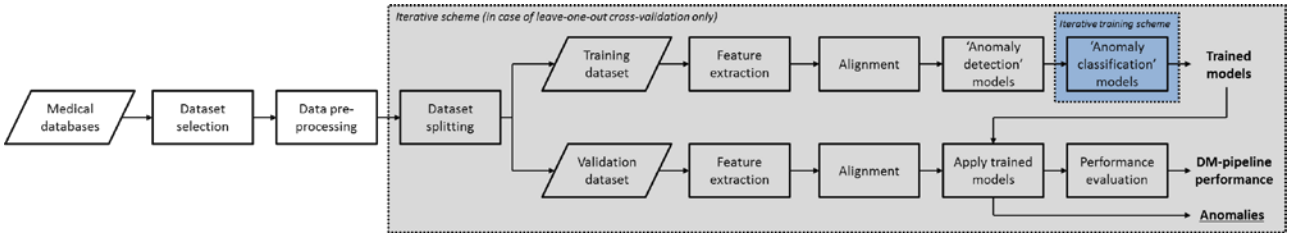

Figure 1 - Overview of the proposed evaluation platform.

In short, the general working principle of the pipeline can be explained as follows. In the training mode, the "Dataset selection" block selects available medical databases, depending on the medical use case considered and the validation mode. All selected signals and associated annotated anomalies obtained from the databases (ground truth) are loaded into the algorithm pipeline. The signals are pre-processed (synchronized, possibly filtered and/or normalized) in the "Data pre-processing" block. The "Dataset splitting" block then separates all data into two subsets: the training dataset and the validation dataset. The "Feature extraction" block is in charge of extracting generic and signal-specific features from the signals. During the training phase of the algorithm pipeline (upper row of the block diagram in Figure 1), the features extracted are aligned with annotated anomalies ("Alignment" block) used to train several 'Anomaly detection' models ("'Anomaly detection' models" block). These models aim at classifying the values of the features extracted in the "Feature extraction" block either as "normal" or "abnormal". When at least one abnormal feature value is encountered, an anomaly is detected. Being able to classify which type of anomaly has been detected based on which type of features have been classified as abnormal is the task of the 'Anomaly classification' models ("'Anomaly classification' models" block). Also, some models ('Anomaly detection and classification' models) will perform both tasks (anomaly detection *and* classification) jointly. The trained 'Anomaly detection', 'Anomaly classification' and 'Anomaly detection and classification' models are then applied to unseen data (validation dataset) during the validation phase of the algorithm pipeline (lower row of the block diagram) in the "Apply trained models" block. The resulting detected and identified anomalies form the main output of the DM-pipeline. Lastly, the performance of the models in terms of detection and classification performance is assessed in the "Performance evaluation" block.

6.2 Feature extraction

This section describes briefly the list of some features which are computed by the AMIGO's solution.

6.2.1 Nonspecific

Nonspecific features are features that are not dedicated to a specific type of data. They can be extracted from any type of signal. The signal can even be a feature itself (e.g. the RR intervals of an ECG). It must be noted that all these features are dependent on the duration of the window of analysis during which they are calculated. It includes mean value, standard deviation, skewness, kurtosis, extrema, quartiles, power of the signal, and average value of the first derivative with respect to time.

6.2.2 ECG-specific

When the source of the data is associated to ECG, the following feature extraction and respective processing is applied.

6.2.2.1 Fiducial point detection and baseline correction

First, the fiducial points, that is, the timing of the onset of the ventricular depolarization, the timing of R waves, the timing of J points and T wave peaks, are derived from the root mean squared signal of multiple ECG signals.

Figure 2 displays the result of the pre-processing and fiducial point detection process. In this same figure, the timing of the onset of the ventricular depolarization, the R wave timing, the J point and T wave peak timings are denoted as q_i, r_i, and k_i, respectively. Locations of QRS complex (A_i), JQ interval (B_i) and cardiac cycle (C_i) are also indicated.

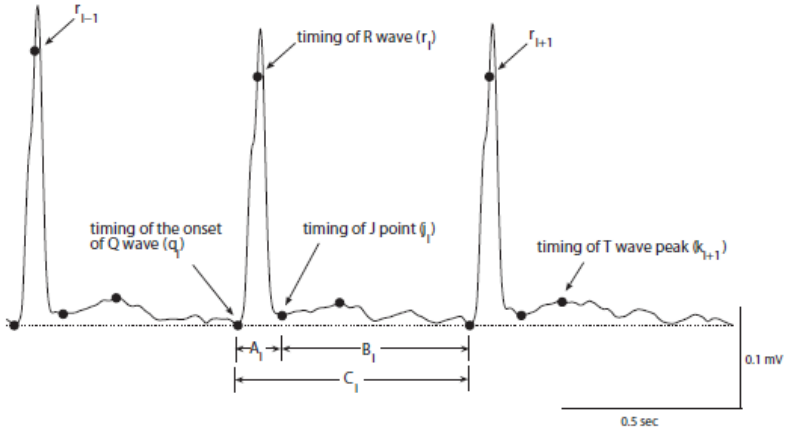

Figure 2 - Fiducial points defined on a two second RMS signal during atrial fibrillation.

6.2.2.2 ECG source separation

In the last decades, numerous studies have separated the atrial and ventricular components on ECG signals to perform arrhythmia detection and classification [2]. This source separation is extremely useful to detect and classify various types of arrhythmia, namely the atrial fibrillation and atrial ectopic beats. The approaches based on averaged beat subtraction are the most used technique for the separate atrial and ventricular activities. A version of such averaged beat subtraction algorithm has been developed for AMIGO's feature extraction solution. Figure 3 displays the results of this source separation on a typical ECG segment.

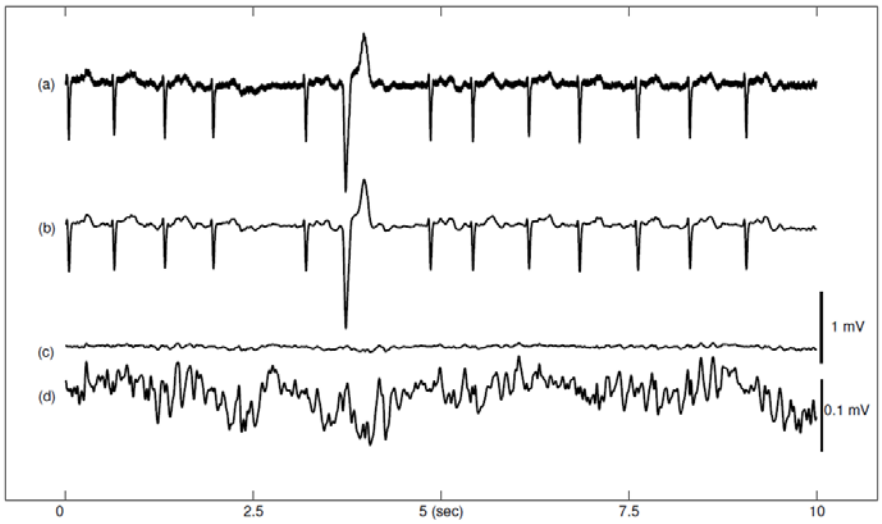

Figure 3 - Raw 10-second real V1 lead signal during AF (a). V1 signal after preprocessing (b). Atrial activity (AA) estimated on lead V1 (c). AA in (c) amplified by a factor of 10 (d).

6.2.3 Respiration-specific

The algorithm used to estimate the respiratory rate [3] can be summarized by the following steps. The measured respiratory signal is filtered in order to restrict the signal bandwidth to possible values of breathing frequencies. Then, a bandpass filter of the fifth order with lower and upper cutoff frequencies at 1/15 Hz and 1 Hz respectively is applied to restrain the frequency domain to possible

respiration rate. These selected respiration boundaries correspond to respiration rate between four and sixty cycles per minute.

6.3 Anomalies versus extracted features

Table 2 justifies the use of the proposed extracted features with respect to each medical anomalies included in the proposed datasets. This justification is based on the literature.

Table 2 – Anomalies versus extracted features.

Anomaly (medical condition) vs extracted features	PR interval	QRS interval	ST segment	RR interval	PP intervals	AA signals (non-specific features)	VA signals (non-specific features)	Eigenvalues of C	Blood pressure-based features	Respiration-based features	EEG-based features
Atrial fibrillation	X			X	X	X		X	X		
Supraventricular ectopic	X				X	X					
Ventricular ectopic				X	X	X	X	X			
Paced beats											
Ventricular tachycardia				X			X	X	X		
Ventricular flutter	X			X			X	X			
Ventricular fibrillation				X				X	X		
Supraventricular tachycardia					X	X					
Ventricular bigeminy		X	X	X							
Ventricular trigeminy		X	X	X							
Idioventricular rhythm		X		X		X	X	X			
Atrial bigeminy	X			X	X						
Sinus bradycardia					X	X					
Sleep apnea syndrome			X	X					X	X	X

6.4 Data-mining techniques

After an extensive review of different classification methods, the linear classifiers are the preferred methods with the biggest advantage that they only need a small dataset for training the classifiers. The methods that were implemented in AMIGO's solution are for the processes or blocks called 'Anomaly detection' and "Anomaly classification" listed in Table 3.

Table 3 - Summary of the investigated data mining schemes

Possible detection & classification scheme	Anomaly detection	Anomaly classification
Scenario 1	Novelty detection models	K nearest neighbors (k-NN)
Scenario 2	K nearest neighbors (*k*NN)	
Scenario 3	Support Vector Machine (SVM)	
Scenario 4	Gaussian Mixture Model (GMM)	

6.5 Module Validation

Each of the process (block) of the evaluation platform showed in Figure 1 has been validated. A combination of a running script without bugs and visual inspection of the performance were sufficient to ensure an **accepted** validation. See section 6 of D5 - Assessment Report for more details.

7 PERFORMANCE

7.1 Novelty detection

As a two-class classification method, the performance of the Novelty Detection algorithm can be evaluated in terms of sensitivity (true positive rate) and specificity (true negative rate). Sensitivity (SE) and specificity (SP) are calculated as:

$$SE = TP / (TP+FN) \quad \text{and} \quad SP = TN / (TN+FP),$$

where

TP	= true positives	= abnormal instances correctly identified as abnormal;
FP	= false positives	= normal instances incorrectly identified as abnormal;
TN	= true negatives	= normal instances correctly identified as normal;
FN	= false negatives	= abnormal instances incorrectly identified as normal.

The closer SE and SP are to 0, the worst the performance. The closer SE and SP are to 1, the best the performance. As both SE and SP are important in binary classification, it was proposed in D2 to use the harmonic mean of both quantities as a global performance metric.

$$HM = 2*SE*SP / (SE+SP)$$

The closer HM is to 0, the worst the performance. The closer HM is to 1, the best the performance. When running the Novelty Detection algorithm on the testing dataset of the De Novo cardiac arrhythmias, the results showed in Table 4 were obtained.

Table 4: Performance of Novelty Detection.

De Novo cardiac arrhythmias	Sleep apnea syndrome
SE = 0.0957	SE = 0.0097
SP = 0.8440	SP = 0.9825
HM = 0.1720	**HM = 0.0193**

7.2 *k*-Nearest Neighbor

With the parameters selected, the *k*NN algorithm was used to process both the data from the MIT-BIH Arrhythmia Database and the MIT-BIH Polysomnographic Database. A binary classification was performed to separate either the normal from the abnormal beats or normal from the abnormal respirations. Table 5 demonstrates the outcome of *k*NN in terms of SE, SP and HM.

Table 5: Two class *k*NN outcomes with the accuracy of cardiac beat and respiration event classification.

De Novo cardiac arrhythmias	Sleep apnea syndrome
SE = 0.8711	SE = 0.4905
SP = 0.7575	SP = 0.4641
HM = 0.8103	**HM = 0.4769**

Table 6 show the ratio of observed beats known to be in a certain class but predicted to be in another. Each row in Table 6 represent the known ration of beats while the columns correspond to the predicted beats by the *k*NN algorithm. If we consider the first row for instance, **74%** of the **Normal** beats were classified as normal, **2.14%** were classified as premature ventricular contraction, **3.72%** were classified as ventricular bigeminy, **2%** were classified as ventricular trigemini and finally **18%** were predicted to be atrial fibrillation.

Table 6: The obtained confusion matrix of De Novo cardiac arrhythmias using the *k*NN algorithm.

	Predicted Class				
Known Class	**Normal**	**Premature ventricular contraction**	**Ventricular bigeminy**	**Ventricular trigeminy**	**Atrial fibrillation**
Normal	**0.7408**	0.0214	0.0372	0.0199	0.1807
Premature ventricular contraction	0.0295	**0.6605**	0.0355	0.0123	0.2623
Ventricular bigeminy	0.0205	0.1087	**0.4437**	0.0205	0.4066
Ventricular trigeminy	0.0961	0.0519	0.3481	**0.0338**	0.4701
Atrial fibrillation	0.1561	0.0235	0.0166	0.0205	**0.7834**

Considering the classification of the sleep apnea syndromes, Table 7 demonstrate the obtained results. In this case, the results of the *k*NN algorithm shows to be less performant compared to the outcomes of Table 6.

Table 7: The obtained confusion matrix of sleep apnea syndrome using *k*NN algorithm.

	Predicted Class						
Known Class	**Normal**	**Hypopnea**	**Hypopnea with arousal**	**Obstructive apnea**	**Obstructive apnea with arousal**	**Central apnea**	**Central apnea with arousal**
Normal	**0.4238**	0.1523	0.2706	0.0082	0.0913	0.0140	0.0398
Hypopnea	0.3916	**0.1871**	0.1156	0.0335	0.0198	0.0662	0.1863
Hypopnea with arousal	0.4352	0.2020	**0.2015**	0.0230	0.0660	0.0313	0.0411
Obstructive apnea	0.4666	0.1436	0.0782	**0.0408**	0.081	0.0752	0.1146
Obstructive apnea with arousal	0.4712	0.1529	0.1532	0.0232	**0.0953**	0.0457	0.0584
Central apnea	0.4973	0.108	0.2531	0.0142	0.0248	**0.0389**	0.0637
Central apnea with arousal	0.3855	0.0870	0.2960	0.0318	0.0761	0.0477	**0.0761**

7.3 Support Vector Machine

The results from the SVM linear method are shown in Table 8. In the case of the cardiac arrhythmias, the performance scores showed promising results compared to the sleep apnea syndrome.

Table 8: The accuracy of cardiac beat and respiration event classification with linear SVM method

De Novo cardiac arrhythmias	**Sleep apnea syndrome**
SE = 0.8890	SE = 0.4536
SP = 0.9535	SP = 0.6664
HM = 0.9201	**HM = 0.5398**

The SVM implementation in Matlab is a binary classifier and can therefore not handle multiple classes at ones. Therefore, to obtain these results, SVM was executed for each pair of classes. For testing five different classes, ten different SVM models were generated. Thereafter, every observation was tested using all the ten models. The predicted class that appeared most for one instance was selected. A confusion matrix was generated to illustrate the results of the nonlinear SVM, Table 9.

Table 9: The obtained confusion matrix of De Novo cardiac arrhythmias using the SVM algorithm with a nonlinear method

	Predicted Class				
Known Class	**Normal**	**Premature ventricular contraction**	**Ventricular bigeminy**	**Ventricular trigeminy**	**Atrial fibrillation**
Normal	**0.8323**	0.0564	0.0003	0.0001	0.1109
Premature ventricular contraction	0.0228	**0.7795**	0.0028	0.0000	0.1949
Ventricular bigeminy	0.0217	0.165	**0.2455**	0.0000	0.5678
Ventricular trigeminy	0.0883	0.2052	0.1247	**0.0000**	0.5818
Atrial fibrillation	0.2687	0.0146	0.0018	0.0000	**0.7149**

The same procedure was applied to the sleep apnea syndromes, Table 10, to construct the confusion matrix with nonlinear method.

Table 10: The obtained confusion matrix of sleep apnea syndrome using the SVM Nonlinear algorithm

	Predicted Class						
Known Class	**Normal**	**Hypopnea**	**Hypopnea with arousal**	**Obstructive apnea**	**Obstructive apnea with arousal**	**Central apnea**	**Central apnea with arousal**
Normal	**0.9297**	0.0001	0.0319	0.0000	0.0377	0.0000	0.0006
Hypopnea	0.7095	**0.0000**	0.0700	0.0000	0.2205	0.0000	0.0000
Hypopnea with arousal	0.6660	0.0000	**0.1516**	0.0000	0.1824	0.0000	0.0000
Obstructive apnea	0.5290	0.0000	0.0280	**0.0000**	0.4426	0.0000	0.0004
Obstructive apnea with arousal	0.6374	0.0000	0.1226	0.0000	**0.2399**	0.0000	0.0001
Central apnea	0.5487	0.0000	0.2619	0.0000	0.1876	**0.0000**	0.0018
Central apnea with arousal	0.4540	0.0000	0.3010	0.0000	0.2441	0.0000	**0.0008**

7.4 Gaussian Mixture Model

The level of sensitivity and specificity obtained with GMM classifier on De Novo cardiac arrhythmias included in the MIT-BIH Arrhythmia Database is illustrated in Table 11. The GMM classifier was also tested on all databases which includes De Novo cardiac arrhythmias and the results are showed in Table 12. These results show a decrease in performance for the GMM, but this is not surprising since some of the data is rather challenging due to amount of noise they contain.

Table 11: The accuracy of cardiac beat classification with GMM on MIT-BIH Arrhythmia Database.

De Novo cardiac arrhythmias
SE = 0.8923
SP = 0.7483
HM = 0.8140

Table 12: The accuracy of cardiac beat classification with GMM on four databases.

De Novo cardiac arrhythmias
SE = 0.4795
SP = 0.4724
HM = 0.4759

Table 13 shows the confusion matrix of **De Novo cardiac arrhythmias** when GMM is utilized. Compared to the results from the *k*NN and SVM classifier, the scores from the GMM are slightly

lower. However, same trends of the results are demonstrated here where the classification of normal beats are easier and the most difficult ones are the ventricular trigeminy.

Table 13: The obtained confusion matrix of De Novo cardiac arrhythmias using the GMM when data from MIT-BIH Arrhythmia Database is used.

	Predicted Class				
Known Class	**Normal**	**Premature ventricular contraction**	**Ventricular bigeminy**	**Ventricular trigeminy**	**Atrial fibrillation**
Normal	**0.7930**	0.0799	0.0262	0.0002	0.1007
Premature ventricular contraction	0.0039	**0.6784**	0.2258	0.0000	0.092
Ventricular bigeminy	0.0064	0.6138	**0.2941**	0.0000	0.0857
Ventricular trigeminy	0.0000	0.8182	0.0987	**0.0000**	0.0831
Atrial fibrillation	0.3261	0.0461	0.2705	0.0000	**0.3574**

Table 14 show the confusion matrix when all four are used. The decrease in performance was expected and can be explained by the large amount of noise in the database MIT-BIH Noise Stress Database.

Table 14: The obtained confusion matrix of De Novo cardiac arrhythmias using the GMM when data from all four the databases are used.

	Predicted Class				
Known Class	**Normal**	**Premature ventricular contraction**	**Ventricular bigeminy**	**Ventricular trigeminy**	**Atrial fibrillation**
Normal	**0.2290**	0.1491	0.2196	0.1803	0.2221
Premature ventricular contraction	0.0161	**0.2404**	0.4556	0.2004	0.0876
Ventricular bigeminy	0.0464	0.0455	**0.4389**	0.4345	0.0347
Ventricular trigeminy	0.0289	0.1350	0.3239	**0.4981**	0.0141
Atrial fibrillation	0.0425	0.2319	0.385	0.2717	**0.0690**

8 DISCUSSION

8.1 Data-mining applicability

The choice of the selected classifiers that are presented here are based on the fact that the available amount of data is relatively small. This eliminates already all the network-based methods such as the ANN and RNN for instance. These methods need an excessive amount of training data to structure the optimal network. On the other hand side, the other extreme (simple classifiers such as linear classifiers) perform surprisingly well on different classifications tasks even if the amount of training data is small. SVM for instance, does not only have the advantage of small training set requirement but as well a flexible nature. It includes regularization parameter, which gives us the possibility to avoid over-fitting. The use of the kernel trick allows us to build in expert knowledge about the problem by engineering this kernel. The choice of the kernel is however not always obvious.

On the other hand, the big advantage of the *k*NN algorithm is that it does not require any parameter training (although its main parameter “k” is often trained via cross-validation). In addition the *k*NN has the advantage of being a very simple and yet a very powerful algorithm. However, depending

on the size of the data, *k*NN may become very heavy both in terms of computation and data storage, as all training data need to be stored and may have a high dimensionality.

Finally, GMMs have the positive aspect of being flexible and manage to model any distribution given, even though they can require a higher number of training data compared to linear classifiers. Its flexibility is an important factor for us since the distribution of our features are still unknown. In addition, our features come with some prior information (*e.g.* confidence indexes) that we can easily exploit to initialize the EM algorithm (used by the GMMs) and therefore avoid overfitting the model.

8.2 ESA technique applicability

As defined in the document, the Novelty Detection technique has been integrated into AMIGO's test workbench as an Anomaly detector. This technique has the advantages of having a user-friendly interface, being accessible, being ready-to-use and providing flexible solution in terms of input parameters.

8.3 Technical and medical validation

By applying the different technical validation steps for each feature and each performance evaluation mode, AMIGO's test workbench solution proposes a multi-level technical validation. Concerning the extracted features, their implementation is crucial to the performance of the final solution. Even the optimal classifier is not able to separate groups when using irrelevant features. Their medical significance has been justified with numerous studies. Moreover, Table 2 cross-checks the presence of "at least" one medically-justified feature for each of the anomalies included in the entire dataset.

8.4 Data-mining performance

Despite the challenge we faced to classify various medical anomalies of different origins or medical causes, the derived results mentioned in this report indicate rather good performances globally.

Concerning the anomaly detection (two classes), the highest sensitivity (0.90) was obtained on test set for De Novo cardiac arrhythmias when nonlinear SVM approach was applied to separate normal versus abnormal classes. The linear SVM on the other hand indicated a higher specificity (0.95) and a harmonic mean (0.92) with a respectable sensitivity (0.89). As for the *k*NN and GMM, these methods are not very far behind in terms of performance, both methods scored a harmonic mean of 0.81 and showed high sensitivity (0.87 and 0.89, respectively) but a slightly lower specificity (0.76 and 0.75, respectively). Note also that, on these dataset, the Novelty Detection approach obtained the lowest performance in terms of sensitivity (0.10), specificity (0.84) and harmonic mean (0.17). This performance probably highlight the limitation of this approach when using dedicated and specific features. This fully generic specifications might be both advantageous and disadvantageous depending on the training context and final application.

Concerning the anomaly classification (multiple classes), the best performance was obtained on test set for De Novo cardiac arrhythmias including atrial fibrillation, premature ventricular contraction, ventricular bigeminy and trigeminy when linear SVM approach was used. AMIGO's best performance is quantified by good classification of normal beats at 94%, premature ventricular contraction at 76%, ventricular bigeminy at 37%, ventricular trigeminy at 1% and atrial fibrillation at 72%.

Compared to the Sleep Apnea Syndromes, the results are significantly worse and this is due to several aspects. First, the number of features used to describe different apnea syndrome are too high compared to the number of observations provided. Secondly, the importance of selecting the proper features describing the anomalies is highlighted by the good performance on the De Novo cardiac arrhythmias where the CSEM expertise has played a major role. On the other hand, the low accuracy when considering the Sleep apnea syndrome is mainly due to the features that were

extracted. These features, in particular EEG and BP related once, were not well selected to describe the apnea syndromes and a future work would be to find the proper features. Thirdly, the annotation in the Physionet database for the different apnea syndrome is not very straight forward as for the beats (i.e. one annotation point to describe an apnea syndrome instead of a window). Some tweaking was necessary to generate windows of apnea syndromes and therefore, more efforts can be done to develop this part further to improve the results. Finally, the selected parameters for each classification method was based on the data from the MIT-BIH Arrhythmia Database. A cross-validation to find the proper parameters for the sleep database needs to be performed. This is particularly important in the case of SVM method since the distribution of the data needs to be known *a priori*.

9 CONCLUSION

The main core of AMIGO is the study of methodology to implement data mining techniques for the development of an automatic diagnosis system. Given this objective and the limited duration of the project, the creation of a finalized system with optimal performances was out-of-the-scope of the project. However a complete system has been implemented and a proof-of-concept established. Moreover, high level of detection and classification has been obtained in the De Novo cardiac arrhythmia medical use case, including AF detection and classification. The large variety of performance between cardiac beat and sleep apnea classification highlights the importance of robust and dedicated features. The proposed methodology can be re-used to improve the performances of the actual system and to extend its detection capabilities to other signals and pathologies.

During the development of the current system several difficulties have been encountered. In the two selected medical use cases the detection is based on events (abnormal cardiac beats or abnormal respiration events). The synchronization of the reference with these events can be problematic for the training and the testing of the system as much as the meaning of quantified performance. In order to ensure an efficient classification development, this problem has to be taken into account from the beginning. Our work has also highlighted the following problem. The development of a generic system to classify abnormal situations requires that the input features has a similar structure including the sampling frequency, independently of the input signals and of the detection tasks. In the two proposed medical use cases the correct detection of abnormal event is not equally related to the context of the event. In the problem of the detection of abnormal cardiac beats, the ECG measurement of each beat contains relevant information for the detection and classification. In opposition, for the detection of breathing problem, the context (variation of the respiration pattern over time) is of tremendous importance for the correct detection and classification of abnormal situations. The generic methodology presented in this project has to be extended further to allow the entry of the global context in the decision process.

The proposed methodology and the developed system represent a solid starting point for the development of a more realistic system. The expertise gained during this project is certainly an asset for a more ambitious development.