

# **AUTONOMOUS MEDICAL MONITORING AND DIAGNOSTICS**

## **EXECUTIVE SUMMARY**

**DELIVERY 4.3 (D8)**

*Client:* ESA-ESOC

*Document N°:* 710008-REP-WP4-T007 – Rev.0

*Date:* 04/03/2016

*Authors:*  
Ing. Jacopo Biancat

*Checked:*  
Chiara Brightenti

Copy    1 of    1  
Pages: 28

*Authorised:*  
Attilio Brightenti

*Jacopo Biancat*

*Chiara Brightenti*

*Attilio Brightenti*

**DISTRIBUTION:**  
ESA-ESOC

**COPYRIGHT & CONFIDENTIALITY**

This document does not contain any copyright nor dissemination restrictions.

**REVISION SHEET**

Sheet	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Rev.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Sheet	16	17	18	19	20	21	22	23	24	25	26	27	28		
Rev.	0	0	0	0	0	0	0	0	0	0	0	0	0		

**Document history**

Rev.	Date of issue	Description
0	04/03/2016	Delivery 4.3

Changes other than typing/grammar proofing, brought with Rev. x, are marked by the following annotation on the right margin and shaded in **yellow**.

Rev. x

INDEX

1	SCOPE .....	4
2	EXECUTIVE SUMMARY .....	5
2.1	Introduction .....	5
2.2	ECG Data.....	6
2.2.1	Dataset .....	6
2.2.2	Use case .....	7
2.2.3	Synthetic Results .....	9
2.2.4	Example of unsupervised method - Orientation of the R wave .....	11
2.2.5	Example of supervised method - Duration of the QRS complex .....	12
2.2.6	Medical validation and autonomy .....	16
2.3	Laboratory data .....	17
2.3.1	Dataset .....	17
2.3.2	Use case .....	18
2.3.3	Synthetic Results .....	19
2.3.4	Example of unsupervised method .....	20
2.3.5	Example of previously unknown correlation .....	23
2.3.6	Medical validation and autonomy .....	23
2.4	Applicability of <i>Novelty Detection</i> .....	24
2.4.1	ECG .....	25
2.4.2	Laboratory.....	25
2.5	Conclusions .....	25
2.6	Suggested future developments .....	27

## 1 SCOPE

This document is the executive summary of the project under contract N° 4000113793/15/F/MOS between S.A.T.E. and ESA regarding the "AUTONOMOUS MEDICAL MONITORING AND DIAGNOSTICS".

This document represents Delivery 4.3 of the Contract, issued in compliance with the revised planning agreed with ESA-ESOC.

## 2 EXECUTIVE SUMMARY

### 2.1 Introduction

The objective of the project was to define, design and validate data mining algorithms that enable advanced or autonomous medical monitoring and diagnostics for astronauts, exploiting standard medical data available from health and sanitary structures.

S.A.T.E. proposed as two broad use cases the analysis of ECG and laboratory data, which are two of the most important medical exams used by physicians nowadays since the former can provide a great deal of information on the normal and pathological physiology of heart activity<sup>1</sup> and the latter provides a broad range of diagnostic information to assess the status of the main organs (e.g. liver, kidney, blood, etc.) influencing more than 70% of health care decisions. Moreover, given their importance in standard medical practice, a significant amount of data related to general patients can be found, which are necessary to adopt a data driven approach as foreseen in this project. Finally, the analysis of the heart activity through the ECG and the analysis of blood and urine allow the diagnosis of the widest range of pathologies and diseases related to the possible in-flight scenarios.

The medical datasets made available by the subcontractor, AUSLMO, consist in a set of analyses collected in the period spanning the period from 2009 to 2014.

Table 3 shows an overview of the amount of ECG and Laboratory data, including an overview of the number of patients attending each type of analysis.

Data Type	Number of tests <sup>2</sup>	Number of patients
ECG Data	62.244	41.056
Laboratory Data	87.273.621	657.918

Table 1 - Total amount of ECGs and Laboratory Data tests and number of patients for each type of analysis.

The medical datasets have been pre-processed by the subcontractor, in order to delete every reference to patient's name, surname and address, replacing that information with an ID, unique for each patient. In this way, datasets have been anonymized, while maintaining their uniqueness and the correspondence between analyses performed in different moments or of different types by the same patient.

Other sensitive data that have been deleted are the name and the surname of the physician performing the diagnosis. Also in this case, a specific ID has been used to refer univocally to each physician.

During the initial phase of the project the scope of the two broad use cases have been refined into two specific use cases which were selected and agreed with ESA:

- Use Case 1 (ECG/Feature) aiming at the development of a classification algorithm based on the features computed automatically by the MUSE system<sup>3</sup>.

The scope of this use case has been expanded during the execution of the project, including also the automatic evaluation of the ECG waveforms (named use case 1.B: ECG/FEATURES/WAVEFORMS), besides the evaluation of the medical features (named use case 1.A: ECG/FEATURES/FEATURES).

<sup>1</sup> It is worth highlighting the fact that the diagnosis of one important cardiovascular diseases, the arrhythmias, cannot be replaced by any other examination.

<sup>2</sup> For "test" it is intended an individual measurement referred to also as *work order* in the following.

<sup>3</sup> MUSE is a medical certified tool used as Cardiology Information System.

- Use Case 4<sup>4</sup> (LAB/Clustering) aiming at the development of a clustering algorithm based on blood and urine analysis without the use of any a priori-knowledge.

To carry out the scope of work both supervised and unsupervised methods were investigated, including Self-Organizing Maps, K-means, Gaussian-mixture models, Adaptive Boosting, Decision trees and Pattern nets<sup>5</sup>. Furthermore for ECG data several pre-processing steps were deemed necessary in order to improve the data used as input to the automatic classification and clustering algorithms. Moreover also cross-correlation analyses were performed both for ECG and laboratory data. Finally the applicability of the ESA *Novelty Detection* software to biomedical data was also investigated.

## 2.2 ECG Data

### 2.2.1 Dataset

The ECG data have been acquired with patients in the resting position using the medical certified tool MUSE Cardiology Information System. This system grants the requirements set forth for an ECG to be clinically valid.

The *clinical validity* of the data acquired by the MUSE system allows the use of such data by the physicians to perform a diagnosis which will in turn be *clinically valid*. This diagnosis with clinical validity is contained in the field “Diagnosis” of data.

Unfortunately this diagnosis have no canonical form, although a standard protocol is used to write it down.

This makes difficult exploiting only the “Diagnosis” field to build meaningful classes with the scope of applying supervised techniques. Indeed the application of supervised techniques require the use of a priori knowledge that allows labelling data into predefined categories aiming at building a function that assign objects to one of those predefined categories.

To do so it was agreed to use the so-called *quantitative* approach that is based on the evaluation of synthetic features extracted by the ECG traces and that are made available by the MUSE system.

Table 1 provides the list of the medical features extracted from the ECG by the MUSE system.

Feature name	Description	Units
SystolicBP	Systolic blood pressure	mmHg
DiastolicBP	Diastolic blood pressure	mmHg
VentricularRate	Ventricular rate	BPM
AtrialRate	Atrial rate	BPM
PRInterval	P – R interval	ms
QRSDuration	QRS duration	ms
QTInterval	QT interval	ms
QTCorrected	Bazett's Algorithm	ms
PAxis	P axis	degrees
RAxis	R axis	degrees
Taxis	T axis	degrees
QRSCount	QRS count	--
Qonset	Q onset	ms
Qoffset	Q offset	ms
Ponset	P onset	ms
Poffset	P offset	ms

<sup>4</sup> The number *four* assigned to this use case is due to the fact that the two selected use cases were selected among a short-list of possible use cases defined during the first phase of the project on the basis of the available dataset.

<sup>5</sup> *Pattern nets* are pattern recognition networks based on neural networks made available by the Mathworks® in the Neural Network Toolbox.

Feature name	Description	Units
Toffset	T offset	ms
ECGSampleBase	ECG sample rate base	--
ECGSampleExponent	ECG sample rate exponent	--
QTcFrederica	QT calculated with the Frederica Algorithm	ms

Table 1 - Medical features extracted from the ECG by the MUSE system.

The criteria that were proposed and agreed with ESA for the classification of the ECG data are:

- *Rhythm*;
- *Ventricular frequency*;
- *Orientation of the R wave*;
- *Duration of the QRS complex*;
- *Duration of the PR interval*;
- *Duration of the corrected QT interval*.

Each of these criteria identify a different number of classes to which an ECG may be assigned, i.e. each ECG may be labelled on the basis of its belonging to a given class, making feasible the application of supervised learning methods.

Table 2 lists the classes identified by each criteria. Each ECG will belong only to one class of each criteria. For example a normal ECG will belong to the class "Sinusal" of the criterion *Rhythm*, to the class "Normal" of the criteria *Ventricular frequency*, *Orientation of the R wave*, *Duration of the QRS complex* and *Duration of the PR interval* and to the class "Not pathological" of the criterion *Duration of the corrected QT interval*.

It is highlighted that, in case of success, the supervised algorithms will allow performing automatically the diagnosis associated to each of the criteria defined above.

Criteria	Class 1	Class 2	Class 3	Class 4
<b>Rhythm</b>	Sinusal	Not sinusal		
<b>Ventricular frequency</b>	Normal	Tachycardia	Bradycardia	
<b>Orientation of the R wave</b>	Normal	Right axis deviation	Left axis deviation	Extreme right axis deviation
<b>Duration of the QRS complex</b>	Normal	Incomplete bundle branch block	Bundle branch block	Pathological
<b>Duration of the PR interval</b>	Normal	Not normal		
<b>Duration of the corrected QT interval</b>	Not pathological	Pathological		

Table 2 - Classes defined for each criteria.

### 2.2.2 Use case

The analysis of the ECG data included the following use case sub-branches or variants:

- Use case 1.A: ECG/FEATURES/FEATURES (evaluation of medical features). For simplicity this use case will be named throughout the document "1.A ECG/FEATURES".
- Use case 1.B: ECG/FEATURES/WAVEFORMS (evaluation of waveforms). For simplicity this use case will be named throughout the document "1.B ECG/WAVEFORMS".



For both use cases supervised and unsupervised methods have been investigated. *Supervised methods* exploit, during the learning phase referred to also as *training*, the available information on the class labels where the classes are known in advance and are those listed in Table 2. On the other side, *unsupervised methods* aim at dividing data into groups (clusters) that are meaningful, without exploiting a-priori knowledge. Therefore the information on the classes are not used during the learning phase but they may be used to evaluate the goodness of the achieved results, i.e. to verify if any of the group automatically identified may be associated to one or more classes defined in advance.

The application of supervised and unsupervised methods aimed at the generation of one or more classification model(s) for each agreed criterion.

The purpose is to train a model that will be capable of automatically assign to every new set of input data a class/diagnosis (associated to the selected criterion, Figure 1).

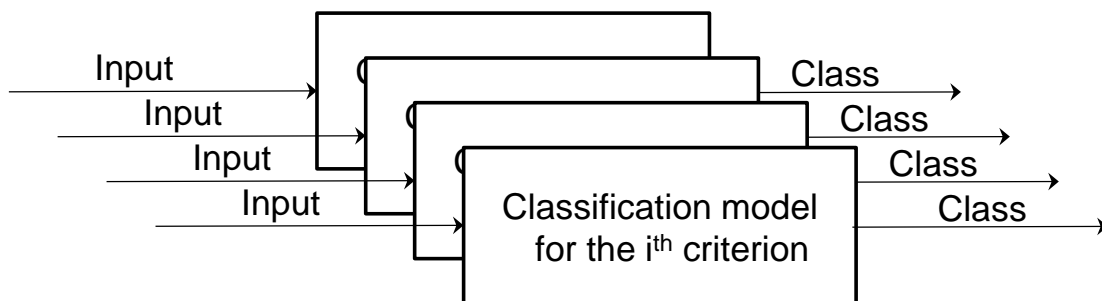


Figure 1- Proposed approach for use case 1.A ECG/FEATURES and use case 1.B ECG/WAVEFORMS.

Use case 1.A ECG/FEATURES and 1.B ECG/WAVEFORMS differ in the type of input: the former (use case 1.A) uses as input a combination of medical features instead the latter uses as input the entire ECG waveform (Figure 2).

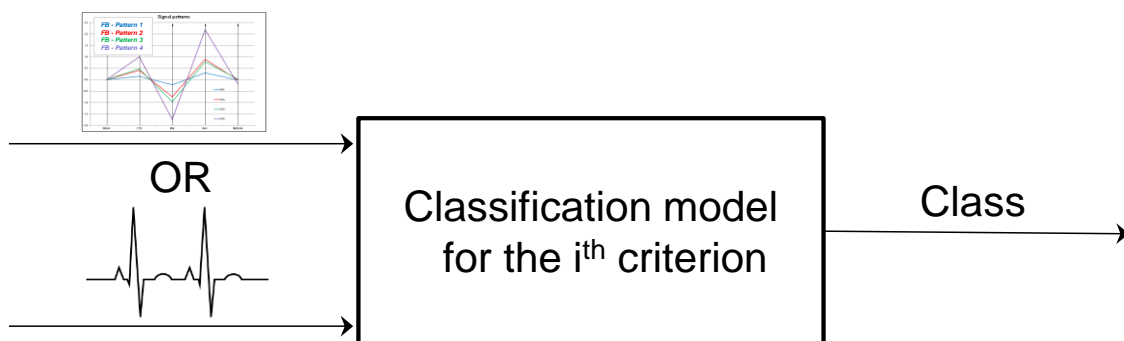


Figure 2 - Possible input to the classification model.

For use case 1.A ECG/FEATURES it is highlighted that the features used as input in the classification model(s) are different from those used to build the classes; otherwise the results would be trivial.

Therefore the scope, in this case, is to evaluate how-well different sets of medical features perform in the diagnosis associated to the i<sup>th</sup> class (compared with those used to build the same i<sup>th</sup> class).



The benefit for astronauts' medical autonomy, in case of success will be that, if some medical features are not available or cannot be computed by astronauts, then these could be replaced by others to determine the diagnosis associated to the non-available features.

On the other side use case 1.B ECG/WAVEFORMS allows verifying if the learning method is capable to automatically distinguish among waveforms belonging to different classes, i.e. being characterised by different diagnoses.

The benefit for astronauts' medical autonomy, in case of success will be that, if no physician may evaluate the ECG recorded by an astronaut and the relevant medical features are not available or cannot be computed by astronauts then the ECG waveform could be automatically evaluated by the supervised model, providing a diagnosis of the ECG trace.

### 2.2.3 Synthetic Results

Table 3 shows the best results obtained for each criterion specifying:

- The type of input used by the classifier, either waveform or feature;
- The lowest *recall*<sup>6</sup> achieved by the classifier in the training and validation datasets;
- The *overall accuracy*<sup>7</sup> achieved by the classifier in the training and validation datasets;

The best results in terms of *overall accuracy* and *recall* of the classifiers have been obtained for both use cases and all criteria by using supervised methods, confirming the importance of the a-priori knowledge represented by the classes related to the ECGs. Indeed if the information on the classes of the data are available, these should be exploited and supervised algorithms should be preferred to unsupervised ones.

Furthermore the results obtained in the two use cases showed also that the two approaches based either on the features or the waveforms are complementary since they perform better on different criteria.

Criteria	Classes	Input used	Training results (%)		Validation results (%)	
			Lowest recall	Overall accuracy	Lowest recall	Overall accuracy
<i>Rhythm</i>	Sinusal	Waveforms	74.4	93.5	69.7	92.6
	Not sinusal					
<i>Ventricular Frequency</i>	Normal	Features	79.5	95.1	79.4	95.3
	Tachycardia					
	Bradycardia					
<i>Orientation of the R wave</i>	Normal	Waveforms	53.0	89.2	46.6	88.7
	Right axis deviation					
	Left axis deviation					

<sup>6</sup> Percentage of elements, belonging to a certain condition or class, which have been correctly predicted.

<sup>7</sup> Percentage of total elements that have been correctly predicted.

Criteria	Classes	Input used	Training results (%)		Validation results (%)	
			Lowest recall	Overall accuracy	Lowest recall	Overall accuracy
	Extreme right axis deviation					
Duration of QRS complex	Normal	Features	100.0	100	99.8	100
	Incomplete bundle branch block					
	Bundle branch block					
	Pathological					
Duration of PR interval	Normal	Features	93.3	98.6	95.7	99.3
	Not normal					
Duration of corrected QT interval	Not pathological	Features	90.9	95.3	90.4	95.5
	Pathological					

Table 3 - Overview of the best results obtained for each criterion (all results refer to supervised methods).

In particular the approach based on features (use case 1.A ECG/FEATURES) achieves good results, being characterized by high *accuracies*, *recall* and *precision* for the following criteria:

*Ventricular frequency*

*Duration of the QRS complex*

*Duration of the PR interval*

*Duration of the corrected QT interval*

On the other side the approach based on waveforms (use case 1.B ECG/WAVEFORMS) achieves promising results being characterized by medium *recall* and *precision* for the following criteria:

*Rhythm*

*Orientation of the R wave*

It is worth highlighting that the most representative measure of the goodness of the classifier is deemed the “lowest *recall*” percentage representing the worse classification percentage of a specific class of a given criterion. On the other side, the *overall accuracy* may be affected by an unbalanced distribution of data over the different classes. Indeed if most data belong to only one class and this class is correctly identified by the classifier, then the *overall accuracy* will be mainly representative of the correct identification of this class only (which will be characterized by a high *recall*), with modest influence of the lower *recall* in the other classes.

The four criteria achieving good results with the feature approach are all characterized by a lowest *recall* higher than 80%. On the other side, the two criteria characterized by promising results with the waveform approach are characterized by a lowest *recall* of 69.7% and 46.6% in the validation

dataset. The 69.7% obtained with the *Rhythm* criterion could still be deemed as an acceptable result, while the 46.6% obtained with the *Orientation of the R wave* criterion is deemed too low.

For both cases it is possible to conclude that enhancements are expected if these methods had to be used for the improvement of astronaut's medical autonomy. Indeed the physicians of the AUSL Modena commented that the use of a reduced number of leads in the automatic evaluation of ECG waveforms could further limit the results because useful information on the ECG traces are lost. Moreover, it is also difficult to retrieve only from the ECG traces all or other surrounding factors and elements that a physician considers to perform the diagnosis, such as the overall status of the patient and the reasons of the request of the clinical exam,

### 2.2.4 Example of unsupervised method - Orientation of the R wave

Unsupervised methods were investigated mainly in the use case 1.B ECG/WAVEFORMS since these methods showed promising results as illustrated below. On the other side for use case 1.A ECG/FEATURES unsupervised methods had shown worse performances than supervised techniques thus they were discarded.

Although the physicians of the AUSL Modena highlighted that different isoelectric values do not have a clinical meaning, the clustering applied to raw data (with no pre-processing step hence also without removing the isoelectric value) seems to obtain good results for the classes belonging to the criterion *Orientation of the R wave*, at least from visual inspection.

Figure 3 compares the mean value of the waveforms of the four classes of the criterion *Orientation of the R wave* (left) and the patterns automatically identified by the unsupervised method, based on Neural Network, partitioning data into four groups (right), without removing the isoelectric values. It is clear the similarity of the four waveforms in the two figures.

However the *overall accuracies* and *recall* reached by the classifier are not high for all the classes which is probably due to the high variability of the human being, which makes difficult the correct assignment of each single ECG to the correct class. In other word the mean waveforms are correctly identified by the automatic algorithm but the high data variability makes difficult the correct classification, i.e. the correct assignment of the single waveforms to their actual class.

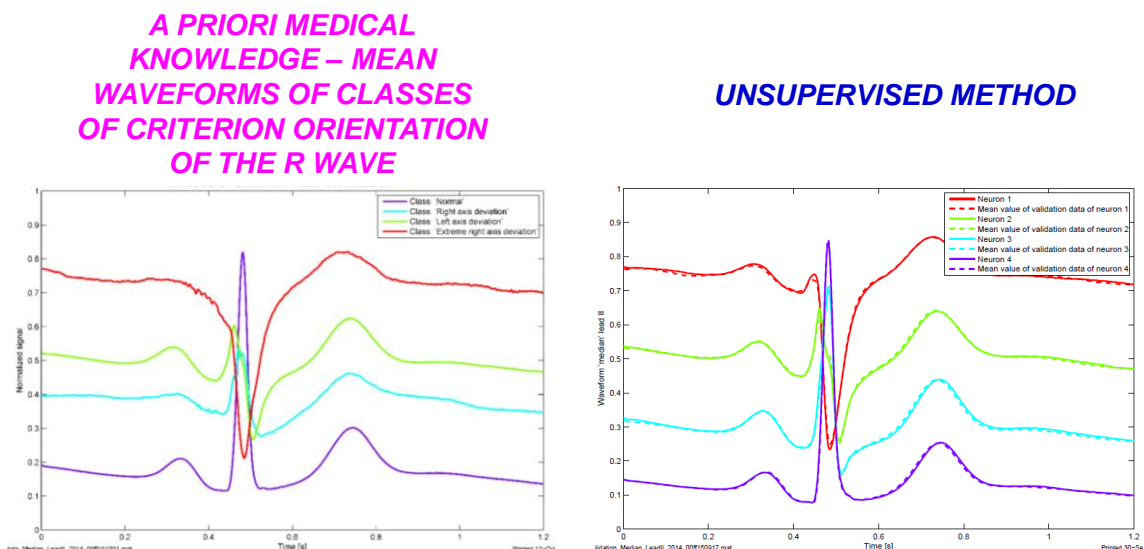


Figure 3-Comparison between the mean value of the waveforms of the four classes of the criterion *Orientation of the R wave* (left) and four patterns identified by the unsupervised method (right).

This high variability of the waveforms may be clearly seen in Table 4 where all the normalized ECGs of the 2014 of the training dataset are plotted overlapped according to their belonging to the four classes of the criterion *Orientation of the R wave*.

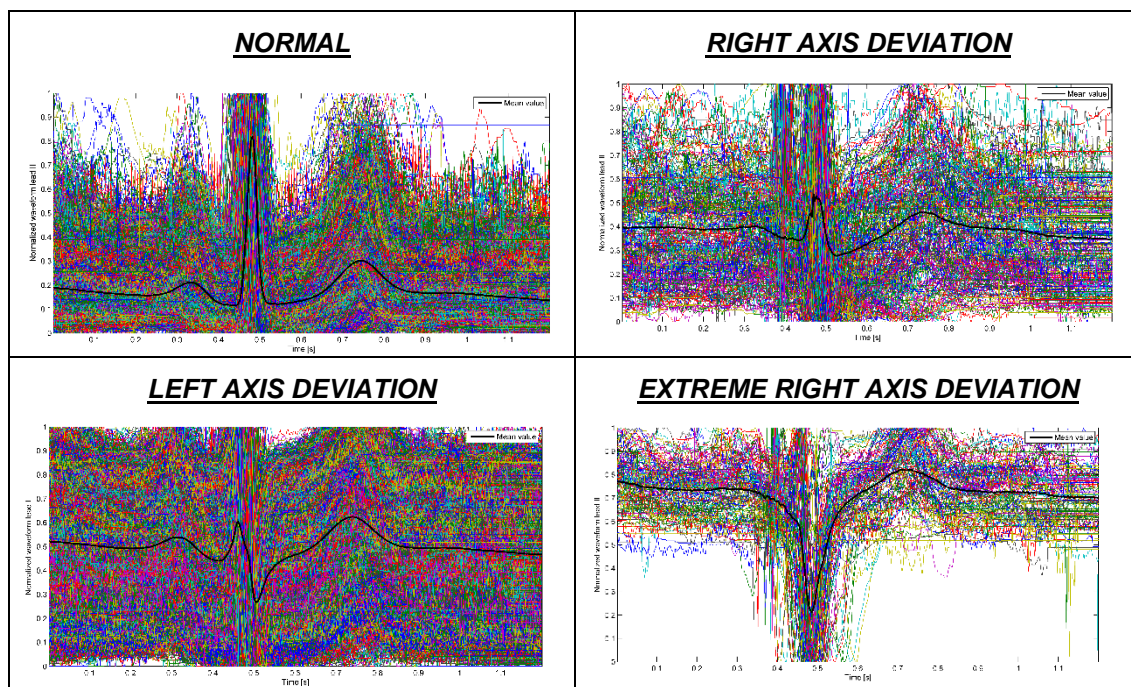


Table 4 - ECG belonging to the four classes of the criterion *Orientation of the R wave*. The thick black line represents the mean waveform.

### 2.2.5 Example of supervised method - Duration of the QRS complex

Supervised methods were investigated in both use cases 1.A ECG/FEATURES and 1.B ECG/WAVEFORMS achieving better results than unsupervised techniques.

This section shows, as example, the detailed results obtained using supervised methods for the *Duration of the QRS complex* criterion in use case 1.A ECG/FEATURES.

Table 5 and Table 6 show the *confusion matrixes* obtained with the training and validation datasets respectively.

The *overall accuracies* and *recall* and *precision* are almost 100% with both the training and validation datasets.

Therefore the results are good and the classifier is deemed suitable for the improvement of astronaut's medical autonomy.

		Actual class				TOTAL ESTIMATED
		Normal	Incomplete bundle branch block	Bundle branch block	Pathological	
Estimated class	Normal	<b>19793</b>	<b>0</b>	<b>1</b>	<b>0</b>	<b>19794</b>
		100.0%	0.0%	0.0%	0.0%	100.0%

		Actual class				TOTAL ESTIMATED
		Normal	Incomplete bundle branch block	Bundle branch block	Pathological	
	Incomplete bundle branch block	<b>0</b> 0.0%	<b>1253</b> 100.0%	<b>0</b> 0.0%	<b>0</b> 0.0%	<b>1253</b> 100.0%
	Bundle branch block	<b>0</b> 0.0%	<b>0</b> 0.0%	<b>3268</b> 100.0%	<b>0</b> 0.0%	<b>3268</b> 100.0%
	Pathological	<b>0</b> 0.0%	<b>0</b> 0.0%	<b>0</b> 0.0%	<b>4965</b> 100.0%	<b>4965</b> 100.0%
TOTAL ACTUAL		<b>19793</b> 100.0%	<b>1253</b> 100.0%	<b>3269</b> 100.0%	<b>4965</b> 100.0%	<b>29280</b> 100.0%

Table 5 - Confusion matrix of the best classifier obtained for the *Duration of the QRS complex* criterion with the training dataset.

		Actual class				TOTAL ESTIMATED
		Normal	Incomplete bundle branch block	Bundle branch block	Pathological	
Estimated class	Normal	<b>18518</b> 100.0%	<b>0</b> 0.0%	<b>1</b> 0.0%	<b>0</b> 0.0%	<b>18519</b> 100.0%
	Incomplete bundle branch block	<b>0</b> 0.0%	<b>1123</b> 100.0%	<b>0</b> 0.0%	<b>0</b> 0.0%	<b>1123</b> 100.0%
	Bundle branch block	<b>0</b> 0.0%	<b>0</b> 0.0%	<b>2834</b> 99.8%	<b>0</b> 0.0%	<b>2834</b> 100.0%
	Pathological	<b>0</b> 0.0%	<b>0</b> 0.0%	<b>5</b> 0.2%	<b>4701</b> 100.0%	<b>4709</b> 99.8%
TOTAL ACTUAL		<b>18521</b> 100.0%	<b>1123</b> 100.0%	<b>2840</b> 99.8%	<b>4701</b> 100.0%	<b>27185</b> 100.0%

Table 6 - Confusion matrix of the best classifier obtained for the *Duration of the QRS complex* criterion with the validation dataset.



Figure 4 shows the patterns of the four classes of the criterion *Duration of the QRS complex*, built using the four selected features.

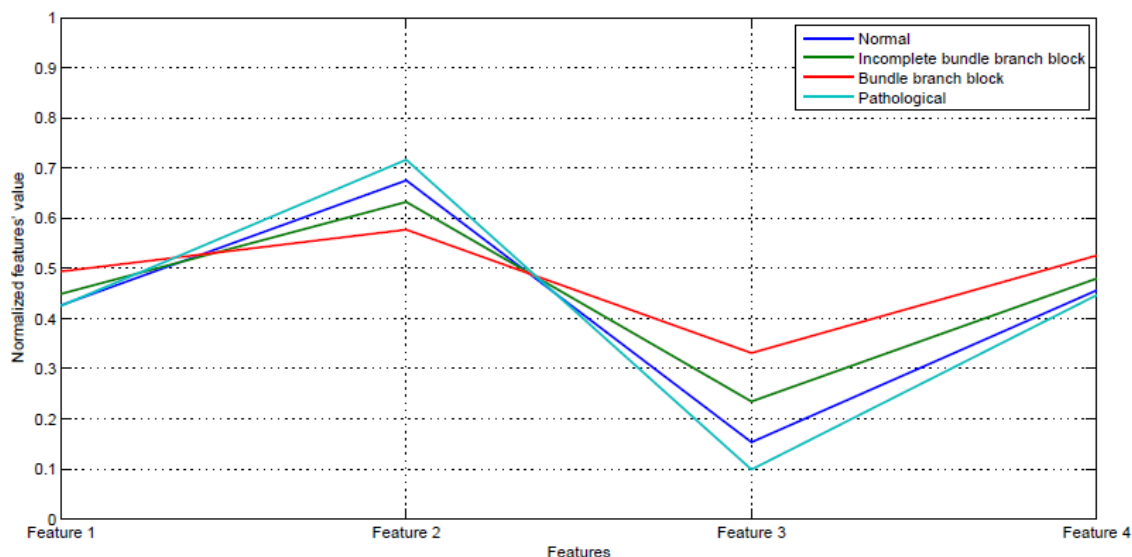


Figure 4 - Patterns of the four classes of the criterion *Duration of QRS complex*.

Figure 5 shows the distributions of the values of each feature used in the best classifier of the criterion *Duration of the QRS complex* by means of boxplot<sup>8</sup>.

Each boxplot shows the distribution of the values of the *i-th* feature in all the four classes where the class 1 (blue) is associated to the class “Normal”, the class 2 (green) is associated to the class “Incomplete bundle branch block”, the class 3 (red) is associated to the class “Bundle branch block” and the class 4 (cyan) is associated to the class “Pathological”. Furthermore the numbers displayed above each boxplot represents the number of items, i.e. ECGs, that are assigned to the each class.

<sup>8</sup> Work tool which allows the graphical representation of a set of measures, highlighting its quantiles.

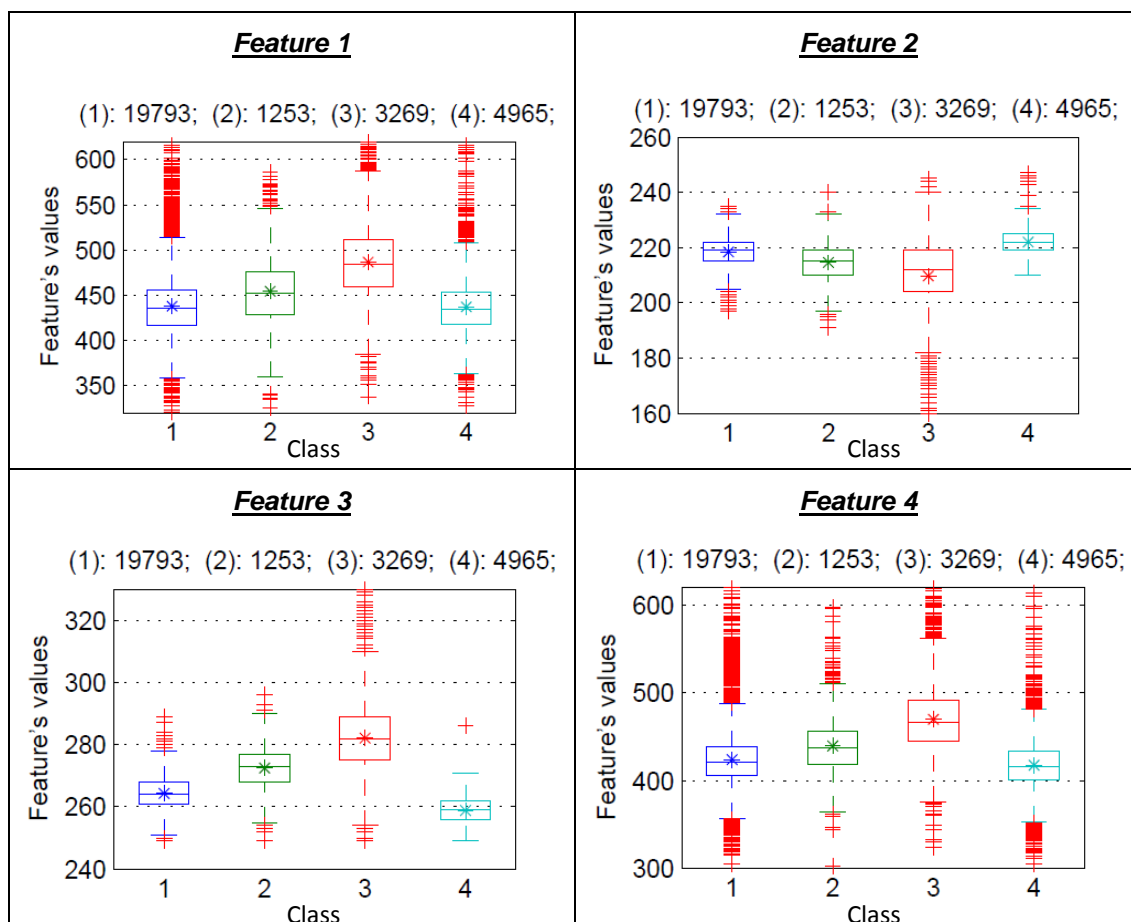


Figure 5 - Boxplot of the features used in the best classifier of the criterion *Duration of QRS complex*.

Figure 5 shows that it is not possible to identify a single feature granting a complete separation of data among the four classes, since for all the four features at least two classes show distributions that are partially overlapped.

However, the accuracies obtained with the best classifier of the criterion proves that the automatic supervised techniques are able to perform linear and non-linear combinations of the selected features that allow separating the four classes. Therefore even if a single feature is not able to grant a separation among the four classes, a suitable combination of the features allows a good separation yielding to a classifier characterized by very high accuracies.

Indeed the analysis of all the possible combinations among the four features showed that by computing the difference between Feature 3 and Feature 2 it is possible to achieve a complete separation of the distributions of data belonging to different classes as shown in Figure 6.



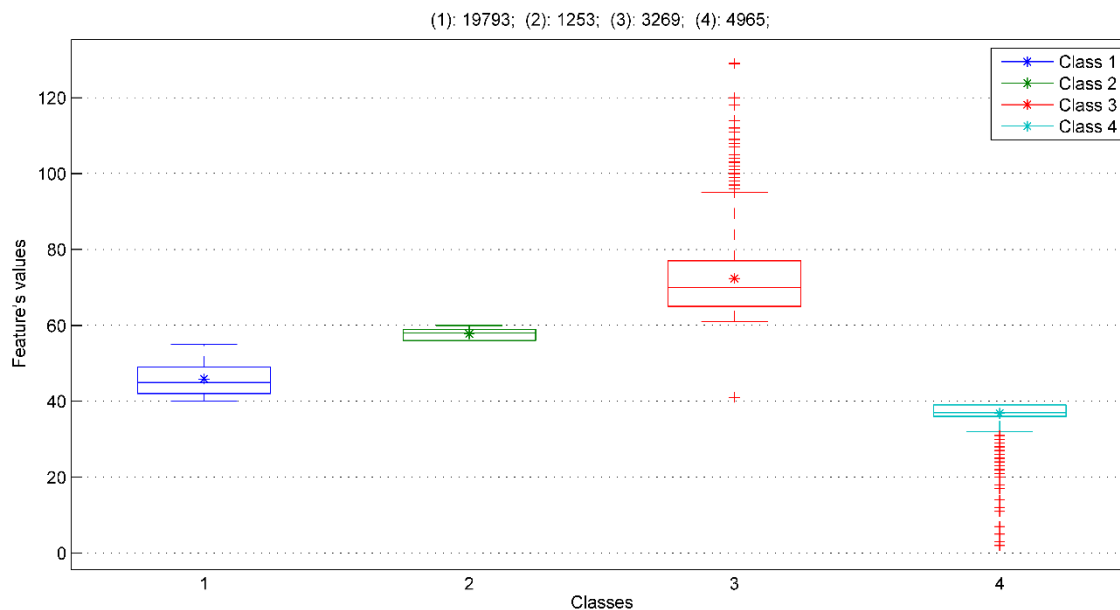


Figure 6 - Boxplot of the values obtained computing the difference between Feature 3 and Feature 2.

## 2.2.6 Medical validation and autonomy

The results described in the previous sections show that the goal of the use case 1.A ECG/FEATURES has been fulfilled at least for four out of the six criteria, since it is possible to detect the classes (and related diagnoses) using a set of features different from those used for the definition of the classes.

Therefore, the medical autonomy of astronauts could be improved by using these identified classifiers. As an alternative, the use of these classifiers may allow implementing a redundant approach that can either confirm the diagnosis provided through standard evaluation or suggest to repeat a specific analysis in case the two diagnoses do not agree. Indeed this redundant approach would improve the reliability of the diagnosis, which is a very important aspect when relying on automatic systems that are subject to possible errors, especially on isolated environments such as the International Space Station or Planetary exploration missions (e.g. Mars), in which the assistance of medical doctors may not be always available.

Figure 7 shows a possible flowchart for the development and implementation of an automatic diagnostic tool for the evaluation of ECGs using data mining methods and aiming at the improvement of astronaut's medical autonomy. The flowchart include three main steps:

1. The first step is associated to the definition of the Medical knowledge. On the basis of the available medical dataset and its characteristics the physicians define a certain number of criteria and classes that allows data labelling. This represents the a-priori medical knowledge that can be exploited in the application of supervised methods.
2. The second step is associated to the Training. Exploiting the available medical dataset and the a-priori medical knowledge defined in the previous step, the use of supervised learning methods allows the identification of automatic classifiers.
3. The third step is associated to the Use of the knowledge generated in the previous two steps. When a new medical datum, also called observation, is available, this is used as input to the classifier identified in the second step that provides automatically the class to which the medical datum belongs and the corresponding diagnosis of the status of the patient.

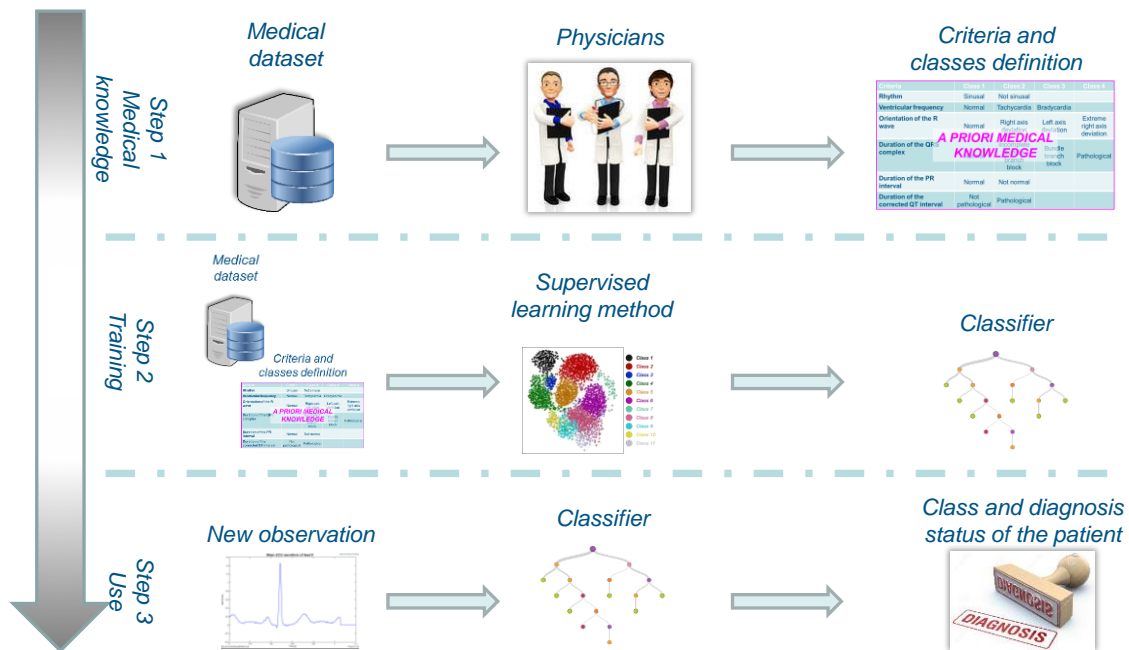


Figure 7-Possible flowchart for the development and implementation of an automatic diagnostic tool for the evaluation of ECGs using data mining methods and aiming at the improvement of astronaut's medical autonomy.

On the other side the results achieved in the automatic classification of the entire waveforms (use case 1.B: ECG/WAVEFORMS) showed an intrinsic difficulty in the application of data mining techniques. This difficulty is associated to the variability (and adaptability) of the human being, which make difficult the creation of "normal" patterns to be exploited for the identification of anomalies of any individual.

This difficulty could be also due to the fact that in usual medical practice the diagnosis is performed by physicians by evaluating all the twelve leads' traces, not limiting to only few of them. Therefore the use of a reduced number of leads in the automatic evaluation of ECG waveforms could limit the results, because useful information on the ECG traces are lost.

## 2.3 Laboratory data

### 2.3.1 Dataset

The Laboratory data are provided in CSV format. CSV files are commonly used for tabular data exchange. In these files, text and numbers are displayed as plain text, allowing the visualization by a standard text editor.

The available Laboratory data files include blood and urine analyses and are not labelled with a diagnosis performed by a physician. Since no a priori diagnosis's knowledge is available, the detection of patterns can only be performed by means of *unsupervised methods*.

In usual clinical practice, the results of the Laboratory data are compared with thresholds, determining if a result is "normal" or "not normal". These thresholds may be either fixed or variable depending on whether they are equal for all patients or if they vary according to the patient's age and/or gender and/or race and/or usual living environment, e.g. local nutrition habits.

### 2.3.2 Use case

The use case agreed with ESA for the analysis of laboratory data was named USE CASE 4: LAB/CLUSTERING<sup>9</sup> aiming at the development of a clustering algorithm based on blood and urine analysis capable of grouping data in a certain number of groups (called *clusters*) without the use of any a priori-knowledge.

The medical knowledge was exploited a posteriori, i.e. only to interpret the results achieved by the knowledge extraction, to verify their significance from a clinical point of view and to be able to associate a diagnosis, if any, to the groups automatically identified by the unsupervised algorithm.

The proposed approach is to use the single laboratory tests as single features to build feature-based patterns, as in the example shown in Figure 8, where the medical features White blood cells, Neutrophils, Lymphocytes, Glucose and Creatinine are being used.

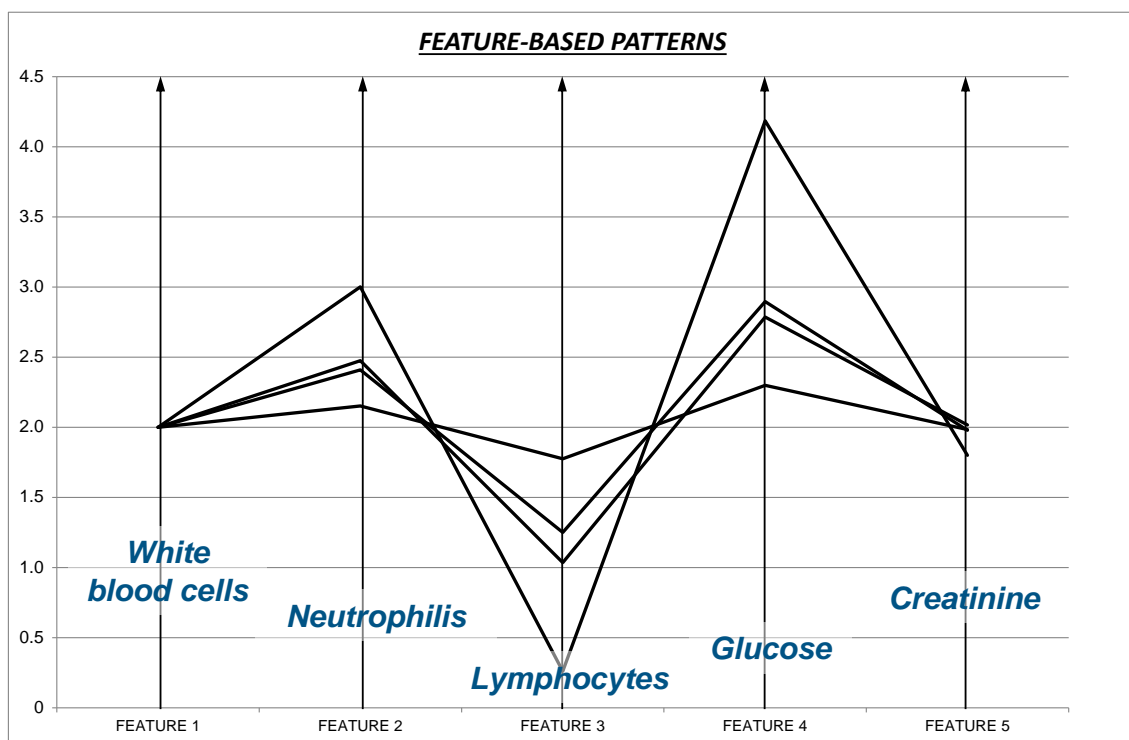


Figure 8 - Example of medical feature based patterns.

The selection of the tests to be investigated is done according to the four panels that were initially defined together with the laboratory physicians of the AUSL Modena according to the following criteria:

- to avoid focusing on specific pathologies, which usually do not affect astronauts (given that their health is continuously checked and the specific environment in which they operate);
- to focus on those laboratory tests capable to capturing the general health status of a person;
- to exploit at the best the large quantity of available data avoiding the use of tests that are executed only rarely.

<sup>9</sup> The number *four* assigned to this use case is due to the fact that the two selected use cases were selected among a short-list of possible use cases defined during the first phase of the project on the basis of the available dataset.

During the execution of the project additional panels were defined in agreement with the physicians of the AUSL Modena to investigate specific diseases and verify the applicability of the method for the detection of unknown correlations among work orders.

These additional panels were named from “01” to “04”.

### 2.3.3 Synthetic Results

Laboratory data were processed exploiting the method defined in the parallel ESA project AUTOMATIC SPACECRAFT STATUS CHARACTERISATION BY DATA MINING MISSION HISTORY, based on unsupervised methods.

The method performs several clustering iterations aiming at identifying the best way to partition data on the basis of a *customized performance measure*, defined in the project synthesising the clusters cohesion and isolation. This *customized performance measure* is a synthetic measure that indicates if the partitioning is suited or not for being used in a diagnostic system providing an indication of its reliability in the classification of a new observation (i.e. the association to a specific pattern generated by the unsupervised method). Low values close to 0% are associated to partitions in which data belonging to the different patterns are completely separated, implying that a new observation would be assigned to only one pattern, thus the associated diagnosis would be highly reliable. On the other side, high values close to 100% are associated to partitions in which data belonging to the different patterns are not separated, implying that a new observation could not be uniquely assigned to a single pattern (i.e. the assignment is performed with uncertainty), thus the associated diagnosis could be unreliable. In conclusion, this *customized performance measure* may be considered as a measure of the uncertainty of the diagnosis.

The results proved that it is possible to automatically identify groups (clusters) that may be used for diagnostic purposes to identify specific diseases or to assess the overall status of a person since it is possible to separate almost completely data belonging to different groups. This reduces the uncertainty in the assignment of data to a specific pattern, i.e. increasing the reliability of the diagnosis.

Moreover the results also proved the possibility to automatically identify meaningful patterns in data without the use of any a priori knowledge, which were later validated by the physicians of the AUSL Modena.

These patterns can be:

- recurrent in multiple panels (A, B, C...).
- clearly visible only for a single panel type.
- likely to correspond to specific pathologies (e.g. renal or liver failure, anaemia).
- likely to correspond to healthy patients.

Finally the results also proved that the defined method may be used for the investigation of possible unknown correlations among *tests*. In particular this analysis could be driven by clinical markers of a specific pathology, such as heart failure, aiming at identifying patterns associated to it.

#### 2.3.4 Example of unsupervised method

Table 7 shows an example of the patterns that were identified on panel D by partitioning data into six groups<sup>10</sup> on a population including male and female patients with age ranging from 25 to 50 years, including 96.363 patients totalling 174.128 panels<sup>11</sup>.

The following patterns were identified and clinically validated by the physicians of the AUSL Modena:

- The **red pattern**, characterized by very high values of creatinine (“Creatinina”, feature number 1) and Urea (“Urea”, feature number 8). These groups have also low values of red blood cells (“globuli rossi”, feature number 3) and haemoglobin (“Emoglobina”, feature number 4).

The physicians recognized this pattern as associated to patients affected by **renal failure** (very high values of creatinine and urea) which is often characterized also by anaemia (low values of red blood cells and haemoglobin).

- The **green pattern**, characterized by:
  - o Low values of red blood cells and Haemoglobin
  - o Normal (high) values of mean corpuscular volume

This pattern is associated to an **anaemia due to lack of the vitamin B12**, which is characterized by large red blood cells.

- The **blue pattern**, characterized by:
  - o Normal values of red blood cells;
  - o Low values of Haemoglobin and Mean corpuscular volume;
  - o Number of platelet greater than the other centres, even if within the standard medical thresholds

This pattern is associated to an **anaemia due to lack of iron**, which is characterized by small red blood cells with a consequent reduction of the haemoglobin and the mean corpuscular volume.

- The **cyan, yellow and magenta patterns** that may be associated to **healthy patients** being all characterized by work orders within the standard medical thresholds indicated by a thick black line. It is interesting to notice that the cyan pattern includes a majority of male patients (89%) while the magenta and yellow patterns include a majority of female patients (83% and 70% respectively). This is confirmed by the fact that male and female patients are characterized by different typical values.

It is observed that the three patterns associated with a clinical disease by the physicians of the AUSL Modena (red, green and blue) are the only ones characterized by some work orders exceeding the standard medical thresholds. In particular, the red pattern is characterized by four work orders exceeding the thresholds while the green and blue patterns by two work orders each.

Noteworthy these pathological patterns are recurrent in the several clustering iterations being detectable using also different patient age classes, different combination of patient gender and different panels.

---

<sup>10</sup> For each of the group a pattern is identified by the unsupervised method.

<sup>11</sup> All accessing the AUSL Modena hospitals in the years from 2009 to 2014.

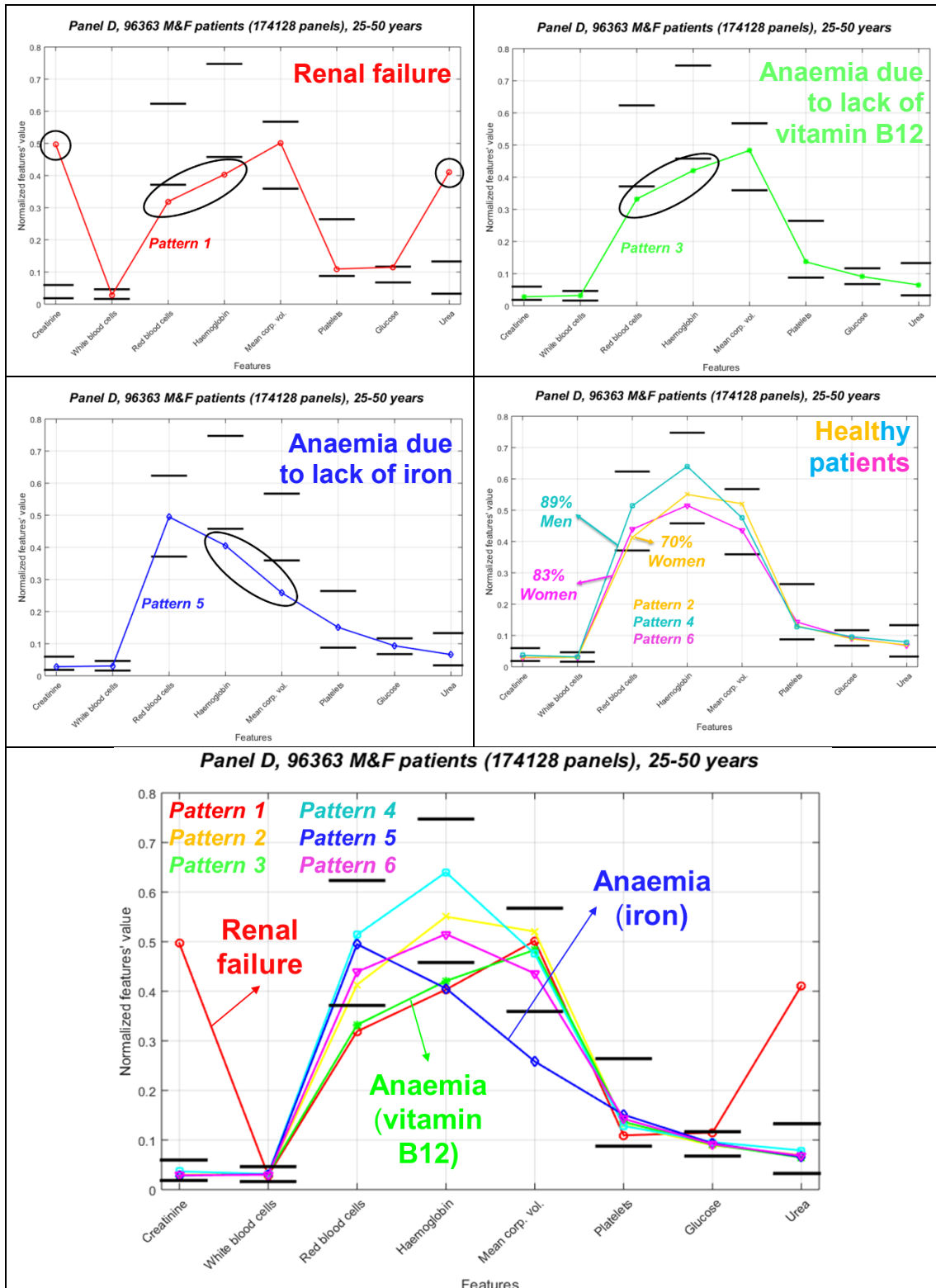


Table 7 - Patterns identified in Panel D partitioning data into six groups for patients with age between 25 and 50 years. The standard medical thresholds are shown by a thick black line.



The patterns shown in Table 7 can be also illustrated by a pictorial representation using polygons or polar diagrams in which each radial axis correspond to a feature (i.e. work order). This graphical representation is also known as *radar chart* and allows displaying multivariate data in the form of a two-dimensional chart of three or more quantitative variables represented on axes starting from the same point. The relative position and angle of the axes is typically uninformative.

The radar chart is a chart that consists of a sequence of equi-angular spokes, called *radii*, with each spoke representing one of the features. A line is drawn connecting the data values, associated to each pattern, for each spoke. This gives the plot a star-like appearance.

Each pattern shows a distinguishable characteristic shape. Figure 9 shows the radar chart of the patterns shown in Table 7, where the standard medical thresholds are marked by a black dot in each axis. It is clear that the red, green and blue patterns, associated to pathological patients, are characterized by different shapes (see also Figure 10 left) while the three patterns associated to healthy patients (yellow, cyan and magenta) are characterized by shapes that are more similar to each other (Figure 10 right), which is ideal for anomaly detection.

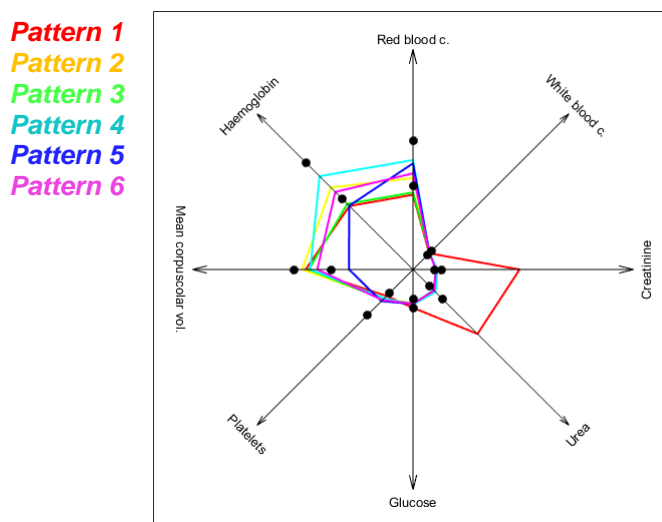


Figure 9 - Radar or polar chart of the patterns shown in Table 7.

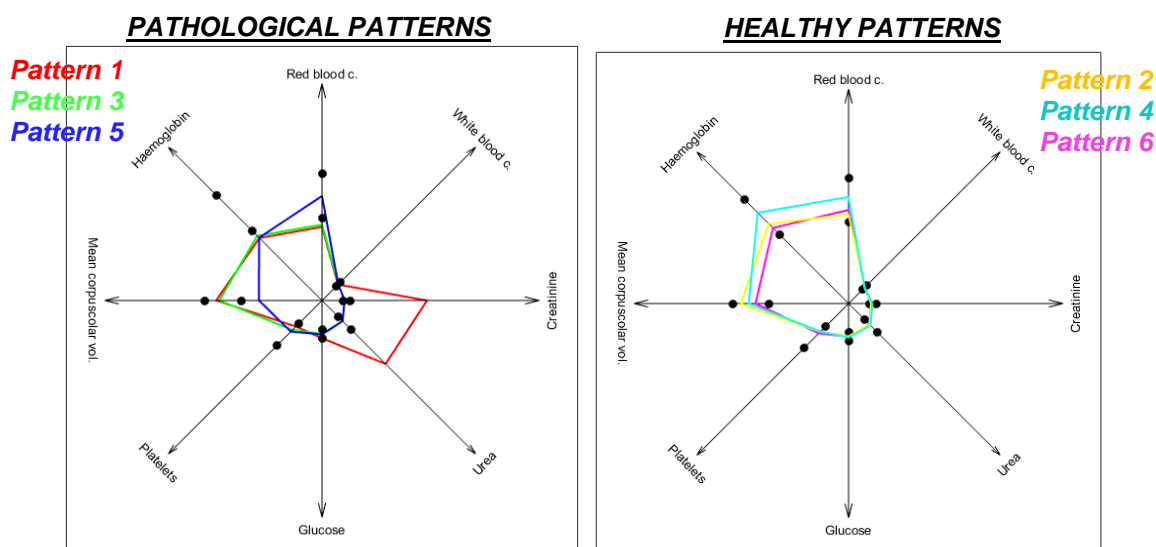


Figure 10 - Radar or polar chart of the patterns associated to pathological (left) and to healthy patients (right).



### 2.3.5 Example of previously unknown correlation

A correlation previously unknown to physicians, as after the state of the science, has been detected by investigating the combination of the laboratory tests including the Troponin that is a pathological marker of heart attack.

This correlation is represented by the blue pattern in Figure 11, in which it is possible to see that high Troponin values are correlated with:

- High values of chlorine, white blood cells and sodium (“cloro”, “globuli bianchi” and “sodio”, features number 1,3 and 10);
- Low values of platelets (“Piastrine”, feature number 7);
- Slightly high values of glucose (“Glucosio”, feature number 8).

The physicians of the AUSL Modena commented that this is an interesting result and that the high values of chlorine and sodium should be further investigated since they were not aware of this correlation.

Indeed this information, once validated by more detailed analyses, could allow improving the interpretation of data, the diagnosis of diseases and the possible treatments.

**Panel C2, 353 M&F patients (446 panels), 25-50 years**

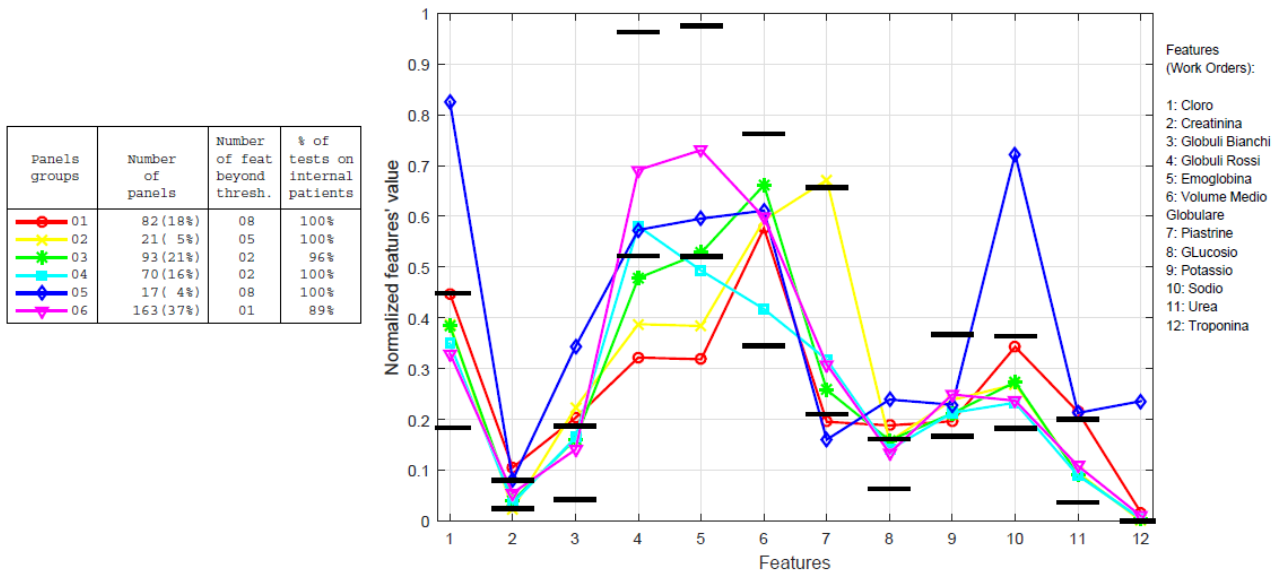


Figure 11 - Example of unknown correlation identified by the blue pattern. The standard medical thresholds are shown with a thick black line<sup>12</sup>.

### 2.3.6 Medical validation and autonomy

The results described in the previous sections show that the goal of the use case 4: LAB/CLUSTERING has been fulfilled since meaningful patterns were identified, without the use of any a priori knowledge, and also clinically validated a posteriori by the physicians of the AUSL Modena, who could associate them to known diseases such as renal and liver failures or anaemia.

Furthermore the results proved that the automatic clusters identified are characterized by a low or null customized performance measure, thus may be used for diagnostic purposes combining

<sup>12</sup> The legend on the left of the figure shows for each group the number of panels assigned to that group, the number of features beyond the standard medical thresholds and the percentage of tests carried out on internal patients (considered as actually sick).

the laboratory tests among them and assigning them to a given pattern to which a specific diagnosis can be associated.

This association may be based on a multidimensional distance metric (to be selected) with respect to all the identified patterns. For example, the identification of a “normal” and a “pathological” pattern identified by clustering into two groups may allow determining if the status of the person is likely to be normal or not.

Therefore the medical autonomy of astronauts may be improved by the use of the clusters characterized by low or null customized performance measure. This allows performing a diagnosis that may provide an indication of a specific disease or of the overall status of the person.

Figure 12 shows a possible flowchart for the development and implementation of an automatic diagnostic tool for the evaluation of laboratory data using data mining methods and aiming at the improvement of astronaut’s medical autonomy. The flowchart include three main steps:

1. The first step is associated to the Training, i.e. the extraction from the available medical dataset of relevant patterns by exploiting unsupervised learning methods.
2. The second step is associated to the Medical Validation. Patterns characterized by low customized performance measures are selected and validated by the physicians assigning a specific diagnosis to each of them.
3. The third step is associated to the Use of the knowledge generated in the previous two steps. When a new medical datum, also called observation, is available, this is used as input in the classifier identified in the first step matching it with one of the selected and validated pattern, to which a specific diagnosis is associated.

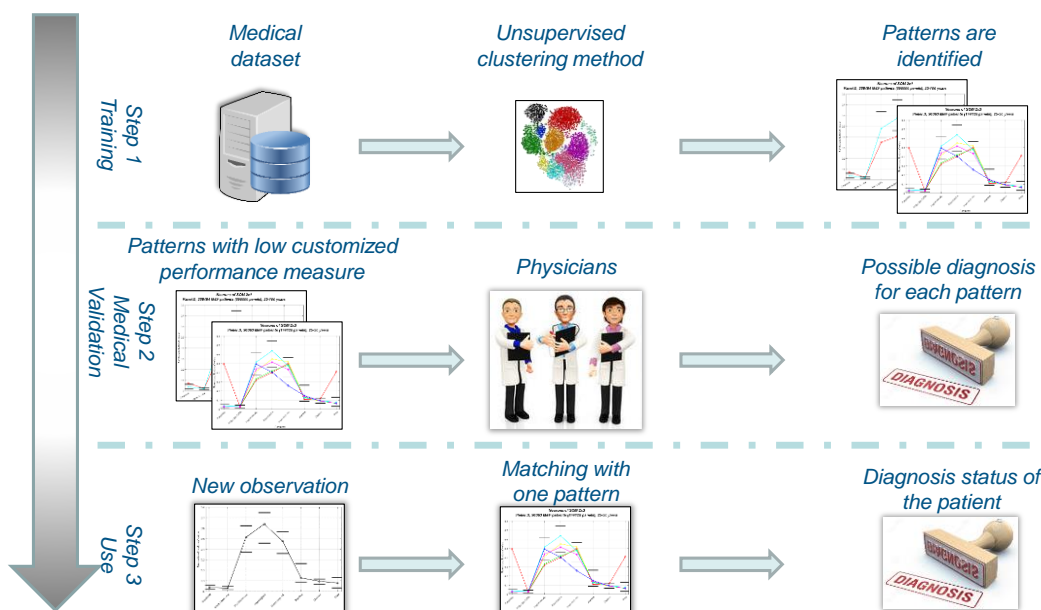


Figure 12-Possible flowchart for the development and implementation of an automatic diagnostic tool for the evaluation of laboratory data using data mining methods and aiming at the improvement of astronaut’s medical autonomy.

#### 2.4 Applicability of Novelty Detection

The project included also the evaluation of the applicability of ESA’s *Novelty Detection* algorithm and software to medical datasets.

This analysis was performed both on the ECG and laboratory data.

The algorithm requires the use of a-priori knowledge for building the so-called *nominal* file that is used as reference behaviour and the *target* file that contains the behaviour to be investigated for novelties.

#### 2.4.1 ECG

The applicability of ESA's *Novelty Detection* algorithm for ECG data has been investigated using as input to the algorithm both waveforms and feature time series.

The results using the waveforms as input showed that *Novelty Detection* does not detect any novelties between two different waveforms, characterized by a different diagnosis, proving that the statistical features (minimum, maximum, etc) currently implemented in *Novelty Detection* are not suitable to detect differences in medical waveforms such as ECGs.

The results using the features as input proved that the *Novelty Detection* identifies the novelties (with low percentages of false and missed alarms<sup>13</sup>) mainly only on the time series associated to the medical feature that has been used to build the classes of each criterion. This allows using *Novelty Detection* to identify ECGs belonging to different classes. However, this distinction allows the separation of ECGs only into two main groups:

- one usually associated to the normal class
- one associated to pathological classes, all being grouped together.

*Novelty Detection* does not allow, instead, distinguishing among the identified novelties, i.e. among other different kinds of pathological classes.

#### 2.4.2 Laboratory

The results show that *Novelty Detection* algorithm seems unsuited for the investigation of novelties in data built through the combination of laboratory tests since:

- No novelty is detected for values that were never seen before, probably because they are considered outliers by the algorithm;
- Values due to a higher resolution of the measurements, even if these are within already observed nominal values, are detected as novelties.

It is also highlighted that the need of defining a *nominal* dataset for the *Novelty Detection* algorithm requires the use of some a-priori knowledge, unlike the unsupervised clustering method proposed in this project for the laboratory data does not require.

### 2.5 Conclusions

The results obtained in the project both in the analysis of ECG and laboratory data proved that medical autonomy of astronauts can be improved by the application of data mining algorithms. Furthermore the results also showed that these data mining algorithms could help physicians, involved in usual clinical practice, in data interpretation supporting them in the disease diagnosis and treatment.

Moreover the project highlighted the need of personalizing the clinical pathway of a patient taking into account his/her most relevant characteristics such as age and gender but not limiting to these two. Indeed the use of tailored *medical references* may allow improving the prevention, diagnosis and treatment of diseases.

---

<sup>13</sup> The percentage of *false alarms* is defined as the ratio between the number of novelties that are not actually a novelty and the number of actual novelties. The percentage of *missed alarms* is defined as the ratio between the number of actual novelties that are not recognized as a novelty and the number of actual novelties.

These *medical references* may be built thanks to the adoption and widespread of electronic health information systems in healthcare institutions, which make available medical information in a more and more organised infrastructure that can be exploited also by automatic data analysis tools.

The use of personal and customized *medical references* allows improving the performances of automatic algorithms since, as reported by Nithya et Al.<sup>14</sup>, accurate results in using data mining methods are usually achieved considering a small subset of medical datasets, remaining difficult to apply such techniques to much larger generalized dataset.

This is also confirmed by the recent (2013) survey of data mining approaches for healthcare that reviewed the studies available in the literature on this topic. It reported accuracies that range from 45% up to 99%, according to the specific medical area addressed, proving that the results that may be achieved in the application of data mining methods to healthcare data are strongly dependent on both the dataset under investigation and on the specific topic addressed.

It is worth highlighting that the use cases defined in the project are characterized by a broad character, which were at the basis of the scope of work. Indeed the project adopted the so-called *inter-patient* approach that consists in evaluating data of a new patient according to a reference database and a model built from data from other patients.

This assure a broad generality but it may limits the results because of the variability (and adaptability) of the human being, which make difficult the creation of “normal” *medical references* which can be exploited for the identification of pathologies of any individual.

On the other side an *intra-patient* approach consists in evaluating data of a same patient according to a reference database and a model built from his/her own data. In this case a relevant quantity of data of that specific patient is required in order to adopt both a *data-centric* (i.e. based on data) *intra-patient* approach.

It is worth underlining that for ESA scopes an *intra-patient* approach could improve the medical autonomy to be provided to astronauts because personal *medical references* could be built allowing more accurate diagnosis. Indeed given that the number of astronauts is very limited and that they are under continuous health monitoring, ESA should have available a large amount of health data related to each specific astronaut making the *intra-patient* approach a viable alternative.

Therefore given the broad generality of the project, the results that have been achieved, which were also supported by medical validation by the AUSL Modena specialists, are deemed of particular relevance taking into account that the extension of the dataset included over 62.000 ECGs belonging to over 41.000 patients and 87.000.000 laboratory tests belonging to over 650.000 patients.

In extreme synthesis the results obtained by the analysis of ECG data showed that high overall accuracies (greater than 95%) are achieved for the criteria *Ventricular frequency, Duration of QRS interval, Duration of the PR interval and Duration of the corrected QT interval.*

On the other side, the results obtained with the *Rhythm and Orientation of the R wave* criteria, led to the conclusion that enhancements are needed in order to apply these methods for the improvement of astronaut’s medical autonomy. Indeed the medical validation proved that the use of a reduced number of leads in the automatic evaluation of ECG waveforms could limit the results, because useful information on the ECG traces are lost. Moreover, it is also difficult to retrieve only from the ECG traces all the surrounding factors and elements, such as the overall status of the patient and the reasons of the request of the clinical exam, that a physician consider to perform the diagnosis.

The results obtained by the analysis of the laboratory data showed that:

---

<sup>14</sup> N.S.Nithya, K.Duraiswamy, P.Gomathy, “A Survey on Clustering Techniques in Medical Diagnosis”, International Journal of Computer Science Trends and Technology (IJCT), Volume1 Issue2, Nov-Dec 2013.

- it is possible to automatically identify meaningful patterns without the use of any a priori knowledge, which were validated by the specialist physicians of the AUSL Modena. For example, patterns related to renal failure, liver failure and anaemia were identified.
- these meaningful patterns highlight the presence of known correlations among clinical data, confirmed by the physicians of the AUSL Modena, which demonstrated the validity of S.A.T.E. solutions (e.g. the known correlation between creatinine and urea in patients affected by renal failure).
- these meaningful patterns proved to be suited for diagnostics purposes, to identify specific disease or to assess the overall status of a person since it is possible to separate almost completely data belonging to different patterns. This reduces the uncertainty in the assignment of data to a specific pattern, i.e. increasing the reliability of the diagnosis.

Moreover the results showed also that:

- some laboratory tests have a different distribution according to age, gender, hospitalization status, which was expected, but still not always considered for the definition of standard medical thresholds.
- some laboratory tests showed an unexpected distribution, mostly outside the standard medical thresholds, which could be due either to an erroneous definition of the standard medical thresholds, or to the methodology used during the test execution, or to inappropriate test equipment set up. Given the thresholds standardization and the diffusion of the methodologies and equipment set-up, these outcomings, which are under investigation by the AUSL Modena, are to be considered of high medical relevance and usefulness beyond the space application intended for this project.

Therefore these latter results proved the need to better characterize laboratory data taking into account the different characteristics of a patient, such as age, gender, hospitalization status, and other factors that may not have been considered so far. Indeed this characterization would allow updating the standard medical thresholds that are currently used for the evaluation of the laboratory results to implement a tailored therapy improving the prevention, diagnosis and treatment of diseases.

Finally the discovery of unknown correlations, which must also be further investigated, proves that the method and approach based on the analysis of data without the use of a-priori knowledge allows reversing the usual approach of medical research based on the definition of a priori hypotheses and their verification on data. Indeed, the availability of large medical datasets opens the way to approaches based on the analysis of data to generate unknown hypotheses, such as correlations, that must be confirmed by physicians for their medical validation. Once validated, these may be exploited to enhance physicians' ability to diagnose and treat disease.

The results obtained using ESA *Novelty Detection* showed that the software is not suited for its application to medical datasets. Indeed the need of using some a-priori knowledge may also affect the results because it allows the exploration of data focusing only on specific objectives aiming at finding differences with respect to a nominal behavior that must be predefined.

## 2.6 Suggested future developments

Given the results of the projects and the indications provided by the physicians of the AUSL Modena during the several iterations of medical validation of SATE findings the following activities are proposed as possible future developments of the project:

### ECG data

- Investigation of techniques for the combination of all the twelve leads in order to exploit all the traces that are evaluated in common medical practice by cardiologist.
- Investigation of alternative methods for data processing and classification such as the Fisher Discriminant Analysis, which is suited to group data into well separated clusters.

- Focus the analysis on specific pathologies that are of interest to ESA for its astronauts. This analysis can be carried out exploiting also the dataset made available by the consortium and exploiting the “Diagnosis” field that is clinically certified, selecting the only ECGs of the patients affected by the disease of interest.
- Analysis of repeated ECGs belonging to the same patient to evaluate possible variations in his/her health status. This could be done using the dataset made available by the consortium. However since in such dataset the number of repeated exams for single patients is limited, it would be preferable to use another dataset.
- Use of the *intra-patient* approach to evaluate the results that may be achieved in the automatic evaluation of ECGs by using data belonging only to one patient for the evaluation of new data of the same patient.
- Further investigation on the combination of the medical features used by the automatic classifiers to minimize the number of used features.

#### Laboratory data

- Focus the analysis on specific panels and/or pathologies that are of interest to ESA for its astronauts, using suitable pathological markers included in data.
- Further validation of the solutions identified in the project to improve astronaut's medical autonomy.
- Further medical validation of the patterns identified in the project some of which highlighting unknown correlations among some laboratory tests.
- Analysis of possible trends or patterns that are symptoms of values exceeding the thresholds in subsequent tests, focusing on a specific pathology of interest to ESA.