Monitoring Chronic Diseases Using Soft Computing Techniques and Rule-based Systems: The CHRONIOUS Case

Piero Giacomelli
Tesan S.p.A.,
email:giacomelli@tesan.it

Giulia Munaro
Tesan S.p.A.,
email:munaro@tesan.it

Roberto Rosso
Tesan S.p.A.,
email:rosso@tesan.it

***Abstract*—CHRONIOUS is an Open, Ubiquitous and Adaptive Chronic Disease Management Platform for Chronic Obstructive Pulmonary Disease(COPD) Chronic Kidney Disease (CKD) and Renal Insufficiency. It consists of several modules: an ontology based literature search engine, a rule based decision support system, remote sensors interacting with lifestyle interfaces (PDA, monitor touch screen) and a machine learning module. All these modules interact each other to allow the monitoring of two types of chronic diseases and to help clinician in taking decision for cure purpose. This paper illustrates how some machine learning algorithms and a rule based decision support system are used in the CHRONIOUS project, to monitor chronic patient. We will analyse how a set of machine learning algorithms can be used in smart devices to alert the clinician in case of a patient health condition worsening trend.**

***Keywords*-Telemedicine; chronic disease management; machine learning; soft computing techniques.**

I. INTRODUCTION

This notes extend the paper [1] that was presented at ETELEMED2011 [2]. Scientific advances over the past 150 years, particularly in the medical field, allowed the extension of life expectancy in western countries and this trend seems to increase in future years. Conservative estimates suggest that by 2030 in EU countries the proportion of people over 60 years regard the entire population will be around 50%; this means that we will see a gradual increase in the number of those subjects with chronic diseases (i.e., diseases not involving healing), that will therefore increase the cost and effort over health care facilities [3].

Such chronic diseases are slowing but constantly replacing malnutrition and infection as primary causes of mortality in the population [4]. The World Health Organization (WHO) has recently emphasized that chronic diseases are a global priority [5].It was calculated that, if governments where able to put in place public health policies that produce a 2% yearly reduction in mortality rates for chronic diseases, 36 million deaths could be prevented worldwide between 2005 and 2015 [6].

Reducing mortality rates caused by chronic diseases is also an economic priority, because it could save about 10% of the loss in income due to death and disability, which amounts to $8 billion in the developing countries only [7].

Chronic diseases are difficult to treat and, apart from deaths, have collateral social impact that are becoming an economic emergency both in western and developing countries. Considering the mean age growing in western countries population, chronic diseases will be a growing emergency in next years. As the number of patient with chronic diseases is rising there will be an increasing cost for hospitalization structure both public and private. Some specific diseases like Chronic Kidney Disease (CKD), sometimes there is, during the treatment of the disease, a non-return point from where the hospitalization is continuous as for dialysed people. The traditional approach that consists in periodic check-ups and periodic lab exams seems a model that won't be sustainable as the population gets older and the total number of patients with chronic diseases rises. At present, the physician deals with an increasing number of chronic patients that are lowering the periodic check-ups and so the reduced frequency is lowering the ability to prevent, if not death, worsening in patient's quality of life.

In the latest years, we have seen a tremendous growth in IT infrastructure, both from the hardware and communication capacity. Nowadays a common mobile phone is much more powerful in terms of hardware and software capacity than the first calculating machine that allowed the man to land on the moon forty years ago. The continuously growth of the World Wide Web (WWW) and, linked to this, the continuous growth in bandwidth capacity for data transmission allows to have cheaper and more widely available bandwidth, for larger portions of the population.

As a consequence of the exponential growth of hardware and software infrastructure, it is possible to rethink the whole approach to the treatment of complex chronic disease [8] [9] by limiting the hospitalization only to situation of severe worsening of patient condition. This was the original idea behind the EU funded CHRONIOUS project [10] [11]: constructing a generic platform to monitor, in an unobtrusive way, patient with chronic disease in two goals [12]:

- Improve the patients quality of life, by reducing as much as possible the hospitalizations.
- Allow the clinician a continuous monitoring the patients, both in standard and potential risk situations.

To gain this two goals, the CHRONIOUS platform has to integrate different technologies both hardware and software modules that need to interact among themselves. This paper is organized as follows: in the first section, the general structure of CHRONIOUS hardware and software modules are described. A deep analysis of the preprocessing algorithms covers the entire second section. Section three is dedicated to illustrate the machine learning algorithms. In last section, we will evaluate possible improvements to the Chronious intelligence system.

II. THE CHRONIOUS SYSTEM: AN OVERVIEW

CHRONIOUS system deals with (Chronic Obstructive Pulmonary Disease) COPD and (Chronic Kidney Disease) CKD. The two diseases were choosen mainly because they are ones the most difficult to treat. There is lot of literature that explain in detail why, but we can summarize in a brutal way by saying that this two diseases are greatly influenced by comorbidities. For example, a patient suffering from CKD tends to have also diabetes [13] and have an increased risk of cardiovascular disease. On the other side the connection between COPD and lung cancer is so deep that some author consider them two manifestation of the same pathology [14]. Even if it is true that not every COPD patient suffers of lung cancer or pulmonary neoplasia, respiratory diseases naturally lead to a fatigue of the cardiorespiratory system with obvious consequences on the health of the heart muscle, that clearly affect all the patient status.

Different comorbidities can lead to different treatments for the patients, and for sure a continuous monitoring of the patient conditions both in terms of measure taken from the patient himself (as, for example, glucose level for diabetic patients) and in terms monitoring general status could surely take advantage over traditional therapies. So the general ideas beside Chronious was to collect different physiological data for different diseases, in a way to allow a better evaluation on well agree indicators for the clinicians. Beside this data collection that is not a novel approach in remote patient monitoring a great attention was dedicated to create a system that is able to mimic the clinician period visits to alert the clinician in a case of a worsening trend. This means that Chronious is not a life saving system but one of the main focus was to create a system that is able to alert the clinician in a way to prevent an emergency hospitalization. The thing we were trying to avoid is to have as in traditional therapies a patient that is controlled 4/5 times a year directly by the doctor, and between two check-up, an emergency hospitalization is required because for example the patient change his diet without telling his clinician. For these different types of chronic diseases, according to the medical guidelines, it is important to monitor different data in a way to check patient health status and to activate suitable emergency alarms for the clinician (for the COPD: ECG, SPO2 and respiratory rate; for CKD: glucose level, body weight and blood pressure) in case of critical event.

The CHRONIUOS platform consists of many modules that act together

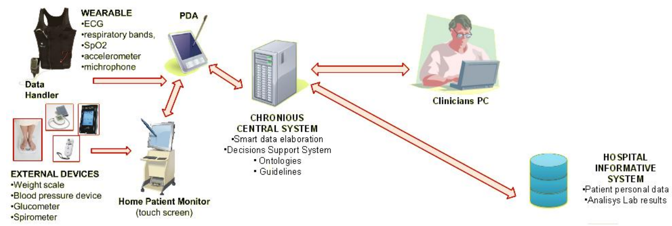

Figure 1. Chronious modules

We can organize them in three main frameworks

- Patient Sensor Framework
- Communication Framework.
- Monitor Framework.

The Patient Sensor Framework consists of hardware devices used to grab data from the patients. The hardware equipments, installed at patient's home, are

- Wearable and external devices.
- Touch screen Home Patient Monitor(HPM).
- Personal Digital Assistant (PDA).

The Wearable Device (WD) is a research project t-shirt equipped with the following sensors:

- a 3-lead Electrocardiogram (ECG).
- A microphone as a context-audio sensor.
- Two respiration bands (thorax and abdominal).
- An accelerometer.
- A sensor for measuring humidity as well as body and ambient temperature.

The external devices are a set of medical certified devices coupled with an Ambient Sensor (AS). The medical certified devices are:

- A weight scale.
- A blood pressure intake device.
- Two respiration bands (thorax and abdominal).
- A glucometer.
- A spirometer.

For the CKD patient, the devices provided are: weight scale, blood pressure device and glucometer. Every patient also interacts with a Touch screen medical PC for directly entering data for questionnaires about diet, psychological status and some question regarding the general patient status that mimic in a reduced way medical question posed by the clinician in a schedule examination(see Fig.2 and Fig.3).

All the data collected can be grouped in two types:

- Silent, when the data recording is automated and it does not involve the patient interaction, as respiratory frequency measurement for COPD patients from the wearable device

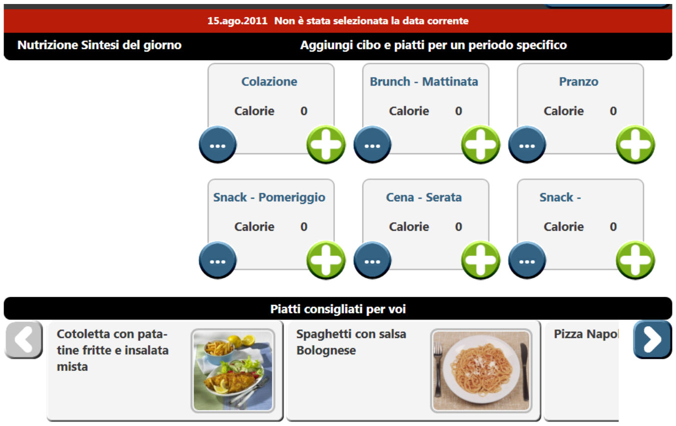

Figure 2. Diet questionnaire

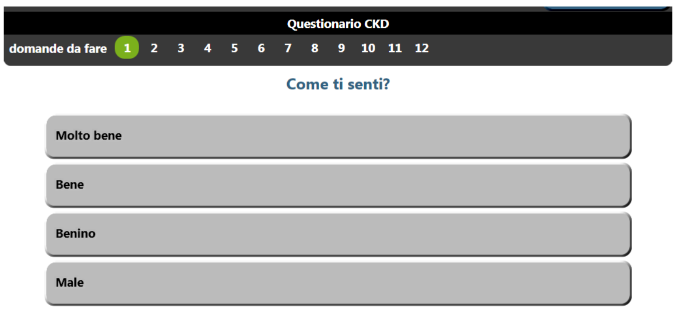

Figure 3. Doctor questionnaire

- no-silent, when it is requested a direct patient or caregiver interaction with the device, as for the blood pressure and questionnaires that are inserted on the HPM: for diet, activity and food intake monitoring. The no-silent data acquisition is particularly important for monitoring CDK patient lifestyle.

During the day there are several measurements, with different time intervals and different frequencies; only one data transmission is done, if there is no worsening in patient parameters. For COPD patient all the data are collected in a silent mode from the wearable t-shirt and transmitted to the PDA via Bluetooth; for CDK patient all the measurements, including a lifestyle questionnaire are stored in the HPM and transmitted in silent mode to the PDA. PDA is in charge of doing the following action

- Collect all the data from the devices.
- Use a set of machine learning algorithms to determinate if it is needed to force a non scheduled transmission to the Monitoring Framework.
- Transmit the collected and analysed information to the Monitoring Framework and receive back the changes, that will affect the interaction between the patient and the other devices.

The Communication Framework is in charge of transmitting data among devices and from PDA to Central DB. This transmission is done using messages in a predefined xml format. The device that is in charge of doing the transmission to the Monitoring Framework of the xmls is the PDA. Apart from the WD for which an ad-hoc proprietary binary transmission protocol has been developed, all other devices use a proprietary communication protocol. So the Communication Framework is also in charge to arrange all these difference in a uniform way. For example, for the weight scale, a proprietary software installed on the HPM is in charge of reading the measures from the weight scale. Once read the measure, the software stores it in a xml format that is different from the one used by the Communication Framework for data transmission to the Central System. Before communicating this measure, an xml transformation phase is necessary for standardizing the data an to tag it with a

device unique identifier that will be used by the Central System to associate measure to patient.

The Communication Framework is not only in charge of sending standardized xml to the central system but also once a day it connects to the Monitoring Framework to receive data from the central system. The data received are of two types:

- Frequency of measurements.
- Changes on drug intake for the patient.

Once the system is installed at patient's home, a standard set of measurements have been assigned to the patient (table I shows the frequency for CKD patient).

Table I
BASES FREQUENCY OF MEASUREMENTS FOR CKD PATIENT

Type	Day / Times per day
Glucometer	Monday once (8:00 am)
Blood pressure	Monday, Wednesday, Friday (8:00 am,5:00 pm)
Weight scale	Monday, Thursday (8:00 am)
Diet questionnaire	Monday (8:00 pm)
Clinical questionnaire	Monday (8:00 pm)

This schedule has been decided by the clinicians based on their experience in monitoring the chronic patient. But the clinician is free to change this schedule based on his experience and according to the suggestions provided by the Monitoring Framework. So, in case a patient needs to have an higher frequency of monitoring in some parameters, the clinician can decide to monitor them daily instead of the based three times per week. This allows a better monitoring of worsening trends in patient conditions. While the alerting system can transmit data from to the Central System without a schedule frequency, the data receiving phase is done once a day. As we underline above, Apart from measures the data received cover also the drug intake. Being that the Central system store also the patient drug therapy, it is possible to change it to the patient. From the patient point of view all the data transfer is silent, the only interface is a software running on the patient HPM that every day display all the measurements that need to be done during the day and a drug reminder that display which pills need to be taken (see Fig.4).

Figure 4. HPM reminders

Move from the Communication Framework to the Monitoring framework consists we have four main parts

- Clinician interface.
- Decision Support System.
- Alerting system.
- Ontology Search Engine.
- Rule based editor.
- Interoperability engine.

Even if it this not possible in this paper to deeply describe every part we will give a brief description of each one, to allow to the reader to grab the complexity of the interaction involved in the Monitoring Framework.

Starting from the simple one, the clinician interface we can describe it as the main web interface for the clinician with the patient. Every clinician has a group of patient, determinated by the disease involved. For every patient, a detail health record is visualized. The clinician can see every alert the system have made for that patient, its history and how it was treated. Every measure that the central system received is visualized and the trend are plotted. Coupled with these information the clinician have the possibility as stated above to change frequency in measurements and drug intake. We can see the main screen in Figure 5.

The Decision Support System is a rule based engine that is able to receive as input nearly 100 different parameters about COPD/CDK patient condition and trends, parse every information and outcome a suggestion based on a set of if-then rules. The Decision Support System is coded as a web-service with different methods that is called by new

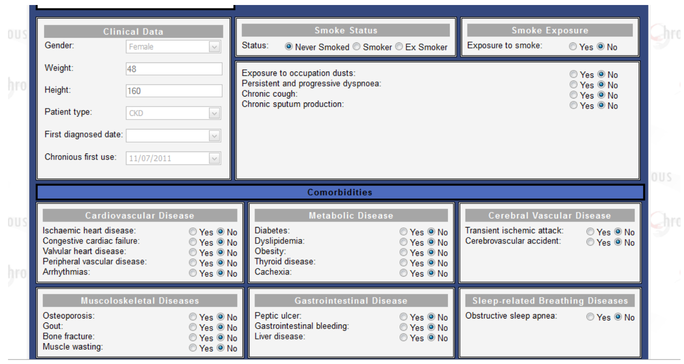

Figure 5. Clinician Interface

data insert into the Chronious Central DB. The web service use a set of JENA rules to infer a suggestion that will be displayed on the Clinician Interface. The mechanism is done using SPARQL [15] query language, on a set of xml based clinical treatment rules. As a basic example, one simple rule is the following one: if the body temperature of the patient is above 38 Celsius degrees the first aid department should be alerted for an hospitalization of the patient for monitoring purpose. The predefined set of rules codifies the clinician's expertise and the guidelines for the medical treatments of the COPD/CDK patients.

It is important to notice that even if the first triggering of an alert is done on the PDA, by the set of trained machine algorithm that will be described later, every alert is double checked by this engine that contains a larger set of rules. This because a new measure that comes to the Central System sent by the PDA could be somehow the first signal of a worsening trend in patient status. So if the PDA does the predefined one day transmission to the Monitoring Framework, even in this case the Decision Support System is called. This double check mechanism is particularly important for chronic patient because such patients can have a long period of stable conditions without any worsening but only with some fluctuations above the limits for an alert. In this situation the PDA will not reveal any worsening trend but in Decision Support System could output a suggestion on intensification of periodical check-up.

As we already pointed out, Chronious was not developed as a life saving system. However, the project faced the problem of a potential life risk situation. The alerting system is the answer to such problem. Every suggestion outputted from the Decision Support System is tagged with a level of severity. This level mimic the level used in first aid department of an hospital in case a chronic COPD / CKD request help.

These levels go from white to red, where red means that a patient is in danger of life. If the Decision Support System tags a suggestion as red the Alerting System sends an sms to a number that fast reacts and eventually go to the patient home to see what is happening. Also, technical problems are arranged by the Alerting System, being that PDA use GPRS network to send data, we can have network problem so even if the data are already stored in the PDA the Central System cannot receive them. In this case, an email is sent to the technical team. This is possible because the Central Database contains the schedules of the measurements, so it is able to understand if a measure was missed. The Alerting System is based on a queue of alerts that are continuously monitored by a windows service and every new alerts is treated according to its severity and based on a set of configurable predefined rule.

Nowadays, the World Wide Web is the primary resource for medical knowledge. Nearly every clinician use it as a source for being up to date with the latest research information on how to treat diseases. Pubmed [16] is probably the first search engine used for such purpose. Lots of information are also stored in documentation that is present only in the medical structure, so we equipped Chronious with a an ontology search engine [17] that starting from raw document uses an personalized ontology to grab meanings for the two diseases covered by Chronious [18] [19] [19]. This with the goal of having a fast way for clinician, to find information related to COPD/CKD.

The ontology was created and fed with the standard guidelines for the treatment of COPD/CDK patient. Using an upload tool it is possible to enrich the original set of documents with new ones, to make more meaningful relationships between the various symptoms (see Fig. 6). It is interesting to notice that, at the time of writing, the first comparison between simple words search in Pubmed and Chronious Search Tool seems to indicate that the ontology search gives better result according to the clinician subjective feeling. This fact is in accordance with previous studies on the subject [20].

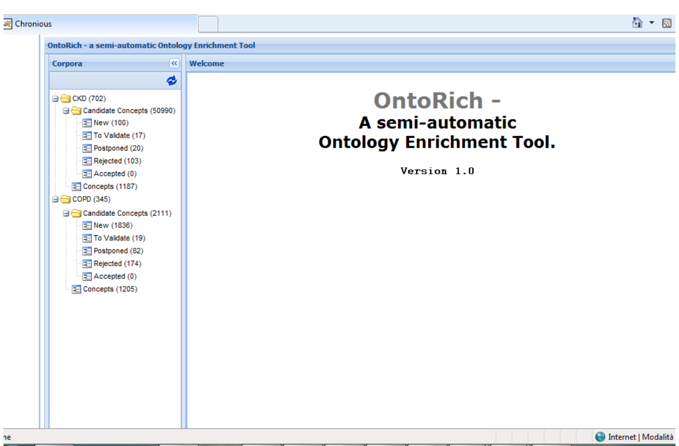

Figure 6. Ontology search

As we told the Decision Support System is a rule base engine based on a large set of if-then rules [21] [22] [23] [24]. The rules are codified using a JENA format [15] but the system gives the possibility to changes these rules according to the clinician intents. The interface that manages the rules, is the Rule Based editor that allow CRUD (Create, Read, Update, Delete) operations on the set of rules in a human readable format. The rule base editor is also equipped with a logical checker that controls if a new rule is in conflict with an existing one. Finally the Monitoring Framework is equipped with an interface written using HL7 messaging ANSI [25], to interface Chronious database with other medical structures. This was done because the medical structure use, in most cases, an RDBMS to store patient information and drug prescriptions. So this interface was needed to avoid duplicate entry by the doctor reducing the chance of errors. In the next section, we will analyse the preprocessing phase needed to activate the intelligence in the PDA that is the first decision system that activates the communication to the Monitoring Framework.

III. The PDA Chronious preprocessing phase

As we pointed out above, the Personal Digital Assistant is a smart phone equipped with WINDOWS MOBILE 6.5 Operating system, a SQL SERVER 2005 COMPACT database and a .NET FRAMEWORK 3.5. The PDA uses Bluetooth connection to send and receive data to and from HPM and Wearable t-shirt. It is equipped with a sim card for being able to do data transmission over GPRS network to the Central System. The data registered by the PDA are the following

- Data from the wearable jacket.
- Answers to questionnaires concerning dietary habits, drug intake and lifestyle of the patient.
- Data from the home patient monitor like blood pressure, glucometer measurements and body weight.

Once these data are collected, they are saved to the PDA database and a set of algorithms are triggered to analyse these data. Since the PDA analyses data for two different diseases, two sets of different algorithms are used. The fundamental data needed by COPD treatment are ECG signals and Respiration data, so in case of a COPD patient we have a first processing Electrocardiogram Pre-processing. After this, a Feature extraction phase is needed and at the end an Evaluation phase of the extracted features is done to determinate if an alert must be triggered to the Central System. For external devices used in particular by CDK patients, there is no need of a preprocessing phase because fundamental measures are the ones provided by the glucometer, the weight scale and the blood pressure measure; so, they are discrete time and directly used by the set of machine learning. Combined with these data, the answers to a set of queries concerning food intake, drug intake, lifestyle and mental status are passed to a set of machine learning algorithms to evaluate the whole patient condition. In the next subsections, we will analyse first the COPD set of algorithms used.

A. Preprocessing of COPD signals

The aim of Preprocessing Phase is to improve the general quality of the ECG, for more accurate analysis and measurement, because there's the possibility to have some noises on the signals. Possible noises in the signal include

- Low frequency Base Line Wandering (BW) caused by respiration and body movements.
- High frequency random noises caused by mains interference (50 or 60Hz).
- Muscular activity and random shifts of the ECG signal amplitude caused by poor electrode contact and body movements.

The preprocessing comprises:

- Removal of base line wandering.
- Removal of high frequency noise.
- QRS detection.

The BW which is is an extrogenous low-frequency activity which may interfere with the signal analysis, rendering its clinical interpretation inaccurate and misleading. Two major techniques are employed for BW removal:

- Linear filtering [26]: involves the design of a LTI high pass filter with cut off in way that the clinical information in the ECG is preserved and the BW is removed as much as possible.
- Polynomial fitting [27]: includes the fitting of polynomials to representative points (knots) in the ECG, with one knot for each beat. Knots are selected from a silent segment, e.g., the PQ interval. A polynomial is fitted so that it passes through every knot in a smooth fashion.

The High Frequency Noise can be caused by the high frequency as well as power supply interference from the ECG signal. Its removal is done using:

- The Daubechies (DB4) wavelet employed on the basis of the resemblance and similar frequency response characteristics of the db4 basis function with the ECG waveform.
- Using wavelets to remove noise from a signal requires identifying which components contain the noise, using optimal methods to threshold them, and then reconstructing the signal using the thresholded coefficients.

The prepocessing phase finally deals with the QRS detection. The main features that should be calculated: the Inter-beat (RR) interval and the Heart Rate Variability (variation in the beat-to-beat interval). For the Inter-beat (RR) interval, two methods have been explored

- Filtering the ECG signal with continuous (CWT) and fast wavelet [28] transforms (FWT)[1].
- Following Pan-Tompkins [29], wavelets are used to remove noise from a signal requires identifying which component or components contain the noise, using optimal methods to threshold them, and then reconstructing the signal using the thresholded coefficients[2].

All the previous features are extracted from ECG signals. For COPD patient also the Respiratory Rate is a fundamental parameter that need to be analysed. In order to calculate the respiration rate using the reference respiration signal, a dominant frequency detection algorithm, based on short-time Fourier transform (STFT) [30], is applied.

The STFT is a localized Fourier transform, utilizing a Hamming window:

$$STFT(f(t)) = STFT(\omega, \tau) = \int_{\infty}^{\infty} f(t)w(t-\tau)e^{-j\omega t}dt \tag{1}$$

were $w(t)$ is the window function, commonly a Hann window or Gaussian hill centered around zero, and f(t) is the signal to be transformed. Because frequency components of the respiration signal are very low (2Hz), a window size of 60 seconds is selected. Every 60s, the hamming window is multiplied to the respiration signal, and the result is transformed to the frequency domain using Fourier transform. The dominant frequency is then detected by finding the maximum amplitude of the spectrums. When the dominant frequency components are found, inverse numbers are calculated in order to obtain the respiration rate. After this first preprocessing phase for COPD patients we wil now analyse the which kind of Features are extracted.

B. Features Extraction for COPD patients

From the Inter-beat (RR) interval and the Heart Rate Variability, several features can be extracted, either in time or in frequency domain.

Dealing with Time domain the values extracted are

1) SDNN(msec): Standard deviation of all normal RR intervals in the entire ECG recording using the following

$$sdnn = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(NN_i - m)^2} \tag{2}$$

[1]The reconstructed ECG signal after denoising contains only spikes with non-zero values at the location of QRS complexes. From this signal, the PQ junction and J point can be located as the boundaries of the spike. If the length of the spike is more or less than a predefined QRS length range it is annotated as noise and if the voltage is below a certain threshold, it is annotated as an artifact. The next stage is the detection of the T wave, and the P wave in the PQ interval. The peaks of Q, R and S waves are identified in the annotated part of the ECG signal from the PQ junction to J point.

[2]The algorithm includes a series of filters and methods that perform lowpass, high-pass, derivative, squaring, integration, adaptive thresholding and search procedures.

where NN_i is the duration of the i-th NN interval in the analysed ECG, n is the number of all NN intervals, and m is their mean duration.

2) SDANM(msec): Standard deviation of the mean of the normal RR intervals for each 5 minutes period of the ECG recording.
3) SDNNIDX (msec): Mean of the standard deviations of all normal RR intervals for all 5 minutes segments of the ECG recording.
4) pNN50 (intervals that are greater than 50 msec, computed over the entire ECG recording.
5) r-MSSD (msec): Square root of the mean of the sum of the squares of differences between adjacent normal RR intervals over the entire ECG recording the formula is

$$rMSSD = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(NN_{i+1} - NN_i)^2} \quad (3)$$

where NNi is the duration of the i-th NN interval in the analysed ECG and n is the number of all NN intervals.

If we now move to the frequency domain the Feature Extraction on the PDA studies two bands:

1) The Low Frequency band (LF), which includes frequencies in the area $[0.030.15]$ Hz.
2) The High Frequency band (HF), which includes frequencies in the area $[0.150.40]$ Hz.

If we now move to the Respiration signal several features can be extracted either directly or indirectly we focused on:

1) Respiration Rate: The number of breaths per minute.
2) Tidal Volume (VT): The normal volume of the air inhaled after an exhalation.
3) Vital capacity (VC): The volume of a full expiration. This metric depends on the size of the lungs, elasticity, integrity of the airways and other parameters, therefore it is highly variable between subjects.
4) Residual volume (VR): The volume that remains in the lungs following maximum exhalation.

After all the preprocessing phase of the data gathered by the wearable devices, all these information are passed to a Classification System. The classification system is responsible for the analysis of the outcome from the preprocessing phase for the COPD patients and of the data gathered by the external devices and questionnaires entered by the patient himself. Below, we will see how the software is used to transform all these rich set of data in an information to be, in case, transmitted real-time or scheduled to the Monitoring Framework.

IV. The CHRONIOUS Classification System

After the collected data have been preprocessed for COPD patient and all the CKD patient input have been acquired, a set of machine learning algorithms are fired up to decide if a potentially risky situation is present. The aim of these tools is alerting the Central System that contains a rule based decision support system, for a better evaluation of the message triggered by the PDA. In case the message containing an alarm for life risk danger, the Central Decision Support System is able to alert the emergency staff or to suggest the clinician to modify the therapy approach. Most of these tools need a preprocessing phase for identifying the correct parameters that need to be validated by the clinicians. This means that in the first validation of the CHRONIOUS project a large amount of efforts has been dedicated to gather feedback from the clinicians about the correctness of the rules / parameters that have been inferred by the algorithms. In thise phase, another important effort has been dedicated by the technicians to evaluate some probability index for fuzzy data measures.

The CHRONIOUS Classification System is composed of the following part:

- A light rule based expert system.
- A supervised classification system.

The light rule based expert system is an xml parser that is able to extract from an xml a set of "'if then"' rules created an validated by the clinician. With these rules combined with the data collected, the rule base system is able to decide if a patient is in a potentially life-risk situation. For example the following rules will generate an immediate alarm to the central system:

- The Hearth Rate is above 120 bpm for both COPD and CKD patient.
- If the weight increase by 2% in the last 24 hour for CDK patient .

These rules are most for alarm triggering. It means that they aren't use for light monitor alerting. In the CHRONIOUS PDA system the Supervised Classification System is composed of the following machine algorithms:

- Support Vector Machines [31]
- Random Forest [32]
- Multi-Layer Perceptron [33]
- Decision Tree [34]
- Naïve Bayes [35]
- Partial decision Trees [36]
- Bayesian Network [37]

Apart from Bayesian Network, the other algorithms have been trained with a dataset to generate a set of rule that have been validated by clinician. The Bayesian Network have been used to identified possible rules about the mental and stress evaluator of the patient. This means that once trained, the rules generate can be used to identify is some stressful condition can alter some parameter and leading to a worsening of the general state of the patient. The use of these algorithms was needed because for the CKD patient the diet covers the most part of the medical treatment, so any factor that can influence a change on the diet intake, would potentially and indirectly lead to a worsening of the patient condition. For example, if a female CKD patient feeld sad, these condition could lead her to eat a bigger piece of pie for satisfaction purpose. In general a bad feeling could lead chronic patients to be uncompliance with the medical treatment. However the kind of rules aren't liked the vital signs so their fuzzyness could be identified by these type of algorithms. Clearly while diet is important to avoid worsening on CDK patient conditions, on the other side the lifestyle could be and indirect cause of COKD worsening condition. Nevertheless for both of them we need static rules to have alarms sending, because for both for example having a body temperature above 38 could be a risky situation were an hospitalization is needed for both COPD and CKD patients. Considering the training dataset a set of 41 attributes have been identified. These data comes from the 2 sets of 2 hours health recordings (11 attributes), food input module (12 attributes), drug intake module (1 attribute), activity input module (2 attributes), questionnaire (13 attributes) and external device (2 attributes). The results of some classifier are shown in table II.

Table II
CLASSIFICATION SYSTEM: MAE: MEAN ABSOLUTE ERROR, RMSE: ROOT MEAN SQUARED ERROR, RAE : RELATIVE ABSOLUTE ERROR, CI: CORRECTLY/INSTANCES

Method	MAE	RMSE	RAE	CI
PART	0.1336	0.312	57.81 %	2.67
J48	0.1336	0.321	57.81 %	2.67
Forest	0.2	0.341	86.52 %	1.75
Naïve	0.127	0.343	54.95 %	2.67

Again we point out that even if there are some errors due to false positive matches, the PDA system in these case would only generate a rule that will send a message to the clinician that most of the times would only say that a light worsening is present. The core intelligence that deals with the central system would be the real suggestion system that would indicate to clinician a suggestion on how to act to possibly revert the trend. The Bayesian Network is the fundamental algorithm for the Mental support tool. It uses a set of attributes that affect a stress index and they weight based on the clinician's feedback as shown in table of Figure 2. When the total stress indicator is above a certain value a light alarm is triggered to the Central Database to inform clinician of a potentially worsening of patient conditions. In the same way a module in the PDA is in charge of the rules concerning the lifestyle od the patient: the lifestyle tool. It collects data entered by the patient or caregiver in a validation phase and using a Bayesian Network is able to compute an index of good or poor lifestyle of the patient also in this case if the poor lifestyle is find a light alarm is sent to the clinician for monitoring purpose.

During the first validation phase the training phase as been reprocessed several times as new data for the patient

Attribute	Different States	Probability of causing Stress (%)
Smoking	YES	90
	NO	10
Environmental Noise and rowded/Noisy places	High	70
	Medium	25
	Low	5
Hypoglycaemia	YES	85
	NO	15
Heart Rate	High	85
	Normal	15
Skin Temperature	Cool	35
	Sweat	65
Breathing asynchrony	YES	90
	NO	10
Sleep Disturbances (Questionnaires)	YES	62.5
	NO	37.5
Mood (Questionnaires)	Better	5
	Same	25
	Worst	70
Activity Comments	Feel sick, nausea	18
	Exhaustion, fatigue	23
	Discomfort in the chest, upper body, or jaw	23
	Irregular or extremely rapid heart beats	28
	None	8

Figure 7. Attributes

were collected. In fact, the knowledge that can be extracted using machine learning algorithms can be increased as new patients use the Chronious System.

V. CONCLUSIONS AN IMPROVEMENTS

In this paper, we presented a set of machine learning algorithms store in a smart phone that, combined with some external devices and patient specific data, can be used for a first monitor/alert system for treatment of patient affected by chronic diseases. Dealing with telemedicine application, these kinds of software, could help to improve patient quality of life and could be also a valid help for clinician to allow a more precise monitoring of patient conditions without need of the physical presence of the clinician. Apart these potentially advantages a PDA that equipped with these kinds of applications could suffer of some limitations. During developing phase we faced these problems:

- Heavy resource consumption of preprocessing algorithms.
- Updating a trained supervised algorithm.

The preprocessing algorithm for COPD parameter denoising is the most memory/CPU consumption. This can became a potentially problem when we deal with life risk situations, because while the algorithm denoise the ECG signal, the patient can loose consciousness and so precious time can be lost in these phase. Other time is lost due to the huge amount of signals transmitted from the wearable, this because we deal with an ECG signal that is composed of a mean of ten measurements per second so this means that 5 minute of ECG signal became nearly 3000 sql commands to a SQL SERVER 2005 COMPACT that is not so performing in this case. Increasing hardware requirements of the PDA can be a first solution however it would be interesting to understand if a relational database is the best solution for storing these types of data or other structure would be more conformable for storing purpose.

PDA algorithms were developed using a Microsoft technology that at the time of writing have been exceeded by the cloud computing paradigm. For future purpose, having a compact RDBMS installed on the PDA, could became be an unnecessary requirement. The data collected could be transmitted directly to a Cloud storage and all the training algorithms could be moved on the cloud.

The other important issue is that once the supervised learning algorithms are trained any little change in some parameter will need to retrain the algorithms and most important, it would need a new validation of the outputs by the clinicians. This retrain can lead to difficulties when after this phase, the new trained algorithms need to be updated on the pda. The Chronious Communication Framework is only able to transmit values from the PDA to Monitor Framework and back, but cannot transmit back the trained configuration for the machine learning algorithms. In this case an interesting solution could be also to allow remote updating of the structure validated. For example a trained neural network could read the weights matrix from a structure upgradable by the Communication Framework. Apart from these improvements, and many others that could lead to a software system closer as much as possible to the clinician and patient needs, it is our opinion that with the smart devices that are closer to a normal PC, the algorithms presented in this paper could become an important part of the telemedicine platforms.

Acknowledgment

A reduced version of this paper was presented during [2] in the wonderful scenario of Le Gosier, Guadeloupe in the French Caribbean. The author that had the privilege to present his first research paper in such this, out of common world country, wishes to thank the IARIA organization for the unforgettable experience.

References

[1] P. Giacomelli, G. Munaro, and R. Rosso, "Using soft computer techniques on smart devices for monitoring chronic diseases: the chronious case," February 23-28,2011, The Third International Conference on eHealth, Telemedicine, and Social Medicine, eTELEMED2011, Gosier, Guadeloupe, France, pp. 77-82, IARIA XPS Press, ISBN: 978-1-61208-119-9.

[2] "ETELEMED2011 web site," http://goo.gl/4HoLj, [retrieved: Dec, 2011].

[3] C. Zoccali, A. Kramer, and K. Jager, "Chronic kidney disease and end-stage renal disease a review produced to contribute to the report the status of health in the european union: towards a healthier europe," *NDT Plus*, vol. 3, no. 2, pp. 213–224, 2010.

[4] A. R. Omran, "The epidemiologic transition: A theory of the epidemiology of population change," *Milbank Quarterly*, vol. 83, pp. 731–757, 2005.

[5] W. H. Organization, "Preventing chronic diseases: a vital investment," http://goo.gl/umxCk, [retrieved: May, 2012].

[6] K. Strong, C. Mathers, S. Leeder, and R. Beaglehole, "Preventing chronic diseases: how many lives can we save?" *Lancet*, vol. 366, pp. 1578–1582, 2005.

[7] D. O. Abegunde, C. D. Mathers, T. Adam, M. Ortegon, and K. Strong, "The burden and costs of chronic diseases in low-income and middle-income countries," *The Lancet*, vol. 370, no. 9603, pp. 1929–1938.

[8] L. Bartoli, P. Zanaboni, C. Masella, and N. Ursini, "Systematic Review of Telemedicine Services for Patients Affected by Chronic Obstructive Pulmonary Disease (COPD)," *Telemedicine and e-Health*, vol. 15, no. 9, pp. 877–883, Nov. 2009.

[9] B. McKinstry, H. Pinnock, and A. Sheikh, "Telemedicine for management of patients with copd?" *Lancet*, vol. 374, no. 9691, pp. 672–673, 2009.

[10] "Chronious official web site," http://www.chronious.eu, [retrieved: Oct, 2011].

[11] M. van der Heijden, B. Lijnse, P. Lucas, Y. Heijdra, and T. Schermer, "Managing copd exacerbations with telemedicine," in *Artificial Intelligence in Medicine*. Springer Berlin / Heidelberg, 2011, vol. 6747, pp. 169–178.

[12] M. Vitacca, L. Bianchi, A. Guerra, C. Fracchia, A. Spanevello, B. Balbi, and S. Scalvini, "Tele-assistance in chronic respiratory failure patients: a randomised clinical trial," *European Respiratory Journal*, vol. 33, no. 2, pp. 411–418, 2009.

[13] A. J. Collins, J. A. Vassalotti, C. Wang, S. Li, D. T. Gilbertson, J. Liu, R. N. Foley, S.-C. Chen, and T. J. Arneson, "Who Should Be Targeted for CKD Screening? Impact of Diabetes, Hypertension, and Cardiovascular Disease," *American Journal of Kidney Diseases*, vol. 53, no. 3, pp. S71–S77, Mar. 2009.

[14] T. L. Petty, "Are copd and lung cancer two manifestations of the same disease?*," *Chest*, vol. 128, no. 4, pp. 1895–1897, 2005.

[15] "w3c sparql reference," http://www.w3.org/TR/rdf-sparql-query/, [retrieved: Aug, 2011].

[16] "Pubmed official site," http://www.pubmed.org, [retrieved: Sep, 2011].

[17] S. Kiefer, J. Rauch, R. Albertoni, M. Attene, F. Giannini, S. Marini, L. Schneider, C. Mesquita, and X. Xing, "An ontology-driven search module for accessing chronic pathology literature," in *On the Move to Meaningful Internet Systems: OTM 2011 Workshops*, P. H. R. Meersman, T. Dillon, Ed. Hersonissos, Crete, Greece: Springer Verlag, 2011, pp. 382–391.

[18] S. Kiefer, J. Rauch, R. Albertoni, M. Attene, F. Giannini, S. Marini, L. Schneider, C. Mesquita, X. Xing, and M. Lawo, "The chronious ontology-driven search tool: enabling access to focused and up-to-date healthcare literature," in *eChallenges e-2011 Conference*, M. C. P. Cunningham, Ed. IIMC International Information Management Corporation, 2011, pp. 1–8.

[19] P. Giacomelli, G. Munaro, and R. Rosso, "Can an ad-hoc ontology beat a medical search engine? the chronious search engine case." January 30- February 04,2012, The Fourth International Conference on eHealth, Telemedicine, and Social Medicine, eTELEMED2012, Valencia, Spain, pp. 215-220, IARIA XPS Press, ISBN: 978-1-61208-179-3.

[20] J. Z. Wang and F. Ali, "An efficient ontology comparison tool for semantic web applications," *Web Intelligence, IEEE / WIC / ACM International Conference on*, vol. 0, pp. 372–378, 2005.

[21] E. Seto, K. J. Leonard, J. A. Cafazzo, J. Barnsley, C. Masino, and H. J. Ross, "Developing healthcare rule-based expert systems: Case study of a heart failure telemonitoring system," *International Journal of Medical Informatics*.

[22] G. Kong, D.-L. Xu, R. Body, J.-B. Yang, K. Mackway-Jones, and S. Carley, "A belief rule-based decision support system for clinical risk assessment of cardiac chest pain," *European Journal of Operational Research*, vol. 219, no. 3, pp. 564–573, 2012.

[23] S. M. Maviglia, R. D. Zielstorff, M. Paterno, J. M. Teich, D. W. Bates, and G. J. Kuperman, "Automating complex guidelines for chronic disease: Lessons learned," vol. 10, no. 2, pp. 154–165, 2003.

[24] J. Chang, C. Ronco, and M. H. Rosner, "Computerized decision support systems: improving patient safety in nephrology." *Nature reviews. Nephrology*, vol. 7, no. 6, pp. 348–355, Jun. 2011.

[25] "HL7 official site," http://www.hl7.org, [retrieved: Jan, 2012].

[26] A. Kachenoura, F. Poree, G. Carrault, and A. I. Hernandez, "Comparison of four estimators of the 3d cardiac electrical activity for surface ecg synthesis from intracardiac recordings," in *Proceedings of the 2009 IEEE International Conference on Acoustics, Speech and Signal Processing*, ser. ICASSP '09, 2009, pp. 485–488.

[27] M. Kaur, B. Singh, and Seema, "Comparison of different approaches for removal of baseline wander from ecg signal," in *Proceedings of the International Conference & Workshop on Emerging Trends in Technology*, ser. ICWET '11. ACM, 2011, pp. 1290–1294.

[28] Y. Chesnokov, D. Nerukh, and R. Glen, "Individually adaptable automatic qt detector," *Computers in Cardiology*, no. 2, pp. 337–340, 2006.

[29] J. Pan and W. J. Tompkins, "A real-time QRS detection algorithm." *IEEE Trans Biomed Eng*, vol. 32, pp. 230–236, 1985.

[30] P. S. Addison, "Wavelet transforms and the ecg: a review," *Physiological Measurement*, vol. 26, no. 5, p. R155, 2005.

[31] N. Cristianini and J. Shawe-Taylor, *An introduction to support vector machines and other kernel based learning methods*. Cambridge University Press, 2000.

[32] L. Breiman, "Random forests," *Machine Learning*, vol. 45, pp. 5–32, 2001.

[33] S. Haykin, *Neural Networks and Learning Machines (3rd Edition)*. Prentice Hall, 2008.

[34] J. R. Quinlan, "Learning decision tree classifiers," *ACM Comput. Surv.*, vol. 28, pp. 71–72, 1996.

[35] N. Friedman, D. Geiger, and M. Goldszmidt, "Bayesian network classifiers," *Machine Learning*, vol. 29, pp. 131–163, 1997.

[36] E. Frank and I. H. Witten, "Generating accurate rule sets without global optimization," in *Proceedings of the Fifteenth International Conference on Machine Learning*, 1998, pp. 144–151.

[37] D. Holmes and L. C. Jain, *Innovations in Bayesian Networks Theory and Applications*. Springer Netherlands, 2008.