Application of Data Mining Techniques to Efficiently Monitor Chronic Diseases Using Wireless Body Area Networks and Smartphones

Ioannis Kouris*, Dimitris Koutsouris

National Technical University of Athens, 9 Heroon Polytechniou str., 15773 Zografou, Athens, Greece

*Corresponding Author: ikouris@biomed.ntua.gr

Copyright © 2013 Horizon Research Publishing All rights reserved.

Abstract This paper presents a wireless body area network platform performing daily physical activity recognition using accelerometers, biosignals and smartphones. Various classifiers have been evaluated to identify the one with the best recognition results. Functional Trees classifier provided the best performance and was used in the real time activity recognition that executed on the smartphone. Geo-location provided by the GPS receiver of the smartphone used to retrieve location based environmental data and Points of Interest via the web. Activity recognition results and environmental data were stored in a database and a cloud-hosted application performed Emerging Patterns search through the data to predict future conditions. The described framework has application in the prevention of short-term complications of metabolic diseases such as diabetes or environmental conditions related diseases such as Chronic Obstructive Pulmonary Disease (COPD).

Keywords Data Mining, Wireless Body Area Networks, Smartphones, Activity Recognition, Open Source Hardware

1. Introduction

The reduction of physical body activity is reported by the World Health Organization (WHO) as the cause of more than 3.2 million deaths per year [1]. A sedentary way of living results in fewer calories per day to be expended and less exercise to be performed. Those factors increase the likelihood of the appearance of metabolic diseases which are related to the reduced physical activity. Increase in the appearance of chronic diseases such as type-2 diabetes [2], obesity [3], cardiovascular diseases [4], coronary disease and even some types of cancer [5] have been reported by the scientific community due to the absence of adequate body physical activity [6].

Sufficient body physical activity is a prerequisite for a

healthy life. Physical activity is widely considered as a pillar for the prevention of type-2 diabetes and cardiovascular diseases (physical activity helps to improve cardiovascular health), as people that are insufficiently physically active are at an increased risk level of developing chronic diseases compared to individuals who keep on having regular physical activity [7]. One of the main challenges in the management of chronic diseases is the efficient and continuous monitoring of the patient's health by the health professionals. In that sense, the patient-doctor interaction is preferred to not happen in a monthly basis scheme, but it can be continuous by using telemedical-telemonitoring tools. This enhances the security of the patient himself, while offers the doctors a continuous and subjective overview of the disease evolution along with the patient's behavior in relation to the medical advices and the prescribed medication. Technology assisted patient monitoring provided via wearable sensors and portable devices (smartphones) provide more accurate and reliable data, acquiring non-invasively data related to the physical activity of the user.

Miniaturization of the electronic sensors has allowed their embedding in multiple electronic devices, such as the smartphones. Micro-Electro-Mechanical Systems (MEMS) attached to the body via straps or sewed in the garment [8] can be used to monitor, the motion of the user, in terms of activity type, duration and intensity and evaluate the conformance on the medical advices. Activity Recognition using MEMS is utilized in many medical or non-medical applications such as in the rehabilitation [9], in the fall detection [10], in the quantification of the body movements of individuals suffering of Parkinson's disease [11] or even for the athletic performance monitoring and improvement [12]. Studies focusing on motion analysis using the stride length, the walking speed or the movement analysis of the lower extremities, measure the momentums of the body parts, using MEMS sensors [13]. Optical and MEMS sensor comparison for human motion analysis has also been evaluated [14]. The usage of MEMS sensors for motion

monitoring and analysis enables an inexpensive way to acquire data under free-living conditions.

Human motion analysis can be performed using commercially available sensors, customized design sensors or the smartphone's embedded sensors. An application where a palmtop computer and a commercial monitoring platform [15] used to record the motion (dual-axis accelerometers), the temperature, the humidity and the light intensity, along with ECG/EMG transmitted via ZigBee to a desktop computer for motion analysis has been tested [16]. Sensors connected via Bluetooth to a mobile phone in order to identify activities such as ascending and descending stairs, walking, running or standing has been also examined [17]. Sensors measuring the pressure of the feet to the sole of a shoe were transmitted to a palmtop computer to study human gait [18]. By just using the smartphone's sensors, recognition of the activities of walking, running, cycling, driving and jumping has been performed with the smartphone attached on the waist of the user and an application running on the smartphone, written in Java code, to calculate burnt calories and detect the performed activities [19].

2. Objectives

This study explores the new capabilities offered by the wearable devices (sensors) in cooperation with the portable devices (smartphones) to accurately monitor physical daily activity and apply new techniques to the prediction of complications of chronic diseases. The main components of the system are: i) the wireless body area sensor network (using open-source microcontrollers and boards) which are wirelessly connected to a smartphone via Bluetooth. User body movements and accelerations, along with biosignals (heart rate) are acquired by the MEMS sensors of the system in order to

perform real-time activity recognition (execution of motion analysis algorithms is performed on the smartphone), ii) the smartphone application running on an Android™ device, used for the initial training of the system and then to execute the model of the algorithm that provides the best prediction results and iii) a cloud-hosted application to perform high processing power statistical calculations to predict disease complications in the short- or the long-term. The prediction is estimated via the evaluation of the physical activities performed, the environmental conditions, the past activities (in relation to the places where performed) and the severity of the disease-related symptoms experienced by the patient. Multiple data mining algorithms have been evaluated for the selection of the one with the best result that is used by the smartphone application and then Emerging Patterns algorithm [20] is applied to identify similarities between current and past patterns related to the appearance of the complications.

3. Methods and Tools

3.1. Telemonitoring Applications Focusing on Chronic Diseases

Each chronic disease requires different parameters to be monitored by a telemedical application. Diabetes self-management requires monitoring of the blood glucose levels, insulin type and dosages, levels of activity, dietary information regarding food portions and carbohydrate equivalents contained in each food. The reaction of the body in regards to the food consumption, the insulin dosages and the calorie expenses having exercise are used to predict the future needs of the diabetic, in an effort to simulate an artificial pancreas [21, 22]. Activities' intensity evaluated by a smartphone application along with monitoring the diabetics' blood glucose levels, insulin injections and exercise has allowed the medical personnel to remotely monitor and modify medication of their patients [23].

The telemonitoring of COPD patients requires a system to measure blood oxygen saturation, lung sounds and discomfort of the patients. Such a system has been evaluated by Foix et al., where a dedicated system for COPD patients has been developed [24]. This system reduces the need of frequent doctor visits as it uses a patented sensor to capture and analyze lung sounds in daily basis. A medically approved questionnaire is also answered by the patient, so that the doctors to assess the need for medication schemas changes to prevent hospitalization. The telemonitoring system is supported by a web-based Electronic Health Record platform, accessible via personal computers or smartphones, allowing the medical personnel to access patients' data virtually from everywhere [25, 26].

Patients suffering from Parkinson's Disease are efficiently monitored via a prototype wearable system that monitors upper and lower extremities' along with the body movement [27]. The system uses four sensors placed on each extremity, on the waist and on the chest to record accelerations and angular velocity, so that to assess the type of the symptom and the severity. The collected data are transmitted via cellular networks to the decision support system that performs signal analysis and extract the results. Relation between the medication intakes and the appearance of the involuntary body movements due to the side effects of the drug are observed by a dedicated computer interface for the doctors and medication changes are provided to the patients [11, 28].

It is clear that activity recognition has applications in the risk assessment of diabetes, COPD, Parkinson's disease and other diseases. In that sense, automated activity recognition [29] using data from accelerometers and ECG, real-time activity recognition using wireless accelerometers and heart rate monitors [30] or mobile phones in various pocket positions [31] have been examined. Some of those studies perform continuous activity recognition [32] recorded by accelerometers, while others examine accuracy of the physical activity measurement and energy expenditure using multiple sensors [33]. Jog Falls [23] provides a platform for diabetes management, through a system that performs

activity recognition (accelerometer and heart rate data) to estimate energy expenditure and promote activity goals. Gyroscopes and accelerometers have been also used to analyze motion patterns in older adults [34, 35], while in another study accelerometer data from RFID readers have been used to recognize activities and correlate them with calorie intake [36].

3.2. Wireless Body Area Network for Activity Recognition

The goal by the development of our Wireless Body Area Network (WBAN) for activity recognition was to develop a customizable wireless network using open-source components. Following that path, we had the freedom to test various components (accelerometers, wireless modules etc) and select the ones that provided the best results, in regards of accuracy, power efficiency and purchase cost. Additionally, the weight was kept as low as possible, enhancing user acceptability. If a commercial sensor network had been selected, there would not have been enough parameters to customize to enable the handling of the transmitted data. The Atmel 328P microcontrollers that were used were loaded with the Arduino™ bootloader and were programmed using a dedicated language similar to C++. Via the code that was uploaded on each sensor, the sensitivity, the sampling rate, the transmission protocol and the calculation of the heart rate was performed before the transmission of the data to the smartphone.

The WBAN was comprised of two sensors, one placed on the waist, one on the lower extremity (shank) and one around the chest (heart beat monitor). The later sensor was the Polar Wearlink®+ sensor which used to provide heart beat detection. This sensor uses an elastic strap with washable electrodes to record the heart beat potential of the heart and wirelessly transmit a 3ms pulse at the frequency of 5.5 kHz, upon each heart beat detection. Each WBAN sensor had similar hardware specifications, although the sensor that was placed on the waist was equipped with an extra electronic component (RMCM01) to detect the pulses transmitted by the chest heart rate monitor. This component placed on the waist sensor, as the manufacturer claims that the pulses of the Polar Wearlink®+ sensor are transmitted to a distance up to 80cm (most common use of these sensors is to monitor the heart beating during aerobic exercise via a Polar wristwatch). The base board selected to develop the sensors was the Arduino Fio board. This board is based on Arduino ATmega328P microcontroller which operates at 3.3V and runs at 8 MHz. Fourteen digital and analog ports are available to connect hardware components. We developed the sensors with the following criteria: i) the board should be small in size; this board is slightly bigger than Arduino Nano board, although ii) offers a base to directly install an XBee [37] or Bluetooth [38] modem and iii) embeds a Lithium Polymer battery charger. This last feature makes this board very practical for everyday use, as the sensors' batteries can be recharged via a USB cable. The accelerometers that used

to monitor the movement were 3-axis accelerometers (ADXL335, $\pm 3g$), one on each sensor, which offered high resolution allowing the capture of abrupt and fine movement of the body. The battery used on each sensor had 1Ah capacity, but was lightweight and slim. The overall weight of each sensor was kept under 50 grams (with the battery included) and the dimensions were 6.8 x 3.0 x 1.0 cm (Figure 1).

Figure 1. WBAN sensor without the casing. The board with the Arduino microcontroller, the accelerometer and the battery are shown

The interconnection of the sensors to the smartphone performed via a Bluetooth modem installed on the XBee base of each Arduino Fio board. In our effort to develop a power efficient system, we tested various sampling rates of the accelerometer measurements and various transmission speeds of the Bluetooth modem. We concluded that if a sampling rate of 20 Hz was selected, it allow the capture of slow movements (when the body remains still – sitting, laying on a bed) or when the person performs more vigorous movements, such as running or cycling in high speed. As long as the transmitted data were less than 1Kbps, the lowest supported speed of the selected Bluetooth modem was chosen to further reduce battery consumption (9600bps baud rate). Bluetooth protocol preferred over ZigBee as there were not available smartphones that implement this protocol. The succeeded autonomy of the sensors was 36 hours, translating in three-day usage (12 hours per day) with a single charge.

Each sensor must initially be paired with the smartphone. The smartphone application that used to connect the sensors and collect the transmitted data will be described in the next chapter. Upon successful pairing of the sensors (password protected), the data packets transmitted to the smartphone contain the absolute value of the difference between the current and the past accelerations on each axis. An identification number is used along with each transmission to differentiate the data of the waist or the shank sensor. Additionally, the waist sensor transmits the heart rate detected every ten seconds. Heart beat rate is calculated counting the time between two consecutive 5.5 KHz pulses detection and averaged every 10 seconds.

3.3. Smartphone Application

The WBAN sensor network was connected to a

smartphone running Android™ operating system. An application developed using Java code to enable the pairing between the WBAN sensors and the smartphone, along with the storage of the transmitted data. To collect data and perform pattern recognition by testing multiple algorithms, annotation of the activities performed is selected via the smartphone application. Each activity performed was registered via the application interface (Figure 2). It was also possible to check if the sensors have been connected successfully, if the heart rate detected and the status of the GPS signal.

Figure 2. Smartphone application screenshot. The main functionalities of the application are available for the user.

After data collection phase, the smartphone application included the code of the trained model for real-time activity recognition. The trained model uses a portable version of the data analysis software (WEKA [39]) modified for execution on Android smartphones. The device used to run the real-time activity recognition was a dual core processor 1.5 GHz Sony Ericsson Xperia S smartphone. Activity recognition was not continuous but performed every 15 seconds. Analyzing each batch of data every 15 seconds preserved battery, without affecting the exported results. The result of the activity recognition stored in the database and mixed with context data collected related to the location of the user (see next paragraph for more details).

The signal from the GPS receiver provided data regarding the location of the user (used to retrieve context-aware information), the elevation and the speed. Due to the great variation of the GPS accuracy, information provided by the sensor was qualified as valid only when the accuracy was less than 6 meters. The speed information was then used to classify the activities in different classes (based on their intensity - Table 1) and the location information was used to calculate the distance of the person from the surrounding Points of Interest (POIs). The POIs were accessible via an open-source database provided by OpenStreetMap (under Creative Commons Attribution Share-Alike 2.0 license) [40]. User location was not updated continuously, in an effort to reduce battery consumption, but user location was updated every 10 minutes. After location update, country, county and city were detected and POIs in a radius up to 500 meters

from the location of the user were retrieved. For our application only places such as theme parks, stadiums, parks, churches, bus stations, commercial stores and similar POIs were considered as valid. The type of each POI along with the distance of the user from it were stored in the database. In addition, the geo-location information was used to retrieve weather data. The WeatherUnderground™ Weather API was used to retrieve current and historical data of the weather for the locations that the user had visited [41]. That information was also stored in the database and was used to enhance prediction (Emerging Patterns).

3.4. Data Collection

The initial phase of the data collection process focused on the collection of data to be used to train the classifiers. We recruited persons with mixed ages, physical conditions and toxic habits (smoking). A total number of ten persons volunteered to participate in the data collection process. Seven out of ten were men and most of them had a sedentary lifestyle. Two persons were performing light exercise once or twice a week (brisk walk) and one male used to have systematic exercise in a gym. The age range spanned from 15 to 43 years old (average age 28 years). Thirty percent of the participants were habitual smokers, but none of them was suffering from a chronic disease. The average Body Mass Index was 24.5 ± 3.4 (minimum 20.3 and maximum 30.4). Considering those data, the number of the participants may be relatively small, although the participants that aided in the data collection for the training of the algorithms can be thought as representatives of the typical Greek citizens [42].

Data collection performed in both indoors and outdoors environments. The activities for which data were collected to train the classifiers were: walking, running, jogging, standing or sitting, ascending/descending stairs, cycling and driving. Our effort was to simulate everyday conditions, so we collected data in both indoors and outdoors locations for walking, standing/sitting and ascending/descending stairs. Running, jogging, cycling and driving data collected only outdoors. Regarding walking activity it subdivided in more classes representing leisure, moderate, exercise and brisk walk while leisure, low, medium, high and extreme intensity and race cycling were identified. The classification of the walking and cycling activities divided in sub-classes based on the speed of the activity which also results in different metabolic equivalents (a measure expressing the energy cost of the physical activities and therefore the rate of energy consumption) (Table 1).

During data collection, the person that was wearing the WBAN sensors was keeping the smartphone in a pocket and was accompanied by a member of the laboratory to validate that the correct activity had been selected from the annotation menu and that the activity was performed without intermissions (check of the quality of the annotation data). Each person was performing each of the activities for at least five minutes but without an upper limit. The total duration of the collected data was seven hours and the time was

distributed between the activities as follows: 64.4% on walking, 2.1% on running, 5.1% on jogging, 18.4% on standing/sitting, 3.2% on ascending/descending stairs, 4% on driving and 2.8% on cycling.

Table 1. Activities recognized based on their speed along with the average metabolic equivalents of each activity

	Activity	Speed (km/h)	Metabolic Equivalents	Description
A	Walking	≤ 3	2,5	Leisure walk
B	Walking	3 to 5	3,0 – 3,3	Moderate walk
C	Walking	5 to 7	3,3 – 5,0	Exercise walk
D	Walking	7 to 9	5,0 – 8,0	Brisk walk
E	Running	> 9	10,0 – 14,0	Running for exercise
F	Jogging	-	6,0 – 7,0	Typical jogging
G	Standing / Sitting	-	1,0 – 2,0	Standing up or sitting
H	Stairs	-	2,5 – 3,0	Ascending or descending stairs
I	Cycling	≤ 16	4,0	Leisure cycling
J	Cycling	16 to 19	6,0	Low intensity cycling
K	Cycling	19 to 22	8,0	Medium intensity cycling
L	Cycling	22 to 26	10,0	High intensity cycling
M	Cycling	26 to 30	12,0	Extreme intensity cycling
N	Cycling	> 30	16,0	Race cycling
O	Driving	-	1,5 – 2,0	Car driving

4. Results

4.1. Application of Signal Processing and Pattern Recognition Algorithms

Signal processing of the annotated data comprises of a series of calculations to remove artifacts from the signal and then prepare the data for algorithm training. Artifacts appeared at the beginning and at the end of each activity or when switching activities, as there was a time lag between activity initiation and annotation via the smartphone application. For that reason, 10 seconds on the beginning and at the end of each activity were omitted. Additionally, when there was no GPS signal available or when the accuracy was not acceptable (greater than 6 meters) then the activity was classified as e.g. walking but without speed information. Sampling rate of 20 Hz was selected (as previously mentioned) so the two sensors were transmitting 40 data packets per second (together). Signal analysis steps included the removal of the gravitational acceleration component from the readings and then the calculation of the anteroposterior, mediolateral and vertical acceleration values

(x, y and z-axes). The data were divided by an 1-second sliding window with 50% overlap between the adjacent windows, resulting in $2n-1$ samples. The average values, the energy (calculated by the sum of the squared Fast Fourier Transform component magnitudes) for each axis along with the heart rate and the speed were calculated for each data window.

The processed data were used as input in the open-source scientific software WEKA [39] to train a series of classifiers and validate the results using the 10-fold cross-validation. Accuracy, recall and F-score for each of the algorithms examined. The data mining algorithms with application in pattern recognition were the Naïve Bayes classifier [43] that assumes conditional independence among all attributes given the class variable and learns from training data the conditional probability of each attribute, given its label class (this classifier has low computational cost and relatively good performance) and the Bayesian Networks classifier [44] which is based on the conditional rather than joint likelihood, attempting to optimize the likelihood of the entire data rather than the conditional likelihood of the class given the attributes. Then, the Support Vector Machines (SVM) which are applied on binary and multiclass classification [45] were tested. The classification process of the SVM focuses on finding a hyperplane which separates the x-dimensional data into its two classes, minimizing the margin error. Finally, Decision Trees classifiers which represent rules that are used to separate the data in classes which resemble a tree structure were evaluated. The C4.5 algorithm [46], which deals very well with missing values (here missing speed) was initially examined. Then we tested the performance of the Random Forests algorithm which is made up of tens or hundreds of decision trees, which tend not to suffer the sensitivity to noise in a dataset that single decision tree induction does [47]. A relatively new algorithm, the Functional Trees [48], was also examined. The execution of this algorithm initiates with an univariate decision tree which is constructed and then the pruning is performed estimating the error of the sub-trees below each node, which is computed as a weighted sum of the estimated error for each leaf of the sub-tree along with the estimated error of the non-leaf node, if it was pruned to a leaf. If the estimated error of the non-leaf-node is lower than the weighted sum of the estimated error for each leaf, the entire sub-tree is replaced to a leaf.

Table 2. Overall accuracy for each of the classifiers using 10-fold cross validation

Naïve Bayes	80,62%
SVM	86,31%
C4.5	93,58%
Functional Trees	97,25%
Random Forest	92,58%
Bayesian Networks	89,75%

Within the categories A to O summarized in Table 1 two

additional categories A' and K' were added to describe data that do not have speed information. A' represents the walking activity and K' the cycling activity respectively. Table 2 summarizes the results for each of the classifiers examined. The classifiers that produce the best recognition accuracy are C4.5, Functional Trees and Random Forests.

For the real-time activity recognition running on the smartphone we selected the trained model produced by the Functional Trees classifier because, due to the better performance. The confusion matrix of the Functional Trees classification is represented in Table 3 and details on the accuracy, the recall and the F-score can be found in Table 4.

Table 3. Confusion matrix of the Functional Trees classification

	A'	A	B	C	D	E	F	G	H	I	J	K'	K	L	M	N	O
A'	2897	0	0	1	0	0	6	18	162	0	3	0	0	0	0	0	0
A	0	10591	0	0	0	0	8	1	36	1	0	12	0	0	0	0	0
B	0	0	8885	1	1	0	11	0	8	4	0	10	0	0	0	0	0
C	1	0	0	1077	0	0	2	0	2	0	0	12	0	0	0	0	0
D	0	0	0	0	124	0	0	0	1	0	0	5	0	0	0	0	0
E	1	4	0	0	0	696	109	0	2	0	0	0	0	0	0	0	0
F	1	15	28	4	1	80	1808	0	14	0	0	0	0	0	0	0	0
G	18	3	0	0	0	0	0	7194	4	0	3	1	0	0	0	0	0
H	185	108	10	0	1	0	20	9	924	0	0	0	0	0	0	0	0
I	0	0	2	0	1	0	0	0	0	1488	0	5	0	0	0	0	0
J	0	0	0	0	0	0	0	9	3	0	14	0	0	0	0	0	0
K'	0	23	12	16	11	0	0	1	3	9	0	817	2	0	0	0	0
K	0	0	0	0	0	0	0	0	0	0	0	2	50	1	0	0	0
L	0	0	0	0	0	0	0	0	0	0	0	0	1	48	1	0	0
M	0	0	0	0	0	0	0	0	0	0	0	2	1	2	49	0	0
N	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	11	1
O	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	2

Table 4. The accuracy, the recall and the F-score for the Functional Trees classifier

Class	Accuracy	Recall	F-score
A'	0,934	0,938	0,936
A	0,986	0,995	0,990
B	0,994	0,996	0,995
C	0,980	0,984	0,982
D	0,892	0,954	0,922
E	0,897	0,857	0,877
F	0,921	0,927	0,924
G	0,995	0,996	0,995
H	0,797	0,735	0,765
I	0,991	0,995	0,993
J	0,700	0,538	0,609
K'	0,943	0,914	0,928
K	0,926	0,943	0,935
L	0,941	0,960	0,950
M	0,980	0,907	0,942
N	0,846	0,917	0,880
O	0,667	0,500	0,571

4.2. Application of Emerging Patterns to Predict Complications of Chronic Diseases

The data provided by the smartphone/WBAN combination are stored in the database so that further patterns to be discovered. The database is located on the smartphone storage memory and those data are uploaded to a cloud-hosted application once per day. The activities performed by the user throughout the day and data related to the environmental conditions around the user are searched by the Emerging Patterns algorithm [20]. Emerging Patterns algorithm offers better recognition performance for activities that happen in parallel and have interruptions while executed, in contrast to Skip Chain Conditional Random Fields [49], which extend the capabilities of the Conditional Random Fields, to predict the current activities based on previous observations. The discovery of Emerging Patterns requires a full database search (always updated as new user data are uploaded to it) to find any matching patterns. Due to the high computational cost of the processes it is preferred to run on a cloud-hosted application which offers scalability of the available resources. The cloud-hosted application offers a framework that can be applied in a diverse range of chronic diseases such as diabetes and COPD for the detection of short-term complications.

The data provided by the activities of the person, the location and the environmental conditions can be combined with disease related information to “learn” the patterns that affect the appearance of the complications. More specifically, regarding diabetes, the physical activity recognition can be used to calculate energy consumption during physical activities, which can be interlaced with inhaled/exhaled gases analysis to build a personalized carbohydrate metabolism model for each individual. Physical activity and energy expenditure can be used as inputs of a Personalized Decision Support System to perform real time carbohydrate metabolism estimation [50] for better insulin needs prediction. Taking into account each individual’s carbohydrate metabolism model the Personalized Decision Support System can be used to aid the optimization of the calculation of the insulin needs [22].

Regarding the prediction of the complications for COPD patients, this framework offers a tool to the scientists to relate diverse information provided by the patient, the activities and the environment (places, weather condition etc) that increase the short-term disease symptoms, such as noisy breathing, cough or shortness of breath, depending on the stage of the disease. The framework uses SOAP web services for the communication between the smartphone and the cloud-hosted application, so it can be used to extend a previous work of our laboratory targeting COPD patients [25, 26]. The telemedical framework, consists of an Electronic Health Record, a specially designed touch screen computer loaded with clinically validated questionnaires to be answered daily and a wireless sensor used by the patients to capture tracheal sounds. The framework described here can be used to monitor ambient conditions while the patient is

out of home and discover patterns between the environmental conditions, the places the patient visits and the activities performed to the short-term complications of the disease and alarm the patient in advance. So, it can be used to uninterruptedly monitor patient indoors and outdoors. We have also tested an extension of the described framework to include micro-climate provided by Internet of Things (IoT) objects. Arduino-based devices equipped with temperature, humidity, CO and dust particles sensors have been developed and placed outside selected buildings of the university campus to monitor and update the database of the cloud application with more detailed environmental data around the location. Data provided by those IoT objects are used if the distance between the smartphone of the user and the sensor is within 100 meters. Till now, we have tested the application of the Emerging Patterns’ discovery in the improvement of the prediction of the physical activities performed. We evaluated the performance with the participation of 6 volunteers who were asked to use the platform while performing daily activities as usual. After two weeks of usage, each of the volunteers asked to perform an one hour evaluation test. The volunteers were demanded to perform various activities with or without intermissions to evaluate the effectiveness of the Emerging Patterns search. The prediction results were collected and analyzed resulting to recognition accuracy of almost 100%. Errors were detected between the transitions of activities with different intensities. Those results prove that Emerging Patterns used in pattern recognition may provide advanced prediction rates in relation to the short-term complications of chronic diseases.

5. Discussion and Future Work

The study presented in this paper focused on the development of a low cost wireless body area network to capture body motion and recognize physical activities performed. The framework has been developed using open-source tools (smartphones, microcontrollers and data mining software) and a cloud-hosted database and application execution to allow the application of the system in various chronic diseases (diabetes, Parkinson’s, dementia etc). High activity recognition rates have been succeeded using the Functional Trees algorithm and the extracted model has been customized to run on a smartphone operating system to allow the performance of real-time activity recognition. The activity recognition data are saved in a cloud-hosted database and then location-aware data are collected to identify POIs and weather conditions around the user. A data mining algorithm based on Emerging Patterns algorithm searches through the database to predict short-term complications related to chronic diseases. A small scale study on the efficiency of the Emerging Patterns algorithm has been performed to evaluate the results. Our framework has been extended to allow interconnection with IoT sensors providing micro-climate data, enhancing the accuracy of the

prediction for environmental conditions-related diseases.

Conflict of Interests

The authors do not have a direct financial relation with the commercial identities mentioned in this paper.

REFERENCES

- [1] Organization WH. Global Health Observatory Data Repository. 2008 [cited 2011 8/12]; Available from: <http://apps.who.int/ghodata/?vid=2463>.
- [2] Kriska AM, Saremi A, Hanson RL, Bennett PH, Kobes S, Williams DE, et al. Physical Activity, Obesity, and the Incidence of Type 2 Diabetes in a High-Risk Population. *American Journal of Epidemiology*. 2003 October 1, 2003;158(7):669-75.
- [3] Booth F, Laye M, Lees S, Rector R, Thyfault J. Reduced physical activity and risk of chronic disease: the biology behind the consequences. *European Journal of Applied Physiology*. 2008;102(4):381-90.
- [4] Katzmarzyk PTC, Timothy S. Craig, Cora L. Bouchard, Claude. Sitting Time and Mortality from All Causes, Cardiovascular Disease, and Cancer. *Medicine & Science in Sports & Exercise*. 2009;41(5):998-1005.
- [5] McTiernan A. Mechanisms linking physical activity with cancer. *Nat Rev Cancer*. [10.1038/nrc2325]. 2008;8(3):205-11.
- [6] Organization WH. Physical Inactivity: A Global Public Health Problem. 2008 [cited 2011 08/12]; Available from: http://www.who.int/dietphysicalactivity/factsheet_inactivity/en/index.html.
- [7] Organization WH. Global recommendations on physical activity for health: World Health Organization; 2010.
- [8] Burns W, Nugent, CD, McCullagh, PJ, Finlay, D, Cleland, I, Altemeyer, B, Parente, Guido, Scotney, BW, McClean, SI, BENTON, S and McCann, J. A Smart Garment for Active Ageing: Mapping User Requirements to Technology. In: Gelderblom GJ, Soede, Mathijs, Adriaens, Leon and Miesenberger, Klaus, editor. *Everyday Technology for Independence and Care*. Amsterdam: IOS Press; 2011. p. 33-40.
- [9] Jasiewicz JM, Allum JHJ, Middleton JW, Barriskill A, Condie P, Purcell B, et al. Gait event detection using linear accelerometers or angular velocity transducers in able-bodied and spinal-cord injured individuals. *Gait & Posture*. 2006;24(4):502-9.
- [10] Doukas CN, Maglogiannis I. Emergency Fall Incidents Detection in Assisted Living Environments Utilizing Motion, Sound, and Visual Perceptual Components. *Information Technology in Biomedicine, IEEE Transactions on*. 2011;15(2):277-89.
- [11] Rigas G, Bougia P, Baga D, Tsiouras M, Tzallas A, Tripoliti E, et al., editors. A decision support tool for optimal Levodopa administration in Parkinson's disease. *Information Technology and Applications in Biomedicine (ITAB), 2010 10th IEEE International Conference on*; 2010.
- [12] Nakashima M, Ohgi Y, Akiyama E, Kazami N. Development of a swimming motion display system for athlete swimmers' training using a wristwatch-style acceleration and gyroscopic sensor device. *Procedia Engineering*. 2010;2(2):3035-40.
- [13] Mariani B, Hoskovec C, Rochat S, Büla C, Penders J, Aminian K. 3D gait assessment in young and elderly subjects using foot-worn inertial sensors. *Journal of Biomechanics*. 2010;43(15):2999-3006.
- [14] Van de Molengraft J NS, Mariani B, Aminian K, Büla C, Penders J, editor. *Wireless 6D inertial measurement platform for ambulatory gait monitoring*. 6th international workshop on Wearable, Micro and Nanosystems for Personalised Health; 2009.
- [15] Polastre J, Szewczyk R, Culler D, editors. *Telos: enabling ultra-low power wireless research*. *Information Processing in Sensor Networks, 2005 IPSN 2005 Fourth International Symposium on*; 2005 15 April 2005.
- [16] Milenković A, Otto C, Jovanov E. Wireless sensor networks for personal health monitoring: Issues and an implementation. *Computer Communications (Special issue: Wireless Sensor Networks: Performance, Reliability, Security, and Beyond)*. 2006;29:2521-33.
- [17] Lara ÓD, Pérez AJ, Labrador MA, Posada JD. Centinela: A human activity recognition system based on acceleration and vital sign data. *Pervasive and Mobile Computing*. (0).
- [18] Saito M, Nakajima K, Takano C, Ohta Y, Sugimoto C, Ezoe R, et al. An in-shoe device to measure plantar pressure during daily human activity. *Medical Engineering & Physics*. 2011;33(5):638-45.
- [19] Bieber G, Voskamp J, Urban B. Activity Recognition for Everyday Life on Mobile Phones Universal Access in Human-Computer Interaction. In: Stephanidis C, editor.: *Springer Berlin / Heidelberg*; 2009. p. 289-96.
- [20] Dong G, Zhang X, Wong L, Li J. CAEP: Classification by Aggregating Emerging Patterns Discovery Science. In: Arikawa S, Furukawa K, editors.: *Springer Berlin / Heidelberg*; 1999. p. 30-42.
- [21] Kouris I, Mougiakakou S, Scarnato L, Iliopoulou D, Diem P, Vazeou A, et al. Mobile phone technologies and advanced data analysis towards the enhancement of diabetes self-management. *Int J Electron Healthc*. 2010;5(4):386-402.
- [22] Mougiakakou SG, Bartsocas CS, Bozas E, Chaniotakis N, Iliopoulou D, Kouris I, et al. SMARTDIAB: a communication and information technology approach for the intelligent monitoring, management and follow-up of type 1 diabetes patients. *IEEE Trans Inf Technol Biomed*. 2010 May;14(3):622-33.
- [23] Nachman L, Baxi A, Bhattacharya S, Darera V, Deshpande P, Kodalapura N, et al. Jog Falls: A Pervasive Healthcare Platform for Diabetes Management. In: Floréen P, Krüger A, Spasojevic M, editors.: *Springer Berlin / Heidelberg*; 2010. p. 94-111.
- [24] Foix LFC, Morillo DS, Crespo M, Gross N, Kunze C, Giokas K, et al., editors. *AMICA telemedicine platform: a design for management of elderly people with COPD*. *Information Technology and Applications in Biomedicine, 2009 ITAB*

2009 9th International Conference on; 2009 4-7 Nov. 2009.

- [25] Crespo LF, Morillo DS, Crespo M, Leon A, Astorga S, Giorkas K, et al., editors. Telemonitoring in AMICA: A design based on and for COPD. Information Technology and Applications in Biomedicine (ITAB), 2010 10th IEEE International Conference on; 2010 3-5 Nov. 2010.
- [26] Giorkas K, Kouris I, Koutsouris D. Autonomy, Motivation and Individual Self-management for COPD Patients, the Amica Project, Wireless Mobile Communication and Healthcare. In: Lin JC, Nikita KS, editors.: Springer Berlin Heidelberg; 2011. p. 49-53.
- [27] Baga D, Fotiadis DI, Konitsiotis S, Akay M, editors. PERFORM: A system for the monitoring and management of persons with chronic motor neurodegenerative disorders: A study on Parkinson's Disease and Amyotrophic Lateral Sclerosis. Pervasive Computing Technologies for Healthcare, 2009 PervasiveHealth 2009 3rd International Conference on; 2009 1-3 April 2009.
- [28] Tsiouras MG, Tzallas AT, Rigas G, Bougia P, Fotiadis DI, Konitsiotis S. An Automated Method for Levodopa-Induced Dyskinesia Detection and Severity Classification XII Mediterranean Conference on Medical and Biological Engineering and Computing 2010. In: Bamidis PD, Pallikarakis N, editors.: Springer Berlin Heidelberg; 2010. p. 592-5.
- [29] Shahriyar A, Priya N. TwitterJacket: An Automated Activity and Health Monitoring Solution for the Elderly. CyLab. 2009;23.
- [30] Tapia EM, Intille SS, Haskell W, Larson K, Wright J, King A, et al. Real-Time Recognition of Physical Activities and Their Intensities Using Wireless Accelerometers and a Heart Rate Monitor. 11th IEEE International Symposium on Wearable Computers; 11-13 October 2007. p. 37-40.
- [31] Sun L, Zhang D, Li B, Guo B, Li S. Activity Recognition on an Accelerometer Embedded Mobile Phone with Varying Positions and Orientations Ubiquitous Intelligence and Computing. In: Yu Z, Liscano R, Chen G, Zhang D, Zhou X, editors.: Springer Berlin / Heidelberg; 2010. p. 548-62.
- [32] Krishnan NC, Panchanathan S, editors. Analysis of low resolution accelerometer data for continuous human activity recognition. Acoustics, Speech and Signal Processing, 2008 ICASSP 2008 IEEE International Conference on; 2008 March 31 2008-April 4 2008.
- [33] Duncan GE, Lester J, Migotsky S, Goh J, Higgins L, Borriello G. Accuracy of a novel multi-sensor board for measuring physical activity and energy expenditure. Eur J Appl Physiol. 2011 Sep;111(9):2025-32.
- [34] Murphy SL. Review of physical activity measurement using accelerometers in older adults: Considerations for research design and conduct. Preventive Medicine. 2009;48(2):108-14.
- [35] Lee Y-D, Chung W-Y. Wireless sensor network based wearable smart shirt for ubiquitous health and activity monitoring. Sensors and Actuators B: Chemical. 2009;140(2):390-5.
- [36] Hong Y-J, Kim I-J, Ahn SC, Kim H-G. Mobile health monitoring system based on activity recognition using accelerometer. Simulation Modelling Practice and Theory. 2010;18(4):446-55.
- [37] Gislason D. Zigbee Wireless Networking. 1 ed: Newnes; 2008.
- [38] The Essentials of the Bluetooth SIG. [16-September-2011]; Available from: <https://www.bluetooth.org/>.
- [39] Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH. The WEKA data mining software: an update. SIGKDD Explor Newsl. 2009;11(1):10-8.
- [40] OpenStreetMap Points Of Interest. [12/11/2011]; Available from: www.openstreetmap.org.
- [41] Weather Underground I. Weather API. 2011 [08/05/2011]; Available from: <http://www.wunderground.com/weather/api>.
- [42] Tzotzas T, Kapantais E, Tziomalos K, Ioannidis I, Mortoglou A, Bakatselos S, et al. Epidemiological survey for the prevalence of overweight and abdominal obesity in Greek adolescents. Obesity (Silver Spring). 2008 Jul;16(7):1718-22.
- [43] Lewis D. Naive (Bayes) at forty: The independence assumption in information retrieval. In: Nédellec C, Rouveirol C, editors. Machine Learning: ECML-98: Springer Berlin / Heidelberg; 1998. p. 4-15.
- [44] Heckerman D. A Tutorial on Learning with Bayesian Networks. In: Holmes D, Jain L, editors. Innovations in Bayesian Networks: Springer Berlin / Heidelberg; 2008. p. 33-82.
- [45] Vapnik V. The Nature of Statistical Learning Theory (Information Science and Statistics). 2 ed: Springer; 1999.
- [46] Quinlan JR. C4.5: programs for machine learning: Morgan Kaufmann Publishers Inc.; 1993.
- [47] Breiman L. Random Forests. Machine Learning. 2001;45(1):5-32.
- [48] Gama J. Functional Trees. Machine Learning. 2004;55(3):219-50.
- [49] Sutton C, McCallum A. Introduction to Conditional Random Fields for Relational Learning. In: Getoor L, Taskar B, editors. Introduction to Statistical Relational Learning: MIT Press; 2006.
- [50] Kouris I, Tsirmpas C, Mougiakakou SG, Iliopoulou D, Koutsouris D. E-Health towards ecumenical framework for personalized medicine via Decision Support System. Conf Proc IEEE Eng Med Biol Soc. 2010;2010:2881-5.