# A Support Vector Machine Approach for Categorization of Patients Suffering from Chronic Diseases

Christos Bellos[1], Athanasios Papadopoulos[1], Roberto Rosso[2], and Dimitrios I. Fotiadis[1]

[1] Foundation for Research and Technology - Hellas, Biomedical Research, Ioannina Greece
[2] TESAN Telematic & Biomedical Services S.p.A., Vicenza Italy
{cbellos,thpapado,Fotiadis}@cs.uoi.gr
rosso@tesan.it

**Abstract.** The CHRONIOUS system is an open-architecture integrated platform aiming at the management of chronic disease patients. The system consists of a body sensor network collecting patient's vital signals, a Personal Digital Assistance (PDA) for the real-time data analysis based on a Decision Support System (DSS) and a central system for the deeper analysis of patient's status and data storing. The DSS combines several data sources to decide upon the severity of patient's current health status. The first pilot study has been designed and carried out using patients suffering from Chronic Obstructive Pulmonary Disease (COPD). The DSS facilitates a one-against-all multi-class Support Vector Machine (SVM) classification system. The performance of the categorization scheme provides high classification results for most of the patient's health status levels. The involvement of a larger number of patients might increase further the performance of the system.

**Keywords:** Personalized treatment, Wearable Monitoring, Management of Chronic Diseases, Real-time classification.

## 1 Introduction

The CHRONIOUS [1] system provides a general patient's management scheme which can be configured and controlled remotely by medical experts or clinicians. The CHRONIOUS system has been designed in order to be flexible so that through the integration of sensors and services covers both patients and caregiver's needs. Its functionalities include the collection of data by sensors which are integrated into a jacket, the control of vital signals, dietary habits and plans, as well as the management and analysis of drug intake, environmental parameters and activity information. Abnormal health conditions are identified using intelligent algorithms running on a portable Personal Digital Assistance (PDA) device and reported to the responsible healthcare professionals supporting decision making and analysis of data.

The flow of data in the proposed system, which is integrated into a mobile application embedded in a PDA device, includes three phases; data preprocessing, analysis and decision making. Data preprocessing is focused on the improvement of

K.S. Nikita et al. (Eds.): MobiHealth 2011, LNICST 83, pp. 264–267, 2012.

signal quality removing several types of noise and artifacts. In the data analysis various types of signals are captured in order to make in real-time a preliminary assessment of the patient's health status. The Data Fusion component fuses all acquired data and forms a multi-dimensional vector to feed the Decision Support System (DSS), triggering the decision making phase.

## 2 Materials and Methods

The aim of the developed DSS component is twofold; initially constructs the training model and trains the algorithms and afterwards classifies the health episodes according to four different levels of severity on an application embedded in the PDA. A supervised Support Vector Machine (SVM) classifier has been implemented. Apart from the increased performance, SVMs have the ability to handle high dimensional data efficiently, using only a subset of training records, called support vectors, in order to represent the decision boundary.

The specific categorization difficulty that the SVM faces is characterized as multi-class due to the four levels of severity that clinical experts have annotated the health episodes and respectively the available data. The SVMs were originally designed for binary classification [2]. Several methods [3] have been proposed to effectively extend it to multiclass classification where a classifier may be constructed by combining several binary classifiers while others propose to consider all classes at once. In the CHRONIOUS system, the one-against-all SVM classification approach has been implemented.

In the performed analysis three different kernels [4] are used; radial basis function (RBF), Polynomial and Sigmoid and their respective parameters (C, degree and gamma) have been optimized using the differential evolution [5] (DE) method. In order to identify the most important clinical-pathological information for individualized health status identification, two feature selection [6] algorithms, Correlation-based Feature Subset Selection and the Gain Ratio attribute evaluation algorithm were applied, to rank the importance of the features and clinical parameters in the studied disease. The Correlation-based Feature Subset Selection algorithm has been selected from the correlation-based methodology and the GainRatioAttributeEval algorithm has been selected from the ranking methodology. The first algorithm evaluates the worth of a subset of attributes by considering the individual predictive ability of each feature along with the degree of redundancy between them. The latter algorithm evaluates the worth of an attribute by measuring the gain ratio with respect to the class.

## 3 Results

In the performed analysis we have recorded data from COPD patients using the wearable jacket [8]. The Feature Extraction process has extracted 18 features from 16 patients in a pilot hospital. The applied correlation-based algorithm reduces the number of features to eight, while the ranking algorithm reduces the number of

**Table 1.** Selected features after the reduction of the initial feature's pool

| Feature | Correlation-based algorithm | Ranking algorithm |
|---|---|---|
| Mean QR Distance | x | |
| Mean Heart Rate | x | x |
| Respiration Rate | x | x |
| Inhalation Duration | x | |
| Exhalation Duration | x | x |
| Oxygen saturation | x | x |
| Body Temperature | x | |
| Environmental Temperature | x | x |

features to five (Table 1). The features are acquired through a standardized protocol which contains: 6 minutes walking, 45 minutes supine position and 45 minutes standing position [7].

In Table 2 the results of the classification are presented after applying different kernels of the implemented SVM using a dataset which has been constructed using real patients' data.

**Table 2.** Correctly classified instances and comparison between different kernels

| Class - Level of classified severity | Applied kernel | Without Feature selection (%) | Correlation-based algorithm (%) | Ranking algorithm (%) |
|---|---|---|---|---|
| Level 1 | RBF | 97 | 97 | **99** |
| | Polynomial | 86 | 83 | 84 |
| | Sigmoid | 87 | 88 | 87 |
| Level 2 | RBF | 86 | 87 | **91** |
| | Polynomial | 78 | 78 | 67 |
| | Sigmoid | 65 | 63 | 56 |
| Level 3 | RBF | 90 | 91 | **94** |
| | Polynomial | 86 | 86 | 89 |
| | Sigmoid | 83 | 77 | 81 |
| Level 4 | RBF | 89 | 90 | **93** |
| | Polynomial | 79 | 74 | 73 |
| | Sigmoid | 88 | 86 | 88 |

The dataset is randomly split into training and testing dataset using the 10-fold cross-validation method. The levels in the first column of Table 2 represent the different annotated levels of severity as well as the different implemented SVM (since we followed the one-against-all approach, we developed four different binary SVMs). The second column displays the three different applied kernels for each SVM while the last three columns display the percentage of the correctly classified instances for different datasets (entire dataset, optimized dataset after applying the Correlation-based algorithm, optimized dataset after applying the GainRatioAttributeEval algorithm). In this preliminary analysis we notice that the RBF kernel provides more accurate results for the tree different datasets as well as for all four levels which represent the four different developed SVMs.

## 4 Conclusions

The CHRONIOUS COPD system provides an estimation of the severity of the patient's condition or detects possible critical health episodes. The aim of the developed SVM is to limit the decision error and to increase the system accuracy. In our preliminary analysis we have obtained positive results of the performance and the accuracy of the implemented system.

Comparing the results acquired from the three different datasets that have been formed and facilitated, we conclude that the performance of the RBF kernel followed by the application of Ranking feature selection algorithm provides the highest categorization result. As for the next steps, the employment of larger datasets as well as the implementation of additional classification methodologies will improve the performance of the CHRONIOUS system.

**Acknowledgements.** This work has been partly funded by the European Commission through IST Project FP7-ICT-2007–1– 216461, www.chronious.eu.

## References

1. Bellos, C., Papadopoulos, A., Fotiadis, D.I., Rosso, R.: An Intelligent System for Classification of Patients Suffering from Chronic Diseases. In: 32nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Buenos Aires, Argentina (2010)
2. Hsu, C., Lin, C.: A Comparison of Methods for Multiclass Support Vector Machines. IEEE Transactions on Neural Networks 13(2) (2002)
3. Kumar, M., Gopal, M.: Reduced one-against-all method for multiclass SVM classification. Expert Systems with Applications (2011), doi:10.1016/j.eswa.2011.04.237
4. Ayat, N.E., Cheriet, M., Suen, C.Y.: Automatic model selection for the optimization of SVM kernels. Pattern Recognition 38, 1733–1745 (2005)
5. Efrén, M.M., Jesús, V.R.: A Comparative study of Differential Evolution Variants for Global Optimization. In: GECCO 2006, pp. 485–492 (2006)
6. Taniar, D.: Data Mining and Knowledge Discovery Technologies. part of the IGI Global series named Advances in Data Warehousing and Mining (ADWM) vol. 2 (2007)
7. Bellos, C., Papadopoulos, A., Rosso, R., Fotiadis, D.I.: Extraction and Analysis of features acquired by wearable sensors network. In: 10th International Conference on Information Technology and Applications in Biomedicine, Corfu, Greece (2010)
8. Papadopoulos, A., Fotiadis, D.I., Lawo, M., Ciancitto, F., Podolak, C., Dellaca, R.L., Munaro, G., Rosso, R.: CHRONIOUS: A Wearable System for the Management of Chronic Disease. In: 9th International Conference on Information Technology & Applications in Biomedicine, Larnaca, Cyprus (2009)