Predictive models to improve the wellbeing of heart-failure patients

Aljoša Vodopija, Miha Mlakar and Mitja Luštrek

Jožef Stefan Institute, Department of Intelligent Systems, Jamova 39, 1000 Ljubljana, Slovenia

Abstract. The paper presents an approach to providing advice on health related quality of life to patients with congestive heart failure, using predictive models built from telemonitoring data. First, by combining machine learning algorithms, feature construction, feature selection and expert knowledge, we built a set of predictive models. We then identified which of the features present in the models can be changed by the patients themselves with an appropriate intervention and modelled the association between them and all the other features using linear models. At the end, by using multi-objective optimization, we found the minimum necessary changes of the modifiable features that improve the patients' feeling of health. This way we can provide a set of appropriate advices for patients. The findings mostly correspond to the current medical knowledge, although some may represent new insights.

Keywords: predictive models, optimization models, congestive heart failure

1 Introduction

Congestive heart failure (CHF) is a chronic disease in which the heart cannot adequately supply the organs and tissues with oxygen and nutrients. It can have various causes, including damaged heart tissue (e.g., due to a heart attack), atherosclerosis, hypertension etc. 1-2% of people in the developed world suffer from CHF [7], and it is the most frequent cause of hospitalization in people aged over 65. CHF cannot be cured, so the management of the disease aims to prolong the lifespan and to increase the quality of life. The latter is often expressed through the so-called patient-reported outcomes (PROs), which are becoming increasingly accepted as one of the measures for the evaluation of medical treatments [1].

In this paper we work on data collected during the Chiron telemonitoring study performed among CHF patients in Italy and UK [10]. In these patients a large number of physiological and ambient parameters were collected by wearable and other devices, together with PROs describing how they felt. We have already built models that predict the patients' feeling of health based on the values of the telemonitored parameters, achieving accuracies around 80% [3]. We used these models to study the relations contained in them, finding some previously known as well as some new ones.

In this paper, we have taken a step further and used similar predictive models to generate advice for the patients on how to improve their feeling of health. To do so, we first selected the features that can be directly modified by the patients with an appropriate intervention or behaviour adaptation. Afterwards, we modelled the relations between these and other features, so that we could propagate the changes in the directly modifiable features to all the affected features. Finally, we used multi-objective optimization to find minimal changes of the directly modifiable features that improve the patients' feeling of health.

It has long been the promise of predictive models in medicine to aid the selection and adjustment of treatments by predicting their outcomes, similarly to what we propose in this paper. However, this promise has not often been realized, most likely becase data needed for actionable models is not readily available. In CHF, most predictive models are long-term and concerned with mortality - a prime example is the MAGGIC score [9], which was developed on data of 39,372 patients to predict 3-year mortality. However, the information contained in such models is quite general and already included in the guidelines for the management of CHF. There are also some medium-term models concerned with hospitalizations - the closest example to our work uses a genetic algorithm to construct a strategy to prevent hospital readmissions [8]. Predictive models concerned with PROs are rare, probably because the interest in PROs is relatively recent. A typical example is the model by Ramos et el. [12], which was developed on data from 103 patients to predict the quality of life. We are not aware of any work that tries to improve the patient's feeling of health through the use of such models.

The rest of the paper is organized as follows. In Section 2 we present the collected dataset and the extracted features. In Section 3 and 4 we present used methods and obtained results. We conclude with Section 5.

2 Dataset

2.1 Data gathering and description

The Chiron project carried out an observational study in which 38 congestive-heart-failure patients from the United Kingdom and Italy were telemonitored [10]. However, some of the data were incomplete, so only the data of 12 patients from the UK and 13 patients from Italy were included in the analysis. These 25 patients together provided a total of 1,068 usable recording days. The patients were aged on average 63±9.6 years. The majority were male (72%) and were categorised as NYHA class 2 (64%) or 3 (36%). The data consists of 15 parameters carefully selected based on their relevance to CHF.

During the study, the patients were wearing vital-signs monitoring equipment [5] for several hours each day. The equipment consisted of an ECG device, two accelerometers placed on the chest and thigh, a body-temperature and a body-humidity (sweating) sensor. The ECG recordings were subsequently analyzed with the Falcon algorithm [6] to extract the fiducial points, enabling us to compute the heart rate as well as to describe each heart beat with additional parameters such as PR interval, QRS duration and T-wave amplitude. The patient's activities and energy expenditure were extracted from the accelerometer recordings [4]. The patients were also provided with a mobile application for generating daily reports about their measurements of systolic and diastolic blood pressure, weight, blood oxygen saturation, ambient temperature and humidity. In addition, they reported their overall feeling of health with respect to the previous day on a daily basis (feeling much worse than yesterday, worse, the same, better or much better).

2.2 Feature construction

For every parameter that was measured continuously or multiple times per day we calculated the average value and the standard deviation during the whole day and during three different types of activities (lying, sitting and moving). From some of the parameters, we further calculated various additional features such as ratios and differences. For more details see [3].

For every feature described above we then calculated a personalized version as the ratio between the daily value of the parameter and its average value for the patient over the whole study period. At last we calculated the change of the feature value in comparison with the previous day.

For the purpose of predictive modelling, we had to select the class to be predicted. If each of the five distinct feelings of health corresponds to one class, the differences between them are too small. Therefore we decided to have only two classes. In the paper [3] we analyzed several class definitions and concluded that the best definitions (also used in this study) is: "Feeling much worse or worse three out of last four days (Class bad) vs. Feeling much better or better three out of last four days (Class good)". Note that for most class definitions, the accuracy of the prediction is only slightly lower, so the selection of the class does not affect the results greatly, as long as only two classes are used and the instances where the patients are feeling the same as yesterday are omitted. However, since the majority of the data instances have the class "feeling the same as yesterday", such a class definition leaves us with only 118 useful instances.

2.3 Feature selection

We constructed several features (309) while the amount of instances is very small (118). Therefore we applied three feature selection methods and as a result obtained three subsets of features:

– The first method was the Correlation-based Feature Subset Selection (CFS) as implemented in the R statistical suite [11]. It is a statistical method which evaluates subsets of features on the basis of the hypothesis that good feature subsets contain features highly correlated with the class value.

– Another subset was chosen manually based on expert medical knowledge and previous experience from the Chiron project.
– Many of the features had a lot of missing values. In the last subset we included only the features with at least 87% values (this threshold was experimentally chosen), because features with a lot of missing values can add noise to the predictions.

These subsets of data were then used for building predictive models described in the next section.

3 Methods

The main purpose of the study in this paper was to construct advice to help CHF patients improve their feeling of health. This was done in the following three steps, which are described in detail in Subsections 3.1–3.3:

– Construction of machine-learning models that predict the patients' feeling of health.
– Selection of features from these models that can be easily changed by the patients themselves with an appropriate intervention or behaviour adaptation (named *modifiable features*), and modelling the relations between them and other features.
– Construction of advice on how to change the modifiable features, by using multi-objective optimization and models from the previous two steps.

3.1 Predictive models

In our previous work [3] we compared several machine learning algorithms. We obtained the best performance using the Random Forest algorithm with the accuracy of 79.3%. Because of that we decided to use the Random Forest algorithm for building the predictive models in this study as well. The class was defined in Subsection 2.2 and the features were selected as described in Subsection 2.3. The algorithm was implemented in the Weka [2] machine-learning suite and run with the default hyper-parameter values.

Since most of the features in the predictive models cannot be easily changed, we identified – based on our judgement – which of them can be changed by the patients themselves with an appropriate intervention or behaviour adaptation. These modifiable features are suitable subject for advice to the patients.

3.2 Correlation between features

Some modifiable features are highly correlated with other features, so if we want to consider the impact of changing one of the modifiable features, the correlated features must also be changed. Because of that we had to model these correlations. The first step was to recognize which of the modifiable features were correlated within other modifiable features, since they cannot be modified independently. We iteratively removed the most correlated modifiable feature according to the coefficient of multiple correlation, until the highest correlation fell below 0.5 as this value represents the standard threshold to distinguish between high correlation and moderate correlation. The removed features were included in a set named *correlated features*.

The second step was to recognize other features correlated with the modifiable ones. We again used the coefficient of multiple correlation (as before we set the threshold to 0.5) to identify them, and we again included the resulting features in the set of correlated features. For the remaining features, we concluded that they are not associated (at least linearly) with the modifiable features, and the set of these features was named *uncorrelated features*. In other words, these features remain the same no matter how we change the modifiable features, and are mostly features that do not change much day to day anyway.

The final step was to model the association between the modifiable features and correlated features. We used linear models as implemented in the R statistical suite [11] and run with the default hyper-parameter values.

3.3 Optimization and advice generation

We represented the task of generating advice to improve the patients' feeling of health as an optimization problem. The optimization algorithm will suggest which modifiable features should be changed and for how much. The desired solutions would be the ones that would change minimum number of features by as small an amount as possible. In this way the patient would need minimal effort to improve his felling of health.

To find such solutions, we defined two-objective optimization problem. Both objectives were to be minimized. The first objective was the number of features that were changed and the second was the overall absolute sum of all changes. Additionally, every obtained solution was determined feasible if the feeling of health improved, otherwise the solution was considered infeasible - this was checked by first calculating the correlated features from the modifiable ones as described in Subsection 3.2, and then feeding all the features into a predictive model described in Subsection 3.1. The experiment runs on average about 1 hour on a computer with specifications: CPU - i7 3.4 GHz and RAM - 8GB.

We used the well-known multi-objective optimization NSGA-II with following settings: a hundred-sized population, simulated binary crossover, polynomial mutation, tournament selection and 10,000 evaluations.

One of the objective values was how many parameters changed in comparison to the current patient's parameter values. Since it would be nearly impossible for the optimization algorithm to find solutions that would have the same real values as the initial ones, we needed to construct an initial population of solutions as the combination of the randomly created solutions and the copies of the current patients parameters. This combination enables the optimization algorithm to combine different parameter values to find combinations that result in better feel of heath, but at the same time it enables the algorithm to keep the number of changed parameters low. The initial population consisted of 60% of randomly generated solutions and 40% of copies of the initial parameter values.

4 Results

In this section we are going to present the results of the evaluation of the three steps described in the previous section.

4.1 Predictive models

To build the predictive models, we analyzed their accuracy on the three different feature subsets described in Subsection 2.3. For every subset, we performed 10-fold cross-validation, which was repeated 30 times with different splits into folds. The average accuracies for different feature subsets are shown in Table 1.

From the table we can see that the best result was obtained by using the set of all features with small amount of missing values. However, because it does not contain all the modifiable features, and because the accuracies for the other subsets are close, we decided to use the union of the three subsets. This union contained both regular and personalized features. Since these are highly correlated, and since the personalized versions proved more robust in the next step, we settled on the personalized features only. On one hand, this set of features contains all the modifiable features, and on the other hand, models built from personalized features are more likely to work for previously unseen patients. We believe these properties outweigh the 3-percentage-point lower accuracy compared to the set of features with small amount of missing values.

Table 1. Comparison of different features selection approaches (the number of features per set and the accuracy obtained using Random Forest classificator)

Features	**No. of features** $[n]$	**Accuracy** [%]
Small amount of NA	13	**86.6**
Union	97	84.7
Only personalized	**40**	**83.2**
CFS selection	11	83.1
Expert selection	92	80.2

The set of all personalized features contains 43 features, where we identified 13 modifiable features listed below (organized by the timing of the measurement):

- Once per day: ambient temperature, ambient humidity, systolic and diastolic blood pressure, weight
- Average over the whole day: heart rate, energy expenditure
- Activity-specific: duration of lying, sitting, moving; average heart rate during lying, sitting, moving

4.2 Correlation between features

Using the method we described in the Subsection 3.2, the duration of lying and the average heart rate during sitting were recognized as correlated with the other modifiable fatures, and were thus moved to the correlated feature set. From the remaining personalized features, the following six were also recognized as correlated:

- Change from the previous day: oxygen saturation, systolic blood pressure
- Average over the whole day: skin humidity, QT interval at average heart rate, ambient temperature/skin temperature, ambient humidity/skin humidity

We then used linear models to describe the associations between the modifiable and the correlated features. The results are shown in the Table 2. The "Linear" column represents the average of the absolute residuals for every correlated feature when predicted with a linear model. The "Mean" column represents the average of the absolute residuals using the patient's average feature value as the prediction. The "Prev. day" column represents the average of the absolute residuals using the previous day's feature value as the prediction. From Table 2 we can see that for every correlated feature, linear models obtained better results comparing to the two baseline approaches.

Table 2. Comparison of different methodes (the average of residulas per method and feature, for simplicity the residuals are multiplied by 100)

Feature\Prediction method	**Linear**	**Mean**	**Prev. day**
Duration of lying	27	65	63
QT interval at average HR	1	3	4
Average of skin humidity	8	14	14
Temperature ratio	4	11	6
Humidity ratio	9	22	17
Average HR during sitting	1	3	3
Change in systolic BP	3	5	7
Change in oxygen saturation	0	0	0

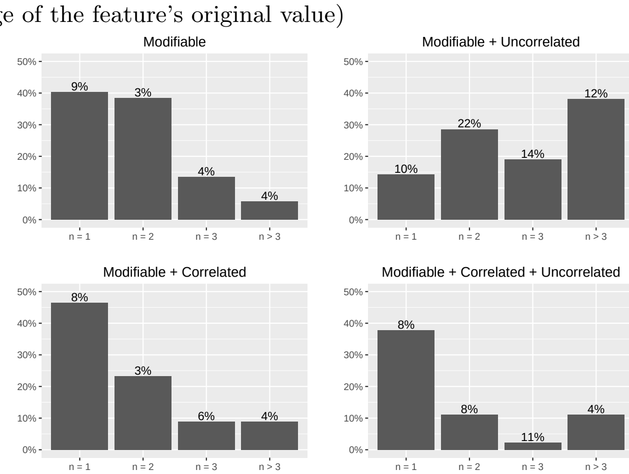
4.3 Optimization and advice generation

We defined four optimization problem based on which features were included in the predictive models.

- Problem M: Only modifiable feature subset was included
- Problem M + C: Only modifiable and correlated feature subsets were included
- Problem M + U: Only modifiable and uncorrelated feature subsets were included
- Problem M + C + U: All feature subsets were included

For every the solution to each of these four problems, we computed several statistics and measures. In some cases the optimization found more than one solutions for the given patient. For the purpose of the statistics and measures in this section we included all the solutions and regarded them as solutions for different patients. However, this has to be further investigated in our feature work. In Figure 1 we can see the number of solutions consisting of $n = 1, 2, 3$ or more changed modifiable features (pieces of advice that concern $n = 1, 2, 3$ or more features). In addition, over every bar we show the average change as the percentage of the feature's original value. For example, in the graph "Modifiable" we see that around 40% of the pieces of advice were concerned with only one feature, which should on average be changed by 9%; around 39% of the pieces of advice were concerned with two features, which should on average be changed by 3%; and so on. From the figure and the table one can see that the best results were obtained using only modifiable features and using modifiable + correlated features. In both problems more than 90% of the solutions required changing only of 1, 2 or 3 modifiable features, and the changes were relatively small (from 3% to 9%).

Fig. 1. Comparison of different optimization problems (the number of solutions consisting of $n = 1, 2, 3$ or more changed modifiable features and the average change as the percentage of the feature's original value)



In addition, for every problem and every modifiable feature we calculated the number of solutions which include that feature, and the average and standard deviation of the changes. Those measures are shown in Table 3. The table shows that many solutions required changing the heart rate, the ambient temperature, the ambient humidity, the blood pressures and the weight. The energy and activity durations seem to be less important. If we concentrate only on the first two problems (which we consider to generate the best advice), we see that the advice suggest to increase the duration of moving and to reduce the duration of sitting. This can be interpreted that the patients would feel better if they were more active, which seems reasonable. The advice also suggests to increase the difference between the systolic and diastolic blood pressure, which makes sense as it is a sign of an adequately functioning heart. Even though the blood pressure is modifiable to some degree, though, such a change may not be possible to achieve in CHF patients. The advice to increase the weight is problematic and probably arose from noise in our relatively small dataset. The advice to increase the temperature is reasonable as CHF patients often complains of cold due to poor circulation in their limbs. We believe that advice concerning environment management is safe and can be used by the patient without any warning. The advice concerning physical activity may have to be reviewed and evaluated by medical practitioners before being relayed to the patients.

Table 3. Comparison of different optimization problems (the number of solutions which include a certain feature, and the average and standard deviation of the changes)

Problem	**M**		**M + C**		**M + C + U**		**M + U**	
Feature	n%	avg ± sd	n%	avg ± sd	n%	avg ± sd	n%	avg ± sd
Energy expenditure	8	-6 ± 12	5	-11 ± 6	4	0 ± 0	43	28 ± 91
Duration of sitting	4	-1 ± 1	4	-1 ± 2	4	0 ± 0	31	-4 ± 15
Duration of moving	2	0 ± 0	4	2 ± 3	2	0 ± 0	31	47 ± 80
Avg. HR over the day	15	2 ± 2	18	5 ± 10	18	-7 ± 8	40	3 ± 9
Avg. HR during lying	17	0 ± 0	12	-1 ± 1	9	0 ± 0	45	0 ± 1
Avg. HR during moving	8	0 ± 0	11	-1 ± 3	53	-8 ± 6	38	-1 ± 4
Ambient temperature	46	6 ± 13	43	6 ± 10	18	5 ± 22	52	14 ± 17
Ambient humidity	23	7 ± 8	21	1 ± 6	2	0 ± 0	79	19 ± 25
Systolic BP	19	6 ± 9	14	1 ± 5	4	-5 ± 7	38	0 ± 2
Diastolic BP	23	1 ± 6	20	-3 ± 13	11	0 ± 0	38	2 ± 8
Weight	38	0 ± 2	18	1 ± 3	9	-3 ± 8	33	1 ± 4

5 Conclusions

This paper shows a preliminary idea on how to identify advice for CHF patients to improve their feeling of health. By using described methods we were able to provide a set of advice for patients. Most results seem reasonable and most of them correspond to the current medical knowledge, although some may represent new insights.

Since the dataset is not very large, the obtained results cannot be considered highly reliable. In the future, we will integrate the presented approach in the HeartMan system, which will provide guidance on disease management to CHF patients. Once the system is finished, it will be piloted on 80 CHF patients. This will provide real-life validation of our advice, as well as new data to improve it.

Acknowledgments

This work was funded by the HeartMan project which received funding from the European Unions Horizon 2020 research and innovation programme under grant agreement No 689660. The HeartMan project partners are Jožef Stefan Institute, Sapienza University, Ghent University, National Research Council, ATOS Spain SA, SenLab, KU Leuven, MEGA Electronics Ltd and European Heart Network.

References

[1] Anker, S. D., Agewall, S., Borggrefe, M., Calvert, M., Jaime, C. J., Cowie, M. R., Ford, I., Paty, J. A., Riley, J. P., Swedberg, K., Tavazzi, L., Wiklund, I., Kirchhof, P.: The importance of patient-reported outcomes: a call for their comprehensive integration in cardiovascular clinical trials. Eur Heart J, 35, 2001–2009, (2014)

[2] Frank, E., Hall, M. A., Witten, I. H.: The WEKA Workbench. Data Mining: Practical Machine Learning Tools and Techniques, Fourth Edition (2016)

[3] Luštrek, M., Somrak, M.: Mining telemonitoring data from congestive-heart-failure patients. Proc. IS 2014, 58–61 (2014)

[4] Luštrek, M., Cvetković, B., Kozina, S.: Energy Expenditure Estimation with Wearable Accelerometers. Proceedings of the 2012 IEEE International Symposium on Circuits and Systems, 5-8 (2012)

[5] Mazomenos, E. B., Rodríguez, J. M., Cavero, C., Tartarisco, G., Pioggia, G., Cvetković, B., Kozina S., Gjoreski, H., Luštrek, M., Solar, H., Marinčič, D., Lampe, J., Bonfiglio, S., Maharatna, K.: Case Studies. System Design for Remote Healthcare, 277–332 (2014)

[6] Mazomenos, E. B., Biswas, D., Acharyya, A., Chen, T., Maharatna, K., Rosengarten, J., Morgan, J. and Curzen, N.: A low-complexity ECG feature extraction algorithm for mobile healthcare applications. IEEE journal of biomedical and health informatics, 17, 459–469 (2013)

[7] Mosterd, A., Hoes, A. W.: Clinical epidemiology of heart failure. Heart, 93, 1137–1146 (2007)

[8] Natale, J.: A strategy for reducing congestive heart failure readmissions through the use of interventions targeted by machine learning prediction. (2015)

[9] Pocock, S. J., Ariti, C. A., McMurray, J. J., Maggioni, A., Køber, L., Squire, I. B., Swedberg, K., Dobson, J., Poppe, K. K., Whalley, G. A., Doughty, R. N.: Meta-Analysis Global Group in Chronic Heart Failure. Predicting survival in heart failure: a risk score based on 39372 patients from 30 studies. European Heart Journal, 34, 1404–1413 (2013)

[10] Puddu, P. E., Morgan, J. M., Torromeo, C., Curzen, N., Schiariti, M., Bonfiglio, S.: A clinical observational study in the Chiron project: Rationale and expected results. Impact Analysis of Solutions for Chronic Disease Prevention and Management, 74–82 (2012)

[11] R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, https://www.R-project.org/ (2016)

[12] Ramos, S., Prata, J., Gonçalves, F. R., Coelho, R.: Congestive heart failure and quality of life. Applied Research in Quality of Life, 9, 803–817 (2014)