Predictive modelling of feeling of health for congestive heart failure patients

Jakob Valič
Jožef Stefan Institute
Jamova cesta 39
SI-1000 Ljubljana
jakob.valic@ijs.si

Miha Mlakar
Jožef Stefan Institute
Jamova cesta 39
SI-1000 Ljubljana
miha.mlakar@ijs.si

Borut Budna
Jožef Stefan Institute
Jamova cesta 39
SI-1000 Ljubljana
budna.borut@gmail.com

Mitja Luštrek
Jožef Stefan Institute
Jamova cesta 39
SI-1000 Ljubljana
mitja.lustrek@ijs.si

ABSTRACT

This paper focuses on predictive modelling of feeling of health (FOH) for congestive heart failure (CHF) patients. The basis for models is the data obtained from HeartMan project clinical trials, which involved 66 patients. The main indicator was FOH, reported repeatedly by patients. We used the Decision trees and Extreme gradient boosting algorithms to build the predictive models. In case of missing data, we used k-Nearest Neighbours imputation method. The algorithms predicted the FOH with around 70% accuracy. The relations in decision tree are in line with medical knowledge about CHF.

Keywords

HeartMan project, Feeling of health, Predictive modelling

1. INTRODUCTION

CHF is a serious disease, affecting 1-2% of population in western world. The percentage rises to more than 10% among people older than 70 years. [7] There is no cure available. However, if the disease happens, proper disease management is crucial as it can improve patient's quality of life.

The main actions to improve the health of CHF patients are quite well established: they include medications, physical exercise, mental exercises, proper nutrition, pill intake monitoring, etc. Modern technology enables monitoring more parameters related to health and quality of life than was possible ever before. Furthermore, quality of life does not depend just on physical health, but is subjective and can only be reported by the patients themselves. Patient-reported outcomes are thus becoming increasingly important in medicine. This paper deals with predicting self-reported FOH from parameters collected with monitoring technology.

The data used in this paper was collected during the HeartMan project. [1] Project developed the application for self-management of the CHF. The data collected was also used to build predictive models of FOH. The idea of predictive modelling is to enable the choice of actions for improving patient's FOH. The HeartMan models provide short-term advice and interventions in comparison to more common long-term ones. [3] Furthermore, the advice could be personalized. The HeartMan project built upon a somewhat similar Chiron project [2], which aimed at developing a decision-support system to estimate the health risk of the patients. The Chiron study included 24 patients, who used wearable devices, and provided information on how they felt each day. [5]

2. HEARTMAN CLINICAL STUDY

HeartMan clinical study was carried out in two trials: the first took place in Belgium, another in Italy. In Belgium, 36 patients were involved, 12 of which were in control group and 24 in intervention group. In Italy, 30 patients were involved. 80% of the patients were male. The mean age of patients was 63 years. Most of the patients had a New York Heart Association (NYHA) functional class II, other had a NYHA functional class III, meaning they had (slightly) limited physical abilities. Most of the patients had been diagnosed with CHF more than 18 months. The intervention group patients used the HeartMan system from 3 to 6 months. They were given the RuuviTag environmental sensor to gather information on air humidity, temperature and pressure. They wore HeartMan watch to gather the PPG signal, skin temperature, galvanic skin response, cortisol level readings, heart rate and RR intervals. Each day they measured their weight and blood pressure. They were given smartphone with HeartMan application.

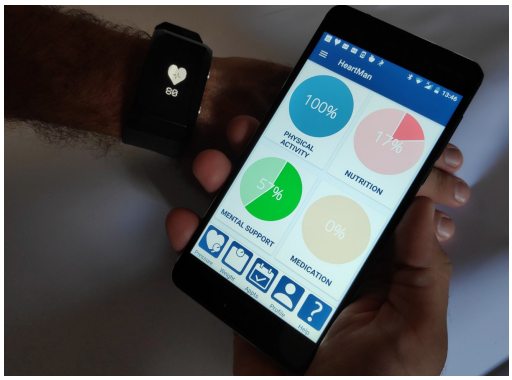

Figure 1: HeartMan watch and application.

HeartMan application gathered data from the HeartMan watch and sent it to the server. A crucial part of HeartMan application was the decision support system, which scheduled mental and physical exercises several times per week for each patient. On scheduled days, a notification was sent to the patient urging him to complete the exercise. Other than that, he could decide to do additional exercises. Each time a patient started one of the mental exercises, he was asked to answer a question: "How do you feel today?" He could then choose among the following answers:

- 1 = much better than usual
- 2 = somewhat better than usual
- 3 = about the same as usual
- 4 = somewhat worse than usual
- 5 = much worse than usual

The patient's answers regarding their FOH served as the input data for predictive models regarding the CHF patient's FOH. The distribution in Figure 2 shows us the number of FOH entries per patient from both Belgian and Italian trial. Some patients provided the information on their FOH every day, whereas some barely provided any.

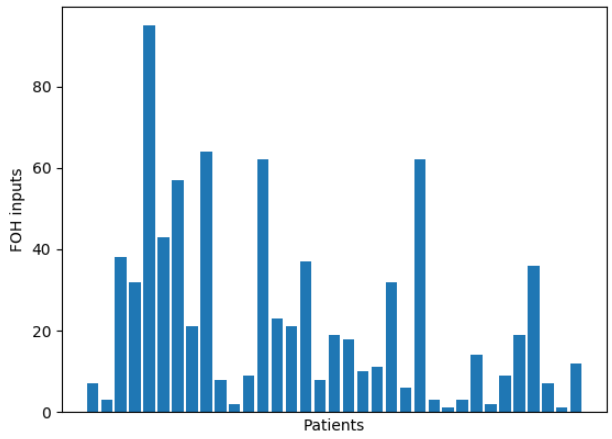

Figure 2: The distribution of FOH inputs per patient

3. PREDICTIVE MODELLING

3.1 Features

We used the following features collected by the HeartMan system:

- Air humidity, air temperature and air pressure, which were collected by RuuviTag sensor.
- Skin temperature, galvanic skin response and heart rate, which were collected by HeartMan wristband.
- Parameters from exercise report. The duration of the exercise, the type of exercise (endurance or resistance) and validity of exercise (check if pre-exercise requirements were met and the exercise was performed correctly) were generated by HeartMan application after exercise was performed by patient. The estimation of exercise intensity was input by patients.
- Systolic blood pressure and diastolic blood pressure, input by patients using HeartMan application.
- FOH, input by patients using HeartMan application.

Since the patients reported their FOH at most once per day, one instance for modelling corresponded to one day. We computed features describing the above parameters for each day. As a basis for calculation of features, we took the timestamp of patients' input on the FOH. Based on this timestamp, we calculated average and standard deviation of the parameters either for the same day as the FOH was recorded or for the previous day, or for last 3 or 24 hours. Table 1 shows the time intervals in which the various features were calculated.

We refer to the features in Table 1 as dynamic, because they were obtained daily via sensors. But we also considered static features from the patients' health records, which generally did not change during the trial, e.g. age, gender, body mass index, heart rate at rest, CHF etiology, patient's comorbidities, ergometry maximum load, ... In this paper we will focus on modelling dynamic features.

3.2 Modelling methodology

We started the modelling with all the data of all the days when the patients reported their FOH. Then we had to address the issue of missing values: most features had the values for some of the days missing, and in many cases the missing values were quite numerous (because the patients did not use the HeartMan system fully on some days, or because of technical problems).

Missing values were filled in with imputation. The best imputation method on Chiron data proved to be k-Nearest Neighbours [6], which sets each missing value to the mean value of the same feature of k most similar instances. However, when a feature had a value missing on many days, imputation proved counterproductive, so we only used it on features with missing values on up to 30 days.

We then selected a number of thresholds for the maximum allowed number of missing values for a feature. We excluded all the features that had more missing values than each threshold, and we only included in further analysis the instances that had no missing values of the remaining features. Consequently, the stricter the threshold, the fewer features were left, and the more instances without missing values we could use.

The FOH was reported using the numerical representation, ranging from 1 (much better than usual) to 5 (much worse than usual). Like on the Chiron data, it proved too difficult to distinguish between all five classes, so we merged 1 and 2, as well as 4 and 5, ending up with three classes. Again,

Table 1: Dynamic features used for predictive modelling.

Parameters	Last 3 hours	The same day	Last 24 hours	Previous day
Skin temperature, Galvanic skin response, heart rate	✓	✗	✓	✗
Exercise parameters: duration, intensity, type, validity	✗	✓	✗	✓
Systolic blood pressure, diastolic blood pressure	✗	✓	✗	✓
Weight change since the previous day	✗	✓	✗	✗
FOH	✗	✓	✗	✗

like on the Chiron data, even this proved challenging, so we removed the middle class in some experiments, reducing the prediction to only two classes.

Two machine learning algorithms were used to build predictive models: Sci-kit Learn implementation of Decision trees (DT) and Extreme gradient boosting (XGB). DT are human-understandable models that can be used to gain an insight into the relations in the data, whereas XGB models are ensembles of Decision trees that typically offer higher accuracy at the expense of understandability. We compared the results of these two algorithms to the majority model, which always returns the class with the most instances.

4. RESULTS

Our experiment consisted of building three- and two-class models on dynamic features. We compared them for various missing-value thresholds and both machine-learning algorithms. The results are shown in Table 2.

Table 2: Classification accuracy of prediction on dynamic features.

Dataset information					
Missing values	$<\infty$	<318	<205	<130	<50
Features	72	30	22	18	13
Instances – 3 classes	221	349	592	686	745
Instances – 2 classes	91	143	229	275	316
Results three classes					
Majority	0.59	0.59	0.62	0.61	0.61
DT	0.56	0.52	0.44	0.43	0.5
XGB	0.65	0.57	0.55	0.56	0.56
Results two classes					
Majority	0.52	0.52	0.53	0.51	0.52
DT	0.67	**0.63**	0.63	0.6	0.55
XGB	0.7	**0.68**	0.54	0.54	0.58

We can see in Table 2 that the results with two classes are not much better than those of the majority model, so we focused our experiments on two classes. To obtain reasonable accuracy, we needed enough features, so we had to accept features with many missing values. We find the most satisfactory results those with the missing-value threshold of 318 (in bold). If we exclude Italian data from our dataset, the XGB accuracy for a comparable missing-value threshold goes up to 0.76. This is probably because of the higher number of missing values in the Italian data.

Table 3: Classification accuracy of prediction for individual patients.

Personal data	**None**	**10%**	**20%**
Majority	0.43	0.42	0.42
XGB	0.58	0.62	0.85

Figure 3 shows the DT with the missing-value threshold of 318. Labels value = [x, y] denote the number of instances when the patient felt good (x) vs. bad (y). Orange colour denotes feeling good and blue colour feeling bad. We can see that low systolic blood pressure generally means feeling bad, which makes sense in CHF patients who have problems with heart output. [4] The main exception is when the standard deviation of the heart rate is high, which is also reasonable, since high deviation means that the heart can adapt to varying demands. When the blood pressure is high, patients nevertheless feel bad when their average heart rate is high, which is also in line with expectation. Other parts of the tree make only minor contribution to the overall prediction.

Since the idea of the HeartMan project was to use predictive models to advise patients on how to improve their (feeling of) health, in our second experiment we tested models on individual patients. For each patient, we first built models on data of other patients, and then added 10% and 20% of that patient's personal data. The results for two classes are shown in Table 20. We can see that person-independent models (personal data = None) did not perform well, although they still outperformed the majority classifier by 15 percentage points. Adding 10% or 20% of personal data to the training data for the models improved the accuracy substantially.

5. CONCLUSIONS

The general models built on HeartMan data proved reasonably accurate. A direct comparison with Chiron models is difficult because the majority classifier there had a much higher accuracy, but the results can be considered comparable. The relations in the models seem in line with the

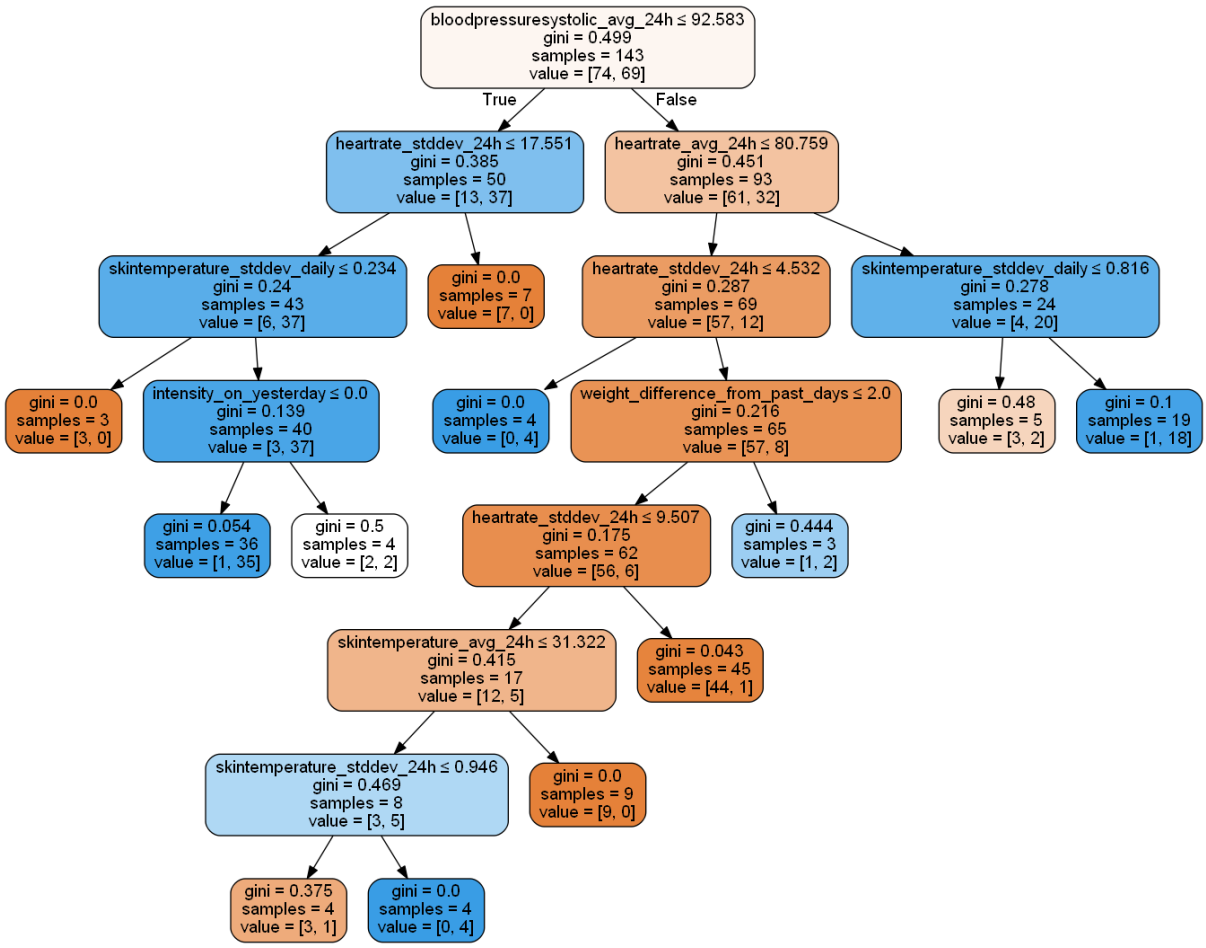

Figure 3: DT for two classes. Orange = feeling good and blue = feeling bad.

existing knowledge about CHF. Interestingly, the relations regarding ambient temperature and humidity, which were quite important in Chiron models, do not appear here.

The personalised models built on Chiron data were a disappointment, while they are fairly accurate in the case of HeartMan. It is difficult to say why this is the case. One possible reason is that the question about the FOH was such that the patients could answer more consistently: they were asked about their FOH compared to the usual one, whereas Chiron patients were asked about their FOH compared to the previous day. Regardless of the reason, these results are very encouraging, since building this type of models was a major objective of the HeartMan project. They certainly warrant further investigation, and are a strong argument for future research on predictive models in personal decision support systems and in health systems.

6. REFERENCES

[1] Personal decision support system for heart failure management, 2019. http://heartman-project.eu/.

[2] The chiron project, 2019. https://medrecord.io/project/chiron/.

[3] A. Baert, D. D. Smedt, J. D. Sutter, et al. Factors associated with health-related quality of life in stable ambulatory congestive heart failure patients: Systematic review. *European Journal of Preventive Cardiology*, 25(5):472–481, 2018. PMID: 29384392.

[4] T. T. Lee, J. Chen, D. J. Cohen, and L. Tsao. The association between blood pressure and mortality in patients with heart failure. *American Heart Journal*, 151(1):76 – 83, 2006.

[5] M. Mlakar, P. E. Puddu, M. Somrak, et al. Mining telemonitored physiological data and patient-reported outcomes of congestive heart failure patients. *PLOS ONE*, 13(3):1–23, 03 2018.

[6] M.-C. Monard. A study of k-nearest neighbour as an imputation method. 10 2002.

[7] P. Ponikowski, A. A. Voors, S. D. Anker, et al. 2016 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure: The Task Force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC)Developed with the special contribution of the Heart Failure Association (HFA) of the ESC. *European Heart Journal*, 37(27):2129–2200, 05 2016.