Hospitalisation Prediction from Telemonitoring Data in Congestive Heart Failure Patients

Božidara Cvetković[1], Aljoša Vodopija[1], Drago Rudel[2], Zdravko Balorda[2], Mitja Luštrek[1]
[1]Jožef Stefan Institute, Jamova cesta 39, Ljubljana
[2]MKS Electronic Systems Ltd., Rožna dolina, Cesta XVII/22b, Ljubljana
boza.cvetkovic@ijs.si

Abstract

Congestive heart failure is a common disease in people aged over 65 whose management can benefit from telemonitoring. In this paper we analyse the data from a year-long telemonitoring trial of 141 patients. The trial itself already reduced the number of hospitalisations, but our goal was to use machine learning to build a classifier that could predict the remaining ones. Such a classifier could then be used for timely interventions that could further reduce the number and/or duration of hospitalisations. By engineering a large number of features from the telemonitored parameters, and experimenting with various feature-selection and machine-learning methods, we built a Naïve Bayes classifier that predicted 7 hospitalisations out of 9, and raised a false alarm in only 1 instance out of 117.

1 Introduction

Congestive heart failure (CHF) is a progressive chronic condition in which the heart cannot pump enough blood to meet the needs of organs and tissues for oxygen and nutrients. As a result, the patients' capacity for physical activity is severely limited, making the disease highly debilitating. Being the most frequent cause of hospitalisation in people aged over 65 [Roger, 2013], it is also very expensive for the society. This makes effective management of CHF of paramount importance.

The management of some diseases, one of which is CHF, can be assisted remotely by telemonitoring [Kvedar *et al.*, 2014]. On one hand, this enables the patients to live at their homes as comfortably as possible (given their disease). On the other hand, it gives the clinicians an insight into the status of their patients' health, so they can make informed decisions regarding the treatment and utilise their time efficiently, making telemonitoring cost-effective.

Telemonitoring systems typically use automatic warnings based on thresholds which are set for each monitored parameter. The system thus warns clinicians to review the parameters and intervene if needed.

Telemonitoring has decreased the mortality and the number of hospitalisations of CHF patients in many cases [Martín-Lesende *et al*., 2013; Rudel *et al.*, 2016], although there are also reports to the contrary from two major trials [Sousa *et al.*, 2014]. However, the telemonitoring in these two trials was not very advanced – the monitored parameters were limited and no intelligent computer analysis was involved. To (further) decrease the number of hospitalisation, one can perform frequent laboratory tests [Pocock *et al*., 2006] or use a telemonitoring system which can detect more complex relations between the monitored parameters that are often not easily seen by clinicians. These relations can be extracted from data with machine-learning methods, which is the objective of this paper.

The analysis done in this paper was performed on the data of 141 patients collected through telemonitoring [Rudel *et al.*, 2016]. The telemonitoring system sent a warning to the clinicians in case some monitored parameter exceeded a predefined personalised threshold. A clinician intervened by calling the patient and inquiring how he/she feels. If the patient reported a deterioration of health, he/she was instructed to change the treatment, come for a check-up, or was hospitalised. This in itself reduced the number of hospitalisations from 0.34 per patient per year to 0.1. In this paper we attempt to predict the remaining hospitalisations from the telemonitoring data. This should enable even more timely interventions, further decreasing the number and/or duration of hospitalisations.

The paper is structured as follows. In Section 2 we present the collected dataset and in Section 3 the extracted features. In Section 4 we present the results of the data analysis aimed at prediction hospitalisations, and then conclude with Section 5.

2 Dataset

The dataset was collected during a telemonitoring trial conducted by the General Hospital Slovenj Gradec [Hospital, 2016] in Slovenia within the United4health project [United4Health, 2016]. They recruited 141 patients with different New York Heart Association (NYHA) Functional Classification for CHF. The trial took place in 2014 and 2015 for an average period of 369 days per patient.

One year before the start of trial each patient was tested in a laboratory to establish the N-terminal pro b-type

natriuretic peptide (primary proBNP) and left ventricular ejection fraction (primary LVEF). The former is an established biomarker of the severity of CHF, and the latter is a measure of cardiac function. Right before the trial the patients underwent the same tests (secondary proBNP and secondary LVEF). They were additionally checked for other cardiovascular diseases relevant for CHF: hypertensive heart disease with heart failure (I110), old myocardial infarction (I252), primary pulmonary hypertension (I270), nonrheumatic mitral (valve) insufficiency (I340), dilated cardiomyopathy (I420), ventricular fibrillation and flutter (I490) and right ventricular failure (I500). For each patient, the weight, height, sex, date of birth, date of the first diagnosis and the history of hospitalisations were also noted.

The patients were given a set of devices and were instructed to take daily measurements in the morning before breakfast and in case of health deterioration to repeat the measurements several times during the day. The measured parameters are the systolic (SYS_BP) and diastolic (DIAS_BP) blood pressure, the heart rate (HR), the blood oxygen saturation (SO_2), the weight (WE) and the occurrence of arrhythmia (ARR). The blood pressure, heart rate and arrhythmia were measured with Cignus Senior Line TD-3128 [Cignus, 2016], the oxygen saturatuion was measured with Nonin Onyx II 9560 [Nonin, 2016] and the weight with Libr-O-Graph [Libro, 2016]. If a patient did not attend the laboratory tests or did not measure some parameter, this is noted as a missing value in the dataset.

Demographic variables	
Men [no.]	102
Women [no.]	39
Age [mean ± sd]	71.5 ± 9.4
General information about the patiens	
Trial duration [mean ± sd]	369 ± 134
Disease duration in years [mean ± sd]	3.5 ± 3.2
Hospitalisation history [mean ± sd]	0.4 ± 0.7
Height [mean ± sd]	167 ± 9.1
BMI [mean ± sd]	30.3 ± 5.2
Clinical CHF variables	
Primary LVEF [mean ± sd]	43 ± 13
Secundary LVEF [mean ± sd]	41 ± 13
Primary ProBNP [mean ± sd]	4179 ± 4926
Secondary ProBNP [mean ± sd]	3406 ± 3507
NYHA class 1 [no.]	1
NYHA class 2 [no.]	101
NYHA class 3 [no.]	35
NYHA class 4 [no.]	1
Other cardiac dieseses	
I110 [%]	59.6
I252 [%]	19.1
I270 [%]	0.7
I340 [%]	1.4
I420 [%]	5.7
I490 [%]	0.7
I500 [%]	29.8

Table 1. Trial data statictics. We present the demographic statistics, the information about the patients, the CHF variables and other dieseses patients have.

The statistics of the trial and the patients involved are presented in Table 1. We can observe that more men than women were involved, and that they were mostly diagnosed with CHF from 3 to 6 years prior to the beginning of the trial. Most of the patients were categorised as NYHA class 2 (patients with mild symptoms – mild shortness of breath) and class 3 (comfortable only at rest – shortness of breath even in light activity). All of them had at least one additional cardiovascular disease – more than half were coping with the hypertensive heart disease with heart failure (I110). The number of hospitalisation prior to start of the trial was mostly 0, while the maximum number was 4.

A sample of the collected raw data is presented in Figure 1 and Figure 2. Figure 1 shows raw signals of body mass index (BMI), systolic blood pressure, heart rate and oxygen

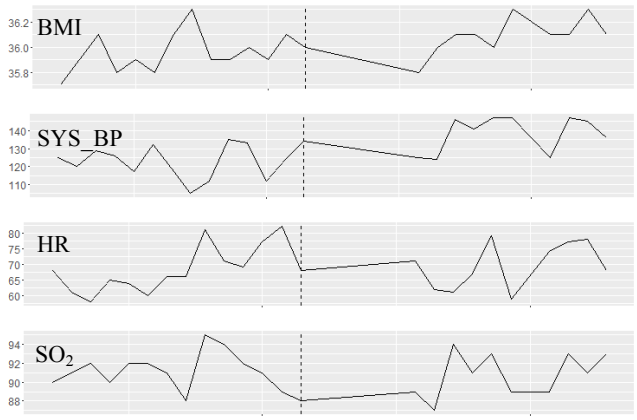

Figure 1. Raw signals of the patient who was hospitalised at the time represented with vertical line. The signals are the body mass index, systolic blood pressure, heart rate and oxygen saturation.

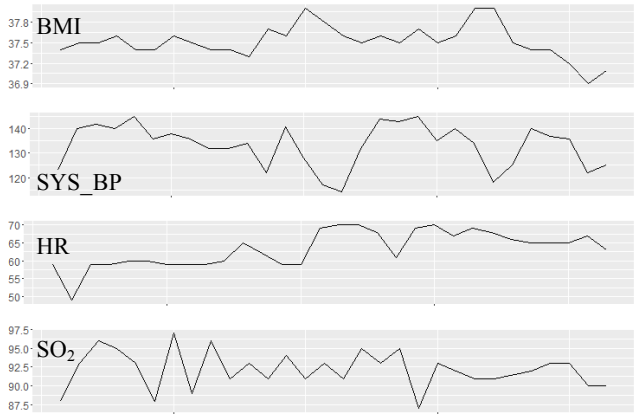

Figure 2. Raw signals of the patient who was not hospitalised. The signals are the body mass index, systolic blood pressure, heart rate and oxygen saturation.

saturation for a patient who was hospitalised at the time point marked with a vertical line. Figure 2 presents the same raw signals, but for a patient who was not hospitalised. We can observe that the BMI of the hospitalised patient before the hospitalisation does not differ much from the BMI of non-hospitalised patient. The same goes for the systolic blood pressure, heart rate and oxygen saturation. From this we can conclude that the dataset is difficult to analyse. Patterns which occur prior to hospitalisations often also occur at other times.

3 Features

The parameters in the dataset can be divided into static parameters, which change little during the trial, and dynamic parameters, which change with each new telemonitoring measurement. The static parameters are the NYHA class, ProBNP (preferably secondary if it exist, otherwise the primary), LVEF (preferably secondary if it exist, otherwise the primary), age, sex, disease duration in years, hospitalisation history and other cardiovascular diseases. The dynamic parameters are the weight, heart rate, systolic and diastolic blood pressure, oxygen saturation, number of interventions (phone calls to the patient), number of detected arrhythmias and number of hospitalisation during the trial until the analysed date.

Each parameter was transformed into features used for machine learning in the following three ways:

- Raw feature values
- Discretised feature values
- Extracted statistical feature values

Raw feature values are the raw values of both the static and dynamic parameters, and additionally two features calculated from the raw values: the BMI and the difference between the systolic and diastolic blood pressure.

Discretised feature values represent the risk carried by the parameter values (high, medium, low). The discretisation is based on the thresholds presented in Table 2. The thresholds were obtained from the research done in the CHIRON project [Chiron, 2016], in which a survey among 32 European opinion leaders in cardiology was performed as a means to obtain an evaluation of the parameters relevant to CHF [Kozina *et al.*, 2013]. Each feature is also characterised by the "shape" of the relation between the parameter values and the risk. If the relation is linear (Figure 3), then the parameter is directly proportional (+) or inversely proportional (–) to the feature (risk). Interestingly, a decrease of systolic blood pressure in CHF patients means a higher risk. If the feature has a U-shaped relation (this is the case only for the BMI in this dataset), it means that it has two high thresholds and two low thresholds (Figure 4).

	Risk			
Parameter	**Low**	**Medium**	**High**	**Relation**
LVEF	50	40	25	linear (–)
NYHA	2	3	4	linear (+)
Age	58	64	81	linear (+)
Systolic BP	140	120	100	linear (–)
Diastolic BP	70	80	95	linear (+)
SO_2	94	88	82	linear (–)
BMI	21–23	18–21 23–30	<18 >30	U-shape

Table 2. The discretisation thresholds.

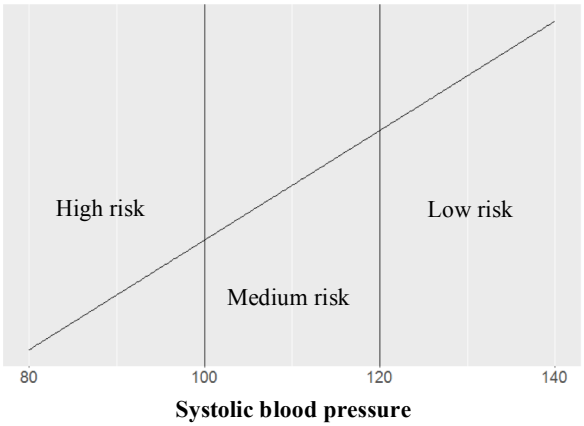

Figure 3. Trasholds and linear relation of the paramter systolic blood pressure.

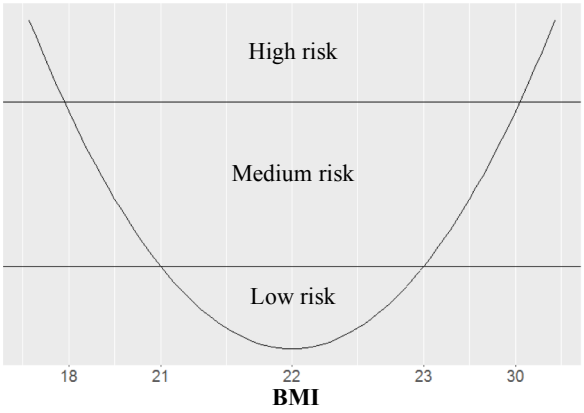

Figure 4. thresholds and U-shape relation of the parameter BMI.

Extracted statistical features were obtained by statistical processing of the raw and discretised features over four time intervals (4, 14, 30 and 90 days). For every time interval we calculated the following statistics: the average value, the discretised average value (with thresholds from Table 2), the number of times the value exceeded the high-risk threshold, the standard deviation, the amount of missing values, the relative deviation from the patient's average value for that feature, and the trend calculated with linear regression as implemented in the R statistical tool [R, 2014].

We ended up with 168 features from which feature vectors were constructed. To these feature vectors we added the class, which indicates whether the patient was hospitalised at the end of the interval for which the features were computed, or not. The features were finally evaluated with the information gain as implemented in the Weka

machine-learning suite [Weka, 2009]. The information gain of a feature corresponds to the amount of information about the class obtained by knowing the value of the feature. The information gains of the top-rated features are shown in Table 3.

To illustrate the computation of the features, Figure 6 and Figure 5 show some features of the two patients whose raw parameters are shown in Figure 1 and Figure 2. The blue line is the trend calculated with linear regression, and the band around it is the standard deviation of the parameter. We can see that there is again no obvious difference between the data of the hospitalised patient before the hospitalisation, and the data of the non-hospitalised patient.

4 Hospitalisation Prediction

Our goal was to construct a classifier which will predict hospitalisations before they occur. Our dataset has only nine occurrences of hospitalisation for eight patients – one patient was hospitalised twice. This gives nine hospitalisation instances. We then randomly selected one interval from each patient to create 117 non-hospitalisations instances (we omitted 15 patients due to many missing values for all the measured parameters).

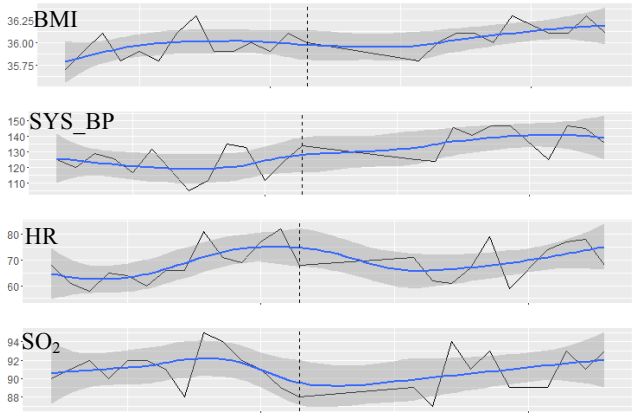

Figure 6. Measured parameters and extracted statistical features for a hospitalised patient.

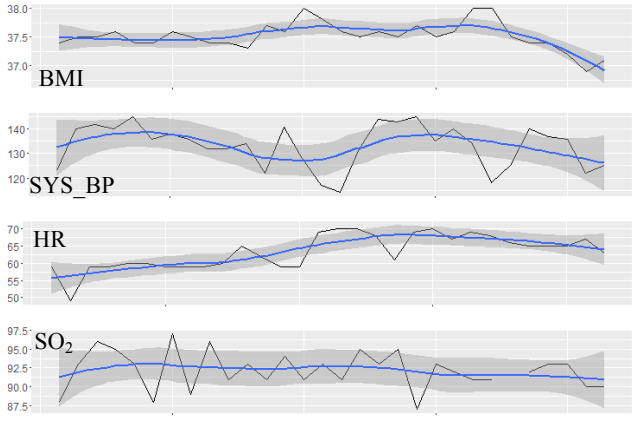

Figure 5. Measured parameters and extracted features for a non-hospitalised patient.

Feature	Information gain
Amount of missing values for SO_2 (90 days)	0.099
Amount of low SO_2 (30 days)	0.099
Relative deviation of DIAS_BP (90 days)	0.098
Trend of BMI (30 days)	0.098
Amount of missing values for BMI (30 days)	0.071
Relative deviation of ARR (4 days)	0.066
The NYHA classification	0.065
Amount of missing values for BMI (14 days)	0.064
Trend of BMI (90 days)	0.064
Trend of diff. beetwen presures (90 d.)	0.060
Trend of SYS_BP (90 days)	0.053
Amount of missing values for HR (90 days)	0.052
Amount of ARR (4 days)	0.051
Trend of SYS_BP (30 days)	0.044
Trend of DIAS_BP (30 days)	0.038
Average value of SO_2 (30 days)	0.031

Table 3. The top-rated features according to the information gain.

The evaluation was performed with the leave-one-hospitalisation-out approach: we created nine training datasets and nine test datasets. The training dataset was always created from randomly chosen 104 instances of non-hospitalisation and 8 hospitalisation instances. To balance the training dataset, we multiplied the hospitalisation instances to the same number as the non-hospitalisation instances. The test dataset contained the left-out hospitalisation instance and the 13 left-out non-hospitalisation instances.

The prediction was performed with four machine-learning algorithms as implemented in the Weka suite: J48 decision trees, Random Forest (RF), Support Vector Machine (SVM) and Naive Bayes (NB). The first experiment was performed on instances containing all the features, and returned rather poor results: we were not able to predict a single hospitalisation, so the recall was 0.

To improve the results, we tried out several feature selection methods, again as implemented in the Weka suite: Information Gain, ReliefF, Correlation-based Feature Subset Selection, Wrapper, and Chi-squared correlation test. The results are shown in Table 4 in terms of recall (the fraction of hospitalisation instances classified as such), precision (the fraction of instances classified as hospitalisation that truly belong to the hospitalisation class), the F-measure (the harmonic mean of the precision and recall), and the area under the receiver operating characteristic (ROC) curve (AUC, explained later alongside Figure 7). The approaches yielding the best results were the wrapper-based approach (using the Naïve Bayes or SVM) and the correlation-based feature subset selection using the SVM machine-learning algorithm. For these we performed the Fisher's exact test for evaluating the statistical significance in terms of the probability that our results were not obtained by chance. The wrapper method using Naïve Bayes gained the highest statistical significance with $p < 3 \times 10^{-9}$, the wrapper

Information Gain				
Algorithm	**J48**	**RF**	**SVM**	**NB**
Recall	0.78	0.00	0.56	0.44
Precision	0.58	NA	0.50	0.57
F-measure	0.67	NA	0.53	0.50
AUC	0.86	0.95	0.91	0.87
ReliefF				
Algorithm	**J48**	**RF**	**SVM**	**NB**
Recall	0.22	0.00	0.11	0.22
Precision	0.29	NA	**1.00**	0.22
F-measure	0.25	NA	0.20	0.22
AUC	0.69	0.75	0.75	0.42
Correlation-based Feature Subset Selection				
Algorithm	**J48**	**RF**	**SVM**	**NB**
Recall	0.56	0.33	0.78	0.78
Precision	0.50	0.75	0.78	0.64
F-measure	0.53	0.46	0.78	0.70
AUC	0.73	0.96	0.95	0.94
Wrapper				
Algorithm	**J48**	**RF**	**SVM**	**NB**
Recall	0.33	0.22	**0.89**	0.78
Precision	0.60	0.67	0.62	0.88
F-measure	0.43	0.33	0.73	**0.82**
AUC	0.71	0.58	0.94	**0.98**
Chi-squared correlation test				
Algorithm	**J48**	**RF**	**SVM**	**NB**
Recall	0.77	0.22	0.56	0.44
Precision	0.64	**1.00**	0.50	0.57
F-measure	0.70	0.36	0.53	0.50
AUC	0.87	0.94	0.90	0.87

Table 4. The results of hospitalisation prediction for features selected with various feature-selection methods, and various machine-learning algorithms. The best results for each metric is shown in bold.

method using SVM gained $p < 8 \times 10^{-9}$ and the correlation-based feature subset selection gained $p < 1.5 \times 10^{-8}$.

Since the Naïve Bayes achieved the highest f-measure and statistical significance, we will give it more attention in the rest of the paper. This method selects the features using an internal cross-validation. The following features were selected:

- Average heart rate in the last 4 days
- Average of the difference between systolic blood pressure and diastolic blood pressure in the last 4 days
- Absence of measurements for diastolic blood pressure, oxygen saturation and BMI in the last 90 days
- Systolic blood pressure trend in the last 90 days
- Trend of the difference between systolic blood pressure and diastolic blood pressure in the last 90 days
- BMI trend in the last 30 days
- Systolic blood pressure trend in the last 30 days

To better see how the best the classifier predicted hospitalisations, we show its confusion matrix in Table 5. The classifier successfully predicted 7 hospitalisations out of 9, and raised a false alarm in 1 instance out of 117.

Predicted / **True**	Hospitalisation	Non-hospital.
Hospitalisation	7	2
Non-hospital.	1	116

Table 5. Confusion matrix of the classifier built with the wrapper-based feature selection and Naive Bayes algorithm.

Figure 7 shows the ROC curve of the best classifier. The classifier's output is the probability that an instance belongs to the hospitalisation class, and the curve display the false and true positive rate for every threshold that can be applied to this probability to transform it into a crisp hospitalisation/non-hospitalisation output. The false positive rate is the fraction of instances mistakenly classified as hospitalisation out of all non-hospitalisations, and the true positive rate is the fraction of hospitalisation instances classified as such (the same as recall). The closer the curve is to the upper left corner, the better the classifier. One can use the curve to select a trade-off between the false and true positive rate suitable for his/her application.

Finally, the best classifier is visualised as a nomogram in Figure 8. A nomogram shows how many "points" each

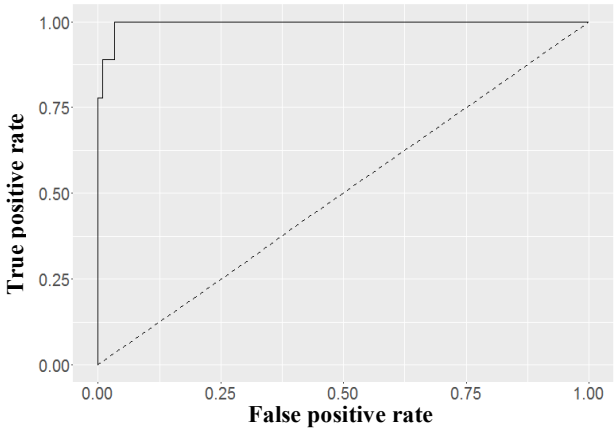

Figure 7. Receiver operating characteristic (ROC) curve of the classifier built with the wrapper-based feature selection and Naive Bayes algorithm.

feature contributes to the hospitalisation class – this is shown by the position of feature values on the line next to each feature, relatively to the scale on the top of the figure (e.g., a heart rate between 100 and infinity contributes –59 points). On the bottom of the figure one can see how the points translate into the probability of hospitalisation (e.g., 0 points correspond to the probability of 0.075 and 23 points to the probability of 0.5).

To better understand a nomogram, we present an example of a patient who was hospitalised. Table 6 presents the features, their values and the number of points they contribute to the hospitalisation classification. In the example, the sum of the points is 38, which indicates that the patient will be hospitalised with the probability of 87%.

Feature	Feature value	Points
Average HR (4 days)	60–100	0
Average diff. between pressures (4 d.)	0–30	20
Missing values for BMI (90 days)	10	8
Missing values for DIAS_BP (90 days)	9	8
Missing values for SO_2 (90 days)	8	–11
Missing values for HR (90 days)	9	1
Trend of SYS_BP (90 days)	–1	–1
Trend of diff. between pressures (90 d.)	0	10
Trend of BMI (30 days)	0	–3
Trend of SYS_BP (30 days)	–1	6
Sum		**38**

Table 6. The number of points each feature contributes to the hospitalisation classification.

According to the nomogram, a high heart rate is a strong indicator that a hospitalisation will not occur, which is unusual since a high heart rate is considered a bad sign in CHF patients. A low difference between the diastolic and systolic blood pressure is a strong indicator of hospitalisation, which is normal, since it also indicates a poor cardiac function. The presence of missing values in the data has a somewhat inconsistent meaning, but it seems to indicate hospitalisation more strongly than non-hospitalisation, which can be explained by missing values being caused by omitted measurements due to poor health. An upward trend of the systolic blood pressure is a strong indicator that a hospitalisation (because of CHF) will not occur, which is reasonable. However, a downward trend of the difference between the diastolic and systolic blood pressure indicates the same, which seems unusual. A change in the BMI in any direction is a strong indicator of hospitalisation – increasing BMI signals fluid retention (a common problem of CHF patients), while decreasing BMI probably signals generally poor health. Finally, a stable systolic blood pressure indicates that a hospitalisation will not occur. In summary, the classifier contains some relations consistent with the current medical knowledge about CHF, and some unexpected relations that are probably wrong because of too little hospitalisation data. However, some of the unexpected relations may also be correct and may represent novel findings about CHF that can only be obtained through the analysis of telemonitoring data – which is still relatively novel.

5 Conclusion

In this paper we analysed the data from a telemonitoring trial of CHF patients with machine-learning methods. Our goal was to build a classifier that can predict hospitalisations from daily measurements of patients' vital signs. Such a classifier can be used to warn of an upcoming hospitalisation, so that clinicians can make a timely intervention to prevent the hospitalisation or reduce its duration. The best classifier was Naïve Bayes which successfully predicted 7 hospitalisations out of 9, and raised

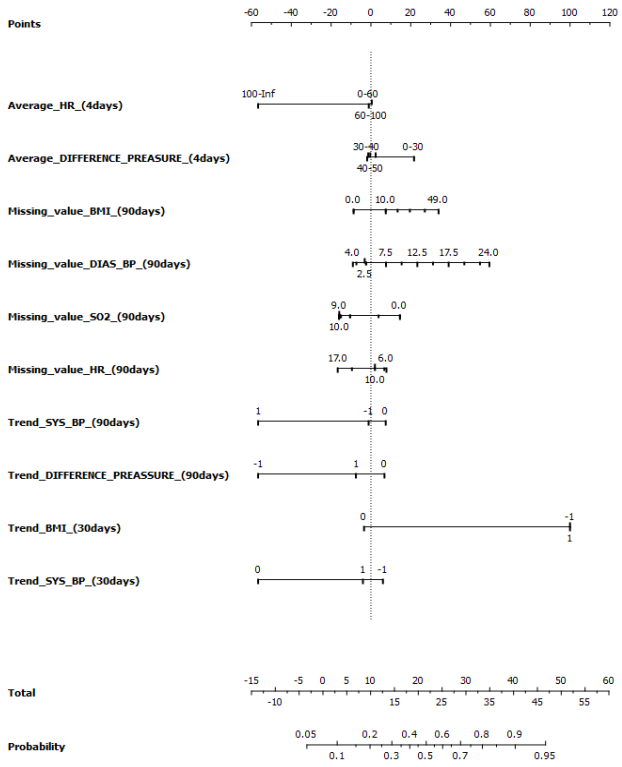

Figure 8. The classifier built with the wrapper-based feature selection and Naive Bayes algorithm visualised as a nomogram.

a false alarm in 1 instance out of 117, which we consider very successful.

In the future we plan to more rigorously check whether it is possible that our classifiers overfit the data and would consequently not be so successful in real life. Furthermore, we will test how far in advance can the hospitalisations be predicted, since an earlier prediction is clearly more valuable than a later one. Finally, we will investigate the possibility of deploying the classifier live in telemonitoring activities of the General Hospital Slovenj Gradec.

Acknowledgments

This work was funded by the United4Health and HeartMan projects. The United4Health project has received funding from the European Union's ICT Policy Support Programme (ICT PSP) under grant agreement No 325215. The HeartMan project has received funding from the European Union's 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

[Chiron, 2016] Chiron project, https://medvision360.com/d/chiron-project/, access: 2015-04-15.

[Cignus, 2016] Cignus Senior Line TD-3128, http://www.taidoc.com/Blood-Pressure-Monitor-TD-3128.html

[Hospital, 2016] Splošna bolnišnica Slovenj Gradec, http://www.sb-sg.si/, access: 2015-04-15.

[Kozina, 2013] Simon Kozina, Paolo Emilio Puddu, and Mitja Luštrek. System for Supporting Clinical Professionals Dealing with Chronic Disease Patients. In *AmI 2013 Workshops, Evolving Ambient Intelligence*, pages 79–85, Dublin, Ireland, December 2013. Springer.

[Kvedar *et al*., 2014] Joseph Kvedar, Molly Joel Coye, and Wendy Everett. Connected Health: A Review Of Technologies And Strategies To Improve Patient Care With Telemedicine And Telehealth. *Health Affairs,* 33(2):194–199, February 2014.

[Libro, 2016] Libr-O-Graph, http://www.iem.de/en/products/libr-o-graph.html

[Martín-Lesende *et al*., 2013] Iñaki Martín-Lesende, Estibalitz Orruño, Amaia Bilbao, Itziar Vergara, Carmen Cairo, Juan Carlos Bayón, Eva Reviriego, María Isabel Romo, Jesús Larrañaga, José Asua, Roberto Abad, and Elizabete Recalde. Impact of Telemonitoring Home Care Patients with Heart Failure or Chronic Lung Disease from Primary Care on Healthcare Resource Use. *BMC Health Services Research*, 13(1):1–15, March 2013.

[Nonin, 2016] Nonin Onyx II 9560, http://www.nonin.com/Onyx9560

[Pocock *et al*., 2006] Stuart J. Pocock, Duolao Wang, Marc A. Pfeffer, Salim Yusuf, John J.V. McMurray, Karl B. Swedberg, Jan Ostergren, Eric L. Michelson, Karen S. Pieper, and Christopher B. Granger. Predictors of Mortality and Morbidity in Patients with Chronic Heart Failure. *European Heart Journal*, 27:65–75, October 2006.

[R, 2014] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, 2014.

[Roger, 2013] Véronique L. Roger. Epidemiology of heart failure. *Circulation Research*, 113:646–659, 2013.

[Rudel *et al.*, 2016] Drago Rudel, Cirila Slemenik-Pušnik, Zdravko Balorda, Stanislav Pušnik, Janez Lavre, and Majda Kladnik. Reducing Hospitalisation by Providing Telemedicine Support to CHF Patients at Home in Slovenia. In *Global Telemedicine and eHealth Updates: Knowledge Resources*, 9:205–208, 2016.

[Sousa *et al.*, 2014] Carla Sousa, Sérgio Leite, Ricardo Lagido, Liliana Ferreira, José Silva-Cardoso, and Maria Júlia Maciel. Telemonitoring in Heart Failure: A State-of-the-art Review. *Revista Portuguesa de Cardiologia,* 33(4):229–239, 2014.

[United4Health, 2016] United4Health project, http://united4health.eu/, access: 15.04.2015.

[Weka, 2009] Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H. Witten. The WEKA Data Mining Software: An Update; *SIGKDD Explorations*, 11(1):10-18, 2009.