Additional Information

Latent states extraction through Kalman Filter for the prediction of heart failure decompensation events

D. Nunes, T. Rocha, V. Traver, C. Teixeira, M. Ruano, S. Paredes, P. Carvalho, J. Henriques*

Abstract— **Cardiac function deterioration of heart failure patients is frequently manifested by the occurrence of decompensation events. One relevant step to adequately prevent cardiovascular status degradation is to predict decompensation episodes in order to allow preventive medical interventions.**

In this paper we introduce a methodology with the goal of finding onsets of worsening progressions from multiple physiological parameters which may have predictive value in decompensation events. The best performance was obtained for the model composed by only two features using a telemonitoring dataset (myHeart) with 41 patients. Results were achieved by applying leave-one-subject-out validation and correspond to a geometric mean of 83.67%. The obtained performance suggests that the methodology has the potential to be used in decision support solutions and assist in the prevention of this public health burden.

I. INTRODUCTION

Cardiovascular diseases (CVDs) were responsible for 31% of all global deaths in 2015, still remaining the leading cause of morbidity and mortality worldwide [1]. Contrarily to the considerable decrease in mortality rates observed for many CVDs (including acute coronary syndromes, valvular and congenital heart disease, uncontrolled hypertension and several arrhythmias), congestive heart failure (HF) syndrome has been associated with an increasing mortality rate [2]. In fact, only 50% of HF patients get the chance to live more than five years after diagnosis [1]. Furthermore, as consequence of gradual ageing of the population and improvements in prognosis of acute cardiac events, projections indicate an increase by 25% of the HF prevalence in the next 20 years [1].

The progression of HF originates recurrent hospitalizations (acute decompensation events) and, even though the symptoms are reduced, the patient's cardiac function continually deteriorates [3]. In consequence, each acute congestion event contributes to a decreased level of cardiac function, leading to further progression of the patient's HF.

Studies reporting HF predictors of morbidity, mortality, destabilizations (sudden alternating periods of apparent stability) and re-hospitalizations can be found in literature [4], [5]. The majority of these studies report the use of blood pressure (BP) as a predictor, which is expected due to the characterization and pathophysiology of the cardiovascular diseases, as well as a cause of decompensation [5]–[14].

Other characteristic widely used in these studies is the body weight (BW), as an increase in this parameter is normally related to fluid retention, being a potential measure of how much fluid is being retained. In this perspective several telemonitoring studies have observed an increase in body weight, associated with an increase in body fluids [14], [15]. Particularly, Koulaouzis and colleagues, report the use of body weight, as also heart rate (HR) and blood pressure for the prediction of HF re-hospitalizations using a telemonitoring dataset. Inspecting the performance obtained with different combinations of those parameters, it was found that the best performance obtained was with body weight and diastolic blood pressure [12], achieving 82% of area under the ROC curve. Besides these two predictors, studies carried out by [6]–[8] also included other physiological data including HR, breathing rate (BR), body temperature and total body water. The results were obtained for clinical data acquired during cardiologists' visits to the patients at home and therefore using a more controlled procedure compared to telemonitoring collection of data. The best performance (sensitivity of 100% and false positive rate of 2.78%) among the works reported in [6]–[8] was obtained in [7], where an independent testing dataset, strongly imbalanced (38 instances and 2 decompensation cases), was analyzed by using the methodology in [7]. The second best performance (sensitivity of 91% and false positive rate of 16.21%) was obtained using a SVM hyper solution framework in a new dataset in [8].

Henriques et al. [11] also investigated the predictive power of four bio signals (BW, BP, HR and BR acquired during a telemonitoring study) for the anticipation of HF decompensation events. The progression of the time series was inspected (trend similarity analysis) and a nearest neighbor classifier was applied to data resulting in a $F_{measure}$ of 78% and a G_{mean} of 80%. Furthermore, intracardiac impedance has also been recently reported to decrease as result of HF worsening being a strong candidate as HF decompensation predictor [16].

Based on the aforementioned, it is crucial the existence of approaches that predict and prevent HF decompensation events and consequently the worsening of HF patients. In the following sections a new method to detect high risk of a decompensation event is described, which is based on the fusion of latent states from different physiological parameters.

II. DATA

Before the introduction of the proposed methodology, the used data is described, as its characteristics are determinant for the methodology obtained. The data used is from MyHeart dataset [11], a heart failure telemonitoring study

*D. Nunes, C. Teixeira, P. Carvalho, J. Henriques and M. Ruano are with the Centre for Informatics and Systems, Polo II Coimbra, 3030-290, Portugal, (e-mails: {dbnunes, cteixei, carvalho, jh}@dei.uc.pt and mruano@ualg.pt).

T. Rocha and S. Paredes are with the Coimbra Polytechnic - ISEC, Coimbra, Portugal, (e-mails: {teresa, sparedes}@isec.pt).

V. Traver is with the Institute for Information and Communication Technologies, Universitat Politècnica de València, Valencia, 46022, Spain, (e-mail: vtraver@itaca.upv.es).

with annotated heart failure episodes, which is a private dataset composed by daily physiological parameters such as body weight (BW), systolic blood pressure (BP), heart rate (HR), breathing rate (BR), external and internal thoracic impedance, (RE) and (RI), respectively, and finally the daily average QRS duration (QRSm).

This study enrolled 148 patients from six clinical centers in Germany and Spain that agreed with data acquisition and processing under anonymous conditions. During the study the patients were requested to measure the physiological measurements referred above, with the exception of the QRSm parameter, which was calculated from the acquired ECG recurring to an automatic algorithm. As a consequence of the ambulatory nature of the study, the data produced is composed by missing values and noise. Consequently, only 41 patients were considered analyzable. Additionally, six cardiologists have analyzed the data, identifying which patients had experienced a decompensation event (16 patients) and which patients had not (25 patients). The data used for the construction and validation of the methodology presented in the next sections, corresponds only to a 40-day snapshot of the physiological parameters for each patient, random in the control case, while in the decompensated patients, the 40 days prior the decompensation event.

III. METHODS

The proposed methodology aims to obtain an instantaneous fused risk score of a decompensation to occur. Instantaneous in this context refers to the fact that the features from the time-series do not require a windowing process, as they are extracted instantaneously as a new sample of the raw signal is available, and this way, the features present the same scale as the original time-series and the risk score is obtained instantaneously through a linear combination of those features. This methodology is achieved recurring to a Kalman Filter formulation for the feature extraction phase, and a logistic regressor for the fusion step and risk score calculation.

The main steps of the proposed methodology are listed in the following bullets:

- Feature extraction: a Kalman Filter formulation;
- Feature Selection: using lasso regularization
- Risk score calculation and validation: fusion of multiparametric latent states

A. Feature extraction: a Kalman Filter formulation

Due to the ambulatory nature of the database, the physiological parameters present a high degree of noise and missing value presence. This characteristic is a challenge in the feature extraction procedure once the calculation of the features may be corrupted by artifacts that do not represent the reality, or even impossible due to the nonexistence of measurements. Conveniently, the Kalman Filter framework tackles these two challenges by relying both on measurements and on a given dynamical system formulation, in order to obtain the optimal state estimation. Using this approach is not only possible to infer about missing observations, but to also extract filtered latent states depending on the system definition. Assuming that the time-series follow a given progression/trend and are subject to some error, both in the system and in the measure-
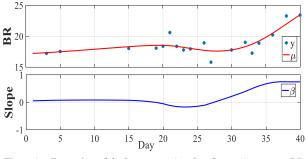

Figure 1 – Extraction of the latent state (used as features) on a raw BR signal.

ments, it is possible to model the physiological parameters by a local linear trend model, i.e., the next filtered value μ_t of a given raw signal y will be equal to the previous value μ_{t-1} plus a trend term β_{t-1} and an error term, allowing the term trend to also vary over time.

$$\mu_t = \mu_{t-1} + \beta_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma_\mu) \tag{1}$$

$$\beta_t = \beta_{t-1} + \tau_t, \quad \tau_t \sim N(0, \sigma_\beta) \tag{2}$$

$$y_t = \mu_t + \omega_t, \quad \omega_t \sim N(0, \sigma_y) \tag{3}$$

The formulation present in the Equation (1-3) allows the decomposition of the physiological parameter into two different components, the filtered state of the time series (μ), and a component of the instantaneous slope values (β), as can be seen in Figure 1. The parameters and latent states are determined by Maximum Likelihood Estimation (MLE). Inspecting the residuals of the filtered process is observed that the majority follow a gaussian distribution, as assumed by the Kalman Filter definition. Although the residuals on some cases do not follow the gaussian distribution, the filtering process still manages to provide a reasonable estimation of the latent states, i.e., the centrality and trend extraction.

In the HF context, the physiological parameters normally present a given trend anteceding a decompensation event, which are respective to the worsening of symptoms and to the increasing difficulty to fulfil the body needs, which eventually may lead to a decompensation event. In Figure 1 is illustrated the onset of an increasing trend around the 30th day, which can be an indicator of worsening of the cardiac status and may have predictive value in the detection of a possible event to occur. Both instantaneous slopes and values are computed for each parameter, as the values of the physiological parameters (aside its trends) may also indicate to some extent the risk of a decompensation event. Aside these two features, the absolute value of the slope is also taken into consideration for each parameter, as an increasing or decreasing trend may have predictive value against a null trend, which represents a steadied condition. In conclusion, for each existing bio-signal are extracted three features with the same scale of the raw signal: the filtered parameter, the instantaneous slope and the absolute value of the slope.

B. Feature selection: using lasso regularization

Given that only 41 patients are available and a total of 21 features (three per bio-signal) are present, a feature selection procedure is crucial in order to prevent high degrees of freedom and overfitting effects. The chosen feature selection procedure was the logistic regression with lasso regularization. A embedded method was chosen in favor of filter and wrapper methods once the filter methods do not take into account the relations between the different features, while the wrapper methods are not suitable for low sample sized datasets once the results obtained are highly biased [17]. By following this procedure, the most relevant features can be identified by the magnitude of the regression weights θ, while the redundant and correlated ones are eliminated due to the regularization factor $(-\lambda \sum |\theta_i|)$ added to the log likelihood function of the logistic regression. The regularization weight λ was set by default to $\lambda_1.10^{-4}$, being λ_1 the largest lambda that provides a nonnull model.

In order to compute the regression weights, the data must be manipulated to obtain a classification task. Each sample may be one of two classes, control or prior decompensation. The control dataset is composed by all the feature samples in the 40 days window from all the control cases. While in the prior decompensation case, only the feature values that correspond to the day anteceding the decompensation are fed to the classifier. This is done to impose the behavior that in the day anteceding a decompensation, the risk must be maximum. Moreover, using this approach, it is not assumed the starting point of high risk decompensation, which may vary given the patient from 1 to 40 days anteceding the event. However, this selection causes a high degree of imbalance in the training procedure, once all 40 days are used in the control patients, while only one day is used in the decompensated patients. To tackle this problem, different error weights are given for each class during the training procedure in order to uniform the prior probabilities of each class.

Another problem lies in the possibility that a given test case do not present all the parameters taken into account in the feature selection procedure, this due to the ambulatory acquisition nature of the physiological parameters, where some parameters may not be acquired possibly due to system failure. This situation is observed in three patients, two of them do not present measures of RI and RE (or at least not the sufficient ones to extract the latent states be Kalman Filtering), and the other only presents measures of BW, HR and BP. In order to prevent the mismatch of the most relevant features, and the ones present in a given test case, the search for the relevant features in the training set is limited to the ones which are present in the test case, this way, the method can adapt itself depending on the situation and not assuming that a given set of features will be present in a real case scenario.

C. Risk score calculation and validation: fusion of multiparametric latent states

Once the most relevant features are identified and the irrelevant ones are discarded based on the magnitude of the regression weights with lasso regularization, a risk score is pursued. This risk score is obtained once again by a logistic regression on the encountered features with uniform prior probabilities, but this time without regularization. At the time of computing the risk scores for unseen cases, the previous trained model can be employed, and an alarm of an incoming decompensation event can be emitted if the risk score surpasses a given threshold, in this case set to 0.5, which is the central value in the possible range of 0 and 1.

The methodology is validated in a Leave-One-Subject-Out (LOSO) manner. Before the results are presented, it is important to define what is considered a true positive (TP), a true negative (TN), a false positive (FP) and a false negative (FN). For validation purposes, a TN is considered when the risk score of a control subject never surpasses the alarm threshold in all 40 days, if it does, it is considered a FP, even if the risk score only surpasses the threshold for just one day. A TP is considered when the risk score of a subject with an incoming decompensation surpasses the given threshold and stays in the alarm region until the decompensation occurs, if the risk score by any chance, returns to normal values (below the threshold) before the decompensation it is considered a FN, once this is not a desired behavior. Using these definitions, the geometric mean (G_{mean}) is calculated as metric of model evaluation.

$$SE = \frac{TP}{TP + FN} \qquad (4)$$

$$SP = \frac{TN}{TN + FP} \qquad (5)$$

$$G_{mean} = \sqrt{SE \times SP} \qquad (6)$$

In Equation (4), SE corresponds to the sensitivity and in Equation (5), SP corresponds to the specificity. The geometric mean is calculated as in Equation (6).

IV. RESULTS & DISCUSSION

In this section the results of the model are presented as also the relative importance and frequency of each feature in the LOSO validation. In Figure 2 is illustrated the variation of the LOSO G_{mean} as the maximum number of features allowed in the final model increase. As can be seen, the performance reaches a G_{mean} plateau of 81,24% at 14 maximum features allowed, this indicates that however more features can be introduced, the regularization factor always clips at this number of maximum features, which can be confirmed by the analysis of the weights obtained in LOSO for each of these runs. However, is using the two most relevant features that the model reaches its maximum performance, a G_{mean} of 83.67%, while a valley of performance can be encountered in between these two conditions. Before discussing this behavior, it is pertinent to inspect the relative weight of each
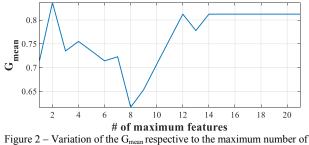

Figure 2 – Variation of the G_{mean} respective to the maximum number of features allowed in the model

TABLE I - The top 5 features with highest absolute weights. The column 'Mean' corresponds to the average weight of all folds, 'StD' its standard deviation and 'Relative Weight' correspond to the relative absolute weight respective to the total absolute weights of all features.

Feature	Mean	StD	Relative Weight
BR slope	-18,49	4,14	0,37
QRSm abs(slope)	6,42	2,3	0,13
BP value	-6,01	1,19	0,12
QRS slope	-4,7	1,43	0,09
BP abs(slope)	2,72	0,73	0,05

feature, or importance, in the plateau region. If the average weight for each feature is calculated across all folds, there are three features that stand out from the rest: the BP filtered value, the BR slope and finally, the QRSm absolute slope, as can be seen in TABLE I. Regarding the physiological validity of the features present in the TABLE I, the negative weight of the feature BR slope suggests that a negative trend on this parameter reflects the worsening of the cardiac status. In patients with HF, a change in respiratory rate can warn of impending pulmonary edema, or fluid in the lungs, which is a common symptom of HF. Additionally, variations in the QRSm duration and in the BP also reflect an increasing difficulty on the heart to satisfy the body needs. Moreover, by analyzing the features that are present in the model with two maximum features, the combination of two features is always the BR slope with the QRSm absolute slope or with the BP filtered value, except for two cases where some physiological parameters are missing, in these two cases the pair of features is the BP filtered value coupled with the BP absolute slope. Given this, one can question why the models with greater performance are the one with two maximum features and the one with no limit of number of features. This behavior is mainly due to the stability of the feature space encountered, i.e., the feature combinations obtained in LOSO are less prone to change in each fold, and in this way the influence of artifacts is reduced.

V. CONCLUSION

The methodology manages to obtain comparable results in respect with the literature and demonstrate the potential of such approach. The characteristic of the obtained features being interpretable and expected in the HF context, a G_{mean} of 83.67%, and finally the capability of capturing the trends onsets of the physiological parameters, makes this a suitable method in the usage of telemonitoring and decision support solutions. For future work is intended to validate the developed strategy in other case studies, such as the prediction of hypertensive crises, and to study how the methodology manages to extract interpretable and meaningful variables for new problems, maintaining the robustness and performance.

ACKNOWLEDGMENT

REFERENCES

[1] M. Metra and J. R. Teerlink, "Heart failure," Lancet, vol. 390, no. 10106, pp. 1981–1995, Oct. 2017.

[2] E. Braunwald, "Heart Failure," JACC Hear. Fail., vol. 1, no. 1, pp. 1–20, Feb. 2013.

[3] M. Gheorghiade, L. De Luca, G. C. Fonarow, G. Filippatos, M. Metra, and G. S. Francis, "Pathophysiologic targets in the early phase of acute heart failure syndromes.," Am. J. Cardiol., vol. 96, no. 6A, p. 11G–17G, Sep. 2005.

[4] E. E. Tripoliti, T. G. Papadopoulos, G. S. Karanasiou, K. K. Naka, and D. I. Fotiadis, "Heart Failure: Diagnosis, Severity Estimation and Prediction of Adverse Events Through Machine Learning Techniques," Comput. Struct. Biotechnol. J., vol. 15, pp. 26–47, Jan. 2017.

[5] B. W. Sahle, A. J. Owen, K. L. Chin, and C. M. Reid, "Risk Prediction Models for Incident Heart Failure: A Systematic Review of Methodology and Model Performance," J. Card. Fail., vol. 23, no. 9, pp. 680–687, Sep. 2017.

[6] A. Candelieri, D. Conforti, F. Perticone, A. Sciacqua, K. Kawecka-Jaszcz, and K. Styczkiewicz, "Early detection of decompensation conditions in heart failure patients by knowledge discovery: The HEARTFAID approaches," in 2008 Computers in Cardiology, 2008, vol. 35, pp. 893–896.

[7] A. Candelieri, D. Conforti, A. Sciacqua, and F. Perticone, "Knowledge Discovery Approaches for Early Detection of Decompensation Conditions in Heart Failure Patients," in 2009 Ninth International Conference on Intelligent Systems Design and Applications, 2009, pp. 357–362.

[8] A. Candelieri, "A Hyper-Solution Framework for SVM Classification: Application for Predicting Destabilizations in Chronic Heart Failure Patients," Open Med. Inform. J., vol. 4, no. 1, pp. 136–140, Jul. 2010.

[9] G. Guidi, M. C. Pettenati, P. Melillo, and E. Iadanza, "A Machine Learning System to Improve Heart Failure Patient Assistance," IEEE J. Biomed. Heal. Informatics, vol. 18, no. 6, pp. 1750–1756, Nov. 2014.

[10] G. Guidi, L. Pollonini, C. C. Dacso, and E. Iadanza, "A multi-layer monitoring system for clinical management of Congestive Heart Failure," BMC Med. Inform. Decis. Mak., vol. 15, no. S3, p. S5, Dec. 2015.

[11] J. Henriques, P. Carvalho, S. Paredes, T. Rocha, J. Habetha, M. Antunes, and J. Morais, "Prediction of Heart Failure Decompensation Events by Trend Analysis of Telemonitoring Data," IEEE J. Biomed. Heal. Informatics, vol. 19, no. 5, pp. 1757–1769, Sep. 2015.

[12] G. Koulaouzidis, D. K. Iakovidis, and A. L. Clark, "Telemonitoring predicts in advance heart failure admissions," Int. J. Cardiol., vol. 216, pp. 78–84, Aug. 2016.

[13] B. F. McBride and C. M. White, "Acute Decompensated Heart Failure: A Contemporary Approach to Pharmacotherapeutic Management," Pharmacotherapy, vol. 23, no. 8, pp. 997–1020, Aug. 2003.

[14] H. Khan, A. P. Kalogeropoulos, F. Zannad, C. N. Marti, P. W. F. Wilson, V. V Georgiopoulou, A. M. Kanaya, A. B. Newman, E. Schelbert, T. B. Harris, S. Kritchevsky, C. Yancy, M. Gheorghiade, G. C. Fonarow, and J. Butler, "Incident heart failure in relation to vascular disease: Insights from the Health, Aging, and Body Composition Study," Eur. J. Heart Fail., vol. 16, no. 5, pp. 526–534, May 2014.

[15] S. I. Chaudhry, Y. Wang, J. Concato, T. M. Gill, and H. M. Krumholz, "Patterns of Weight Change Preceding Hospitalization for Heart Failure," Circulation, vol. 116, no. 14, pp. 1549–1554, Oct. 2007.

[16] H. Suzuki, M. Nodera, M. Kamioka, T. Kaneshiro, Y. Kamiyama, and Y. Takeishi, "Intracardiac impedance after cardiac resynchronization therapy is a novel predictor for worsening of heart failure," Heart Vessels, vol. 32, no. 8, pp. 926–931, Aug. 2017.

[17] L. I. Kuncheva and J. J. Rodríguez, "On feature selection protocols for very low-sample-size data," Pattern Recognit., vol. 81, pp. 660–673, 2018.