Predicting Heart Failure patient events by exploiting saliva and breath biomarkers information

Evanthia E. Tripoliti, Georgia S. Karanasiou, Fanis G. Kalatzis, Dimitrios I. Fotiadis
Department of Biomedical Research, Institute of Molecular Biology and Biotechnology
Foundation for Research and Technology-Hellas, Ioannina, Greece
etripoliti@gmail.com, g.karanasiou@gmail.com, tkalatz@gmail.com, fotiadis@cc.uoi.gr

Silvia Ghimenti, Tommaso Lomonaco, Francesca Bellagambi, Roger Fuoco
Department of Chemistry and Industrial Chemistry, University of Pisa
Pisa, Italy
silvia.ghimenti@ifc.cnr.it, tlomonaco@ifc.cnr.it, fbellagambi@ifc.cnr.it, fuoco@deci.unipi.it

Katerina K. Naka, Aris Bechlioulis
2nd Department of Cardiology, University of Ioannina, Ioannina, Greece
drkknaka@gmail.com, md02798@yahoo.gr

Mario Marzilli, Maria Chiara Scali
Azienda Ospedaliera-Universitaria Pisana
Cardiothoracic and Vascular Department, Pisa, Italy mario.marzilli@med.unipi.it, chiara.scali@yahoo.it

Abdelhamid Errachid
Université de Lyon,
Institut de Sciences Analytiques, Villeurbanne France
abdelhamid.errachid@univ-lyon1.fr

Yorgos Goletsis
Department of Economics
University of Ioannina
Ioannina, Greece
goletsis@cc.uoi.gr

Abstract—The aim of this work is to present a machine learning based method for the prediction of adverse events (mortality and relapses) in patients with heart failure (HF) by exploiting, for the first time, measurements of breath and saliva biomarkers (Tumor Necrosis Factor Alpha, Cortisol and Acetone). Data from 27 patients are used in the study and the prediction of adverse events is achieved with high accuracy (77%) using the Rotation Forest algorithm. As in the near future, biomarkers can be measured at home, together with other physiological data, the accurate prediction of adverse events on the basis of home based measurements can revolutionize HF management.

Keywords-heart failure; event prediction; saliva biomarkers; breath biomarkers; data mining

I. INTRODUCTION

Heart failure (HF) is a chronic life-threatening condition characterized by high rates of mortality and re-hospitalizations. The European Society of Cardiology reports that 26 million people worldwide suffer from HF and 74% of them present at least one comorbidity [1]. HF is characterized by frequent re-admissions to hospital. HF accounts for 1-3% of all hospital admissions, while almost the 24% of hospitalized patients are re-hospitalized within a 30-day and the 46% within a 60-day, post discharge period. Across the world, the 2-17% of patients admitted to hospital with HF die while in hospital and the 17-25% die within one year of admission [2]. The cost of HF management is driven by hospitalizations, corresponding to 1-2% of total healthcare expenditure.

The ability to accurately predict the aforementioned undesirable events enables the effective risk stratification of patients and allows the clinical decision making. This valuable prognostic information can guide the clinical experts in the adaptation of patient management and in the selection of the best treatment plan that should be followed. In turn, this is expected to improve the quality of care provided to the patients, while in parallel result in better health outcomes. Towards this direction, different factors have been studied for their predictive ability in HF morbidity and mortality, destabilizations and re-hospitalizations. In addition, several studies have been conducted focusing on the simultaneous examination of multiple factors using statistical methods (e.g. multi-variable Cox regression models). Such studies resulted in the creation of acknowledged in the clinical practice scores: (i) for the estimation of risk for mortality, the Heart Failure Survival Score [3], the Get With the guidelines score [4], the Seattle Heart Failure Model [5], the EFFECT [6], (ii) for re-hospitalizations [7], and (iii) for morbidity [8].

The prediction of re-hospitalizations [9-15] and mortality [11], [16-23] has gained the interest of researchers who developed prediction models exploiting different categories
of information (e.g. sociodemographic, clinical examination, medical condition, lab tests, medication intake, phenotypic data, sensor data) along with machine learning techniques. A description of the methods reported in the literature and a comparison of them are presented in [24].

Recent research has identified certain biomarkers which strongly correlate with the HF severity, progression and mortality [25-42]. Moreover, progress in analytical chemistry and biosensor development allows some of them to be detected in saliva and breath [43-46]. Uric Acid, Tumor Necrosis Factor Alpha (TNF-a), a-Amylase, Lactate, Cortisol and 8-iso-prostaglandin F2a, are among the most important saliva biomarkers, while Acetone (2-Propanon) and 2-methy-1,3-butadiene (isoprene) are indicative examples of breath biomarkers that play a key role in the patient diagnosis and prognosis.

The goal of this study is to introduce such biomarkers in the adverse event prediction process. Obtaining saliva and breath biomarkers is non-invasive and in a future setting can be performed at home [47], becoming on this way a significant tool for HF patient management. In our study, we employ these breath and saliva biomarkers in a machine learning approach which combines heterogeneous patient data (i.e. sociodemographic, clinical, sensor data and biomarkers) for the prediction of adverse events.

II. MATERIALS AND METHODS

A. Dataset

The proposed method is evaluated using a dataset of 27 patients collected by the clinical center of the Universita Di Pisa (UNIPI), Italy within the framework of the HEARTEN project [47]. The criteria for patient selection are reported in Table I.

| Criterion | Description |
|-----------|-------------|
| I         | Diagnosed with HF (Framingham criteria) who have continuous symptoms with frequent recurrence |
| II        | Belonging to the functional NYHA I-IV class followed by an optimal treatment |
| III       | Recently hospitalized, (at least one in the last six months) |
| IV        | Undergone one electrocardiogram (in the last 12 months) and have HF symptoms |
| V         | Underage, with very severe HF, patients with obesity and advanced chronic kidney failure are not included |

The features recorded for each patient can be grouped to the following categories (Table II).

| Category    | Description                                                                 |
|-------------|-----------------------------------------------------------------------------|
| General Information | Age, gender, existence/relation with caregiver, education level, ethnicity etc. |
| Allergies   | Features expressing the presence or not of allergies                         |

Based on clinical studies on biomarker behavior and influence, performed during the HEARTEN project, the following biomarkers are selected as most prominent marker compounds for monitoring HF conditions: (i) acetone in breath, (ii) cortisol and TNF-a in saliva. More specifically, acetone mirrors metabolism, as well as metabolic stress and the concentrations of acetone are elevated in HF patients compared to healthy subjects. A very significant increase in salivary cortisol levels is observed in peculiar cases when a sudden worsening of patients happened during hospitalization. In HEARTEN studies, after therapy adjustment, the cortisol decreased by a factor of about 2. As a consequence, cortisol was considered a good candidate for monitoring the HF patients. Additionally, chronic HF
patients have shown high circulating levels of TNF-a, which correlate with the severity of their disease, since TNF-a levels show a linear relation with HF prognosis. Moreover, several studies have demonstrated that concentrations of TNF-a in saliva may reflect those in serum, making the TNF-a an ideal HF-related salivary biomarker.

In total, 263 features are recorded for each patient (3 features corresponding to biomarkers, 151 features extracted from sensor data and 109 features corresponding to the other categories).

These features are recorded from the first time of patient’s hospitalization until discharge, every second day. Thus, 141 instances are collected from all patients. It must be mentioned that the number of the hospitalizations days is not the same for all patients. On average, each stay lasts approximately 4 days. The dataset given as input to the proposed method is created with the assumption that the discharge instance of each patient is considered as event free (“no event” class), while the instance of the patient when the event was presented and/or the first hospitalization took place is considered as event (“event” class). This results to a set of 54 instances, with 25 corresponding to an event and 29 to a no event.

B. The proposed method

The proposed method consists of three steps: (i) preprocessing, (ii) feature selection, (iii) classification. A schematic representation of the proposed method is shown in Fig. 1 and a detailed description of each stage is provided below.

In the first step, missing values are addressed. Features with more than 60% of missing values are removed, since imputation of missing values cannot be performed due to the nature of the data. Furthermore, features where the distribution between the values is greater than 80% are not retained. In the second step, the identification of features that can act as discriminators between the two expected situations (presence of an event or not) are selected following a wrapper approach [53], in combination with the classifiers employed in step 3. Two different approaches are tested. The first one (Fig. 1i) takes as input all the features, while in the second (Fig. 1ii), the method is applied separately to features extracted from sensor data and to features corresponding to categories (i)-(vii). Finally, in the third step, nine classifiers are employed and tested [54]: (i) Random Forests (RF), (ii) Logistic Model Trees (LMT), (iii) J48, (iv) Rotation Forest (ROT), (v) SVM, (vi) Radial Basis Function Network (RBF Network), (vii) Bayesian Network (BN), (viii) Naïve Bayes (NB), (ix) Simple Classification and Regression Tree (Simple CART).

III. RESULTS

The proposed method is evaluated on a dataset of 54 instances, while the number of features is differentiated depending on the outcome of the feature selection step. The obtained results in terms of accuracy (Acc), sensitivity (Sens), specificity (Spec) and area under curve (AUC) are presented in Table III. Both approaches produce rather similar results (accuracy 76% and 77%, respectively), with 12 features to be finally selected in the first approach and 23 features in the second approach. Rotation Forests (ROT) seem to be the best performing classification algorithm (see Table IV).

In order to evaluate the contribution of biomarkers in the prediction of adverse events, the following experiments are made: (i) all available features are given as input, (ii) features only from sensors are employed, (iii) features only from biosensors are utilized as predictors, and (iv) features from sensors and biosensors are met. The results without feature selection are presented in Table V. The results

| Classifiers | Evaluation measures |
|-------------|----------------------|
|             | Acc | Sens | Spec | AUC |
| 1st approach of feature selection approach |
| NB          | 76% | 76%  | 76%  | 76% |
### Table IV. Event Prediction Results per Classifier Following the Second Approach

| Classifiers | Evaluation measures |
|-------------|----------------------|
|             | Acc  | Sens  | Spec  | AUC  |
| ROT         | 77%  | 77%   | 77%   | 74%  |

#### Table V. Event Prediction Results on Different Combinations of Features (Without Feature Selection).

| Case | Model | Acc |
|------|-------|-----|
| (i) All features without feature selection | ROT | 57% |
| (ii) Features only from sensors without feature selection | RF | 60% |
| (iii) Features only from biosensors without feature selection | CART | 66% |
| (iv) Features from sensors and biosensors without feature selection | ROT | 59% |

#### Table VI. Event Prediction Results on Different Combination of Features (Using Feature Selection).

| Case | Model | Acc |
|------|-------|-----|
| (i) All features with feature selection | NB | 76% |
| (ii) Features only from sensors with feature selection | RF | 71% |
| (iii) Features only from biosensors with feature selection | CART | 66% |
| (iv) Features from sensors and biosensors with feature selection | RF | 72% |
| The proposed method | ROT | 77% |

The proposed method presented in Table VI are extracted from the same experiments, as those presented in Table V, but this time following feature selection.

It should be noted that in case (iii) where only biomarkers are utilized, the feature selection approach is not applied due to the already small number of biomarkers used (i.e. three).

As shown in Tables V and VI, our approach yields superior results. The positive effect of feature selection is also clear (accuracy improvement from 66% to 77%).

A comparison of the proposed method with those reported in the literature (Table VII) cannot be directly performed since the studies reported in the literature: (i) predict the presence or not of one specific adverse event only (destabilizations, re-hospitalizations, mortality) and not the presence or not of HF adverse event in general like the proposed method, and (ii) do not utilize biomarkers.

Focusing on specific adverse events, of course it has advantages for clinical practice; still it requires a much larger dataset. In this sense, this can be considered as a limitation of the proposed method. This will be addressed in the future through the data that will be collected during the pilot phase of the HEARTEN project. The utilization of breath and saliva biomarkers is the innovative feature of the proposed method.

#### Table VII. Comparison with the Literature.

| Study* | Evaluation measures |
|--------|----------------------|
| Candelieri et al. 2008 | Acc | 92.03% |
| Candelieri et al. 2009 | Acc | 82.06% |
| Candelieri et al. 2010 | Acc | 87.35% |
| Guidi et al. 2014 | Acc | 87.60% |
| Guidi et al. 2015 | Acc | 71.90% |
| Zolfaghar et al. 2013 | Acc | 87.12% |
| Vedomske et al. 2013 | AUC | 84.00% |
| Koulaouzidis et al. 2016 | AUC | 82.00% |
| Kang et al. 2016 | AUC (c-statistic) | 59.00% |
| Tugerman et al. 2016 | AUC | 84.20% |
| Shah et al. 2015 | AUC | 71.80% |
| Fonarrow et al. 2005 | odds ratio for mortality | 12.9 |
| Bohacik et al. 2013 | Acc | 77.66% |
| Panahiazar et al. 2015 | 1 year AUC | 68.00% (baseline set) 81.00% (extended set) |
| 2 years AUC: | 70.00% (baseline set) 74.00% (extended set) |
| 5 years AUC: | 61.00% (baseline set) 73.00% (extended set) |
| Taslimitehrani et al. 2016 | 1 year Acc | 91.40% |
| 2 years Acc | 83.00% |
| 5 years Acc | 80.90% |
| Austin et al. 2012 | AUC | 79.00% |
| Bohacik et al. 2015 | Sens | 63.27% |
| Spec | 65.54% |
| Ramirez et al. 2015 | Sens | 18.00% |
| Spec | 79.00% |
| Subramanian et al. 2011 | AUC(c-statistic) | 84.00% |

* Studies references are presented in reference [24] of the current study.

### IV. Conclusions

An automated method for the prediction of adverse events related to HF, utilizing information from saliva and breath biomarkers, is presented. Different experiments are
conducted in order the proposed method to be evaluated and the contribution of biomarkers to the prediction problem to be estimated. The results confirm the prediction ability of biomarkers either if they are employed as the only input (Acc: 66%) or in combination with other categories of features (Acc: 77%). Among the biomarkers, TNF-α is the one that presents the largest correlation with the prediction of an adverse event. However, the small number of instances does not allow the extraction of “safe” conclusions. The collection of biomarkers measurements from a larger number of patients will lead to a more in-depth evaluation.

As in the near future biomarkers can be measured at home, together with biosensor data, the accurate prediction of adverse events on the basis of home based measurements will revolutionize HF patient management. Such an approach can become the core of a chronic care model, allowing for early action by both, patients and physicians.

ACKNOWLEDGMENT

This work is supported by the HEARTEN project that has received funding from the European Union’s Horizon 2020 research and innovation program under grant agreement No 643694.

REFERENCES

[1] C. Cook, G. Cole, P. Asaria, R. Jabbour, and D. P. Francis, “The annual global economic burden of heart failure,” *International Journal of Cardiology*, vol. 171, no. 3, pp. 368–376, 2014.
[2] P. Ponikowski *et al.*, “Heart failure: preventing disease and death worldwide,” *ESC Heart Failure*, vol. 1, no. 1, pp. 4–25, Sep. 2014.
[3] E. S. Ketchum and W. C. Levy, “Multivariate risk scores and patient outcomes in advanced heart failure,” *Congest Heart Fail*, vol. 17, no. 5, pp. 205–212, Oct. 2011.
[4] P. N. Peterson *et al.*, “A validated risk score for in-hospital mortality in patients with heart failure from the American Heart Association get with the guidelines program,” *Circ Cardiovasc Qual Outcomes*, vol. 3, no. 1, pp. 25–32, Jan. 2010.
[5] W. C. Levy *et al.*, “The Seattle Heart Failure Model: prediction of survival in heart failure,” *Circulation*, vol. 113, no. 11, pp. 1424–1433, Mar. 2006.
[6] D. S. Lee, P. C. Austin, J. L. Rouleau, P. P. Liu, D. Naimark, and J. V. Tu, “Predicting mortality among patients hospitalized for heart failure: derivation and validation of a clinical model,” *JAMA*, vol. 290, no. 19, pp. 2581–2587, Nov. 2003.
[7] E. F. Philbin and T. G. DiSalvo, “Prediction of hospital readmission for heart failure: development of a simple risk score based on administrative data,” *J. Am. Coll. Cardiol.*, vol. 33, no. 6, pp. 1560–1566, May 1999.
[8] S. J. Pocock *et al.*, “Predictors of mortality and morbidity in patients with chronic heart failure,” *Eur. Heart J.*, vol. 27, no. 1, pp. 65–75, Jan. 2006.
[9] K. Zolfaghari, N. Meadem, A. Teredesai, S. Basu Roy, C. Si-Chi, and B. Muckian, “Big data solutions for predicting risk-of-readmission for congestive heart failure patients,” presented at the IEEE International Conference on Big Data, 2013.
[10] M. A. Vedomske, D. E. Brown, and J. H. Harrison, “Random Forests on Ubiquitous Data for Heart Failure 30-Day Readmissions Prediction,” in *2013 12th International Conference on Machine Learning and Applications*, 2013, vol. 2, pp. 415–421.
[11] S. J. Shah *et al.*, “Phenomapping for novel classification of heart failure with preserved ejection fraction,” *Circulation*, vol. 131, no. 3, pp. 269–279, Jan. 2015.
