Estimation of New York Heart Association class in Heart Failure Patients Based on Machine Learning Techniques*

Evanthia E.Tripoliti, *Member, IEEE,* Theofilos G. Papadopoulos, Georgia S. Karanasiou, Fanis G. Kalatzis, Aris Bechlioulis, Yorgos Goletsis, *Member*, *IEEE*, Katerina K. Naka, Dimitrios I. Fotiadis, *Senior Member, IEEE*

***Abstract*—The aim of this work is to present an automated method for the early identification of New York Heart Association (NYHA) class change in patients with heart failure using classification techniques. The proposed method consists of three main steps: a) data processing, b) feature selection, and c) classification. The estimation of the severity of heart failure in terms of NYHA class is addressed as two, three and, for the first time, as four class classification problem. Eleven classifiers are employed and combined with resampling techniques. The proposed method is evaluated on a dataset of 378 patients, through a 10-fold-cross-validation approach. The highest detection accuracy is 97, 87 and 67% for the two, three and the four class classification problem, respectively.**

I. INTRODUCTION

Heart failure (HF) is described by the inability of the heart to fulfill the circulatory demands of the body due to progressively impairment of the ventricle to fill with or eject blood. HF leads to damage of the cardiovascular system and becomes one of the major causes of mortality and morbidity [1]. This in combination with the severe consequences, in terms of quality of life, recurrent hospitalizations and escalating healthcare costs that HF disease induces for the patients and the healthcare systems, intensify the need for effective and efficient management of HF that includes early detection of HF, recognition of HF subtype, estimation of HF severity and treatment.

In clinical practice, several criteria (e.g. Framingham, Boston, the Gothenburg and the European Society of Cardiology-ESC criteria) are utilized to determine the presence of HF [1]. Once HF is detected, the etiology or the subtypes of HF can be estimated based on the measurement of the left ventricular ejection fraction (LVEF). The experts classify the severity of HF using either the New York Heart Association (NYHA) or the American College of Cardiology/American Heart Association Guidelines (ACC/AHC) [1] classification systems that provide useful and complementary information. ACC/AHA stages of HF emphasize on the development and progression of HF, whereas NYHA focuses on the exercise capacity of the patient and the symptomatic status of the disease [1, 2].

Although the patho-physiology of HF has been understood in great extent by the medical community, the huge amount of data that should be analyzed and the complexity of them, transform the process of HF diagnosis and treatment selection to quite challenging and complicated tasks. Furthermore, the utilization of NYHA classification system, as a tool for HF severity estimation, introduces high intra-observer variability, since it is based on subjective evaluation [3]. The effort to overcome those issues led to the utilization of machine learning techniques that has the potential to analyze, predict and classify medical data accurately and efficiently [4].

Regarding HF severity estimation, which is the aim of the current work, Akinyokun *et al.* [5] proposed a neuro-fuzzy expert system. For each patient, six variables were utilized, out of seventeen variables that were recorded, expressing signs and symptoms of HF. Guidi *et al.* [6] classified a patient as mild HF, moderate HF and severe HF utilizing four different classifiers, Neural Network (NN), Support Vector Machines (SVM), Decision Tree (DT) and a Fuzzy-Genetic (FG) algorithm. The classifiers were trained and tested using anamnestic and instrumental data. Two years later, the same research team [7] tested Classification and Regression Tree (CART) and Random Forests (RF) classifiers. For the evaluation of the classifiers, a subject based cross validation approach was followed to address the fact that the dataset included correlated data (baseline and follow-up data of the same patient). In [8] a multi-layer monitoring system for clinical management of congestive HF (CHF) is presented where a decision support system was developed providing prediction of de-compensations and assessment of the HF severity based on the RF algorithm. A scoring model allowing classification of a subject to three groups, healthy group (without cardiac dysfunction), HF-prone group (asymptomatic stages of cardiac dysfunction) and HF group (symptomatic stages of cardiac dysfunction) was presented by Yang *et al.* [9]. The model was based on the SVM classifier. Pecchia *et al.* [10], Mellilo *et al.* [11] and Shahbazi *et al.* [12], exploited the discrimination power of

*Research 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.

E.E Tripoliti, G.S. Karanasiou, and F.G. Kalatzis are with the Department of Biomedical Research, Institute of Molecular Biology and Biotechnology, FORTH, GR 45110, Ioannina, Greece, (e-mail:etripoliti@gmail.com, g.karanasiou@gmail.com, tkalatz@gmail.com)

T.G. Papadopoulos is with the Unit of Medical Technology and Intelligent Information Systems, University of Ioannina, GR 45110, Ioannina, Greece (email: tpapado2011@gmail.com).

Y. Goletsis is with the Department of Economics, University of Ioannina, GR 45110, Ioannina, Greece (email: goletsis@cc.uoi.gr)

K.K. Naka and A. Bechlioulis are with the Michaelidion Cardiac Center, University of Ioannina, and 2nd Department of Cardiology, University of Ioannina, GR 45110, Ioannina, Greece (email: drkknaka@gmail.com, md02798@yahoo.gr).

D.I. Fotiadis is with the Department of Biomedical Research, Institute of Molecular Biology and Biotechnology, FORTH, Ioannina, Greece and the Dept. of Materials Science and Engineering, Unit of Medical Technology and Intelligent Information Systems, University of Ioannina, GR 45110, Ioannina, Greece (e-mail: fotiadis@cc.uoi.gr)

Preprint submitted to 2017 IEEE International Conference on Biomedical and Health Informatics. Received November 29, 2016.

long-term heart rate variability (HRV) measures that can be extracted by electrocardiogram (ECG) in order to estimate the severity of HF. In [10], the selected HRV measures were given as input to the CART algorithm, while Mellilo *et al.* [11] modified the CART algorithm by incorporating a feature selection algorithm to address the issues of small and unbalanced dataset. Furthermore, they compared the results of the modified CART algorithm with those produced by simple CART, C4.5 and RF classifiers which were evaluated with and without the application of SMOTE algorithm [13]. Shahbazi *et al.* [12] employed *k*-Nearest Neighbor (k-NN) classifier in order to classify the patient as low or high risk.

The aim of our work is to early detect the change of NYHA class of the HF patient and inform the experts in order to modify and adjust the management of the HF patient thus avoiding thus serious side effects. It is developed within the NYHA class detection module of the HEARTEN Knowledge Management System (KMS) [14]. HEARTEN KMS aims to effectively assess patient status, assess and enhance patient adherence, support patient (self-) management by (i) exploiting real patient data coming from multiple sources (sensors, bio-sensors, clinical data, personal data), (ii) performing computational analysis using data mining techniques, (iii) providing useful information (including suggestions and alerts) tailored to the needs of each actor regarding patient status, patients' HF severity, risk for adverse events, as well as adherence in terms of medication, nutrition and physical activity.

The current work addresses the estimation of HF severity as a two, three, and for the first time, as four class classification problem allowing the experts to adjust the treatment of patient and minimize the risk of adverse events. In the studies reported in the literature and presented above, classifiers separate low risk from high risk patients or classify the patient as mild moderate and severe HF by merging NYHA classes without clarifying which NYHA classes correspond to each group. In this work, all the cases are examined by aggregating different categories of features and employing different classifiers. In case of the two class problem the following cases are examined: *a1)* NYHA I & II *vs.* NYHA III & IV, *a2)* NYHA I *vs.* NYHA II & III & IV, *a3)* NYHA I & II *vs.* NYHA III, *a4)* NYHA II *vs.* NYHA III & IV, *a5)* NYHA II & III *vs.* NYHA IV. In case of the three class problem the following cases are examined: *b1)* NYHA I *vs.* NYHA II *vs.* NYHA III & IV, *b2)* NYHA I *vs.* NYHA II & III *vs.* NYHA IV, *b3)* NYHA II *vs.* NYHA III *vs.* NYHA IV. In case of the four class problem each NYHA class corresponds to a different group (*c1:* NYHA I *vs.* NYHA II *vs.* NYHA III *vs.* NYHA IV).

II. MATERIALS AND METHODS

A. Dataset

The proposed method is evaluated using a dataset of 378 patients retrospectively collected by three different clinical centers, Universita Di Pisa (UNIPI) Italy, Servicio Andaluz de Salud (SAS) Spain and the 2nd Department of Cardiology, University Hospital of Ioannina (UHI). The datasets consists of patients: (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) who have been recently hospitalized, (at least one in the last six months), (iv) who have undergone one electrocardiogram (in the last 12 months) and have HF symptoms. Patients who are underage, with very severe HF, with obesity and advanced chronic kidney failure are not included.

The dataset includes: 18 patients in NYHA class I, 125 patients in NYHA class II, 173 and 62 patients in NYHA class III and IV, respectively. The features recorded for each patient can be grouped to the following seven categories: (i) General Information, (ii) Allergies, (iii) Medical Condition, (iv) Drugs, (v) Biological data related with the HF disease, (vi) Clinical Examinations, vii) Adherence. Totally, 102 features are recorded for each patient that according to the literature and experts knowledge are correlated with the severity of the HF.

B. The proposed method

The proposed method consists of three steps. The first step focuses on the pre-processing of data in order issues such as missing values induced by different clinical practices in each center and consequently different type of collected retrospective clinical data of the three clinical centers to be addressed. The second step aims to identify features that have the potential to discriminate NYHA classes. The third step estimates the severity of HF and informs the experts if the patient has changed NYHA class. A detailed description of the three steps is provided below.

Step 1 - Data pre-processing: The input of the first step is 378 instances each one consisting of 102 features plus the NYHA class label. Seven NYHA class labels were assigned by medical experts to the patients defined as NYHA I, NYHA I-II, NYHA II, NYHA II-III, NYHA III, NYHA III-IV, NYHA IV. The labels NYHA I-II, NYHA II-III, NYHA III-IV are considered as intermediate severity to the corresponding classes. Nevertheless, these intermediate classes were strongly unbalanced in the dataset since only 3 instances are labeled as NYHA I-II, 6 as NYHA II-III and 3 as NYHA III-IV. For this reason each one of those instances is merged with the straight following class of higher risk. From the 102 features, 26 features were removed due to high missing value rate (more than 60%). Imputation of missing values cannot be performed due to the nature of data. Thus, after the first step, a dataset consisting of 378 instances and 76 features (class label is not included) is retained.

Step 2 - Feature selection: Feature selection can be performed either utilizing a filter or a wrapper approach. Both approaches have been tested in the second step of the proposed method. In the case of the filter approach, Info Gain, Gain Ratio, Symmetrical Uncertainty (SU), Relief-F, One-R and Chi-squared feature selection measures are employed, while wrapper and the Correlation-based Feature Selection (CFS) algorithm [15, 16], in combination with the classifiers employed in step 3, are utilized.

Step 3 - Classification: Eleven classifiers are tested [15]: (i) RF, (ii) Random Tree (RT), (iii) Logistic Model Trees (LMT), (iv) J48, (v) Rotation Forest, (vi) SVM, (vii) Radial Basis Function Network (RBF Network), (viii) Bayesian Network (Bayesnet), (ix) NaiveBayes, (x) Multiple Layer Perceptron (MLP), (xi) Simple CART.

III. Results

The proposed method is evaluated on a dataset of 378 instances each one consisting of 76 features plus the class label. The proposed method is repeated nine times, each time addressing either a two (cases *a1* to *a5*), a three (cases *b1* to *b3*) or a four class classification problem (case *c1*). The models (classifiers and feature selection approach) that provide the best results are presented in Table I. The results of the proposed method in terms of accuracy (ACC), positive predictive value (PPV), sensitivity (SENS), specificity (SPEC), area under curve (AUC) and F-measure (FM) are reported in Table II. For the evaluation of the classifiers, 10-fold stratified cross-validation procedure is applied. SMOTE is applied to the training set during the 10-fold cross-validation procedure to address the imbalanced class problem. It is applied only to the training set and not to the test set avoiding thus the incorporation of the instances which are used for evaluation to the creation the synthetic instances in SMOTE. In case *a3* the resampling of the dataset is not performed since the two classes (first class includes instances of NYHA class I and II and the second class includes instances of NYHA class III) are balanced. More specifically, the first class includes 143 instances and the second class includes 173 instances. The results of the step by step evaluation of the proposed method is presented in Table III.

TABLE I. Models provided the best results (Highest accuracy in bold)

Case	Filter	Wrapper	CFS
a1	Rotation Forest and ReliefF	MLP	Simple CART
	79%	75%	73%
a2	Rotation Forest and SU	RBFNetwork	Rotation Forest
	97%	92%	96%
a3	Rotation Forest and ReliefF	RF	MLP
	77%	**78%**	72%
a4	Rotation Forest and OneR	RT/Simple CART	SVM
	78%	77%	75%
a5	Rotation Forest and Gain Ratio	SVM	SVM
	91%	89%	85%
b1	Rotation Forest and Gain Ratio	Rotation Forest	Rotation Forest
	76%	**79%**	71%
b2	Rotation Forest and Info Gain	SVM	SVM
	87%	84%	81%
b3	Rotation Forest and ReliefF	LMT	Rotation Forest / SVM
	69%	**69%**	65%
c1	Rotation Forest and SU	Rotation Forest	SVM
	67%	**68%**	60%

TABLE II. Evaluation measures of the proposed method using filter approach

Case	PPV	FM	SENS	SPEC	ACC	AUC
a1	79%	79%	79%	76%	79%	82%
a2	97%	97%	97%	58%	97%	88%
a3	77%	77%	77%	76%	77%	82%
a4	77%	77%	78%	70%	78%	81%
a5	91%	89%	91%	57%	91%	83%
b1	75%	74%	76%	70%	76%	81%
b2	87%	85%	87%	55%	87%	78%
b3	69%	69%	69%	78%	69%	79%
c1	67%	66%	68%	80%	67%	79%

IV. Discussion

An automated method for the estimation of HF severity is presented. HF severity estimation is addressed as two, three and four classification problem. A dataset of 378 instances, each one consisting of 102 features is given as input in the proposed method. During the first step, 76 features are retained. The feature which contribute to the discrimination of the NYHA classes are selected in the second step of the proposed method. Filter, as well as wrapper approaches are employed. The selected features are differentiated depending on the nature of the classification problem (two class, three class and four class) and the case *(a1-a5, b1-b2, c1*) that it is examined. The best results are obtained using the Rotation Forests classifier in combination with feature selection techniques based on the filter approach.

The proposed method is differentiated from other methods reported in the literature since the patients are not classified in mild, moderate and severe, by merging NYHA classes to these three groups, but each NYHA class is treated as a separate group. This provides valuable clinical information to the experts in terms of the suggestions that they should provide for the personalized management of the patients.

A comparison of the proposed method with those reported in the literature is presented in Table IV. The comparison can be achieved only for case *a1* since the definition of classes utilized in [11] and [12] is the same with the definition followed in the current work. The results indicate that the proposed method presents similar and in most of the cases better results than those reported in the literature. The classification accuracy is decreased in cases where the instances of NYHA class II and III are not merged into one class (e.g. cases *a1* and *a3*), expressing thus a difficulty of the proposed method to discriminate those two classes. This can be attributed to the fact that in the current work HRV features are not included in the dataset, on the contrary to the studies described in [11] and [12]. Furthermore, features indicative for the estimation of the severity of HF, like NT-proBNP, have been removed due to the high rate of missing values. The effect of the missing values is confirmed when the proposed method is applied separately to each one of the datasets provided by the three different clinical centers (corresponding to different features with missing values). Additionally, the dataset will be enhanced with features expressing breath and saliva HF biomarkers.

V. Conclusions

Severity estimation of HF, in terms of four NYHA classes, through the utilization of machine learning techniques is presented. Automated patient classification (in term of NYHA class) will overcome the problem of subjective evaluation of patient condition and of self-reporting bias. The early identification of such changes can be used for treatment adjustment and can become a valuable tool in an automated system for HF patients (self-) management, reducing the high rates of hospitalizations and decrease significantly the corresponding healthcare costs.

TABLE III. RESULTS OF THE STEP BY STEP EVALUATION OF THE PROPOSED METHOD

Case	Before feature selection	After feature selection
a1	NYHA I & II *vs.* NYHA III & IV	
	75%	79%
a2	NYHA I *vs.* NYHA II & III & IV	
	94%	97%
a3	NYHA I & II *vs.* NYHA III	
	73%	77%
a4	NYHA II *vs.* NYHA III & IV	
	74%	78%
a5	NYHA II & III *vs.* NYHA IV	
	89%	91%
b1	NYHA I *vs.* NYHA II *vs.* NYHA III & IV	
	72%	76%
b2	NYHA I *vs.* NYHA II & III *vs.* NYHA IV	
	84%	87%
b3	NYHA II *vs.* NYHA III *vs.* NYHA IV	
	67%	69%
c1	NYHA I *vs.* NYHA II *vs.* NYHA III *vs.* NYHA IV	
	61%	67%

TABLE IV. COMPARISON WITH THE LITERATURE

Classification problem	Literature		Proposed work	
Two class	Melilo *et al.* [11] Low risk (NYHA I or II) *vs.* High risk (NYHA III or IV)	85%	*a1*	79%
	Shahbazi et al. [12] Low risk (NYHA I or II) *vs.* High risk (NYHA III or IV)	97%		
	Pecchia *et al.* [10] Mild HF *vs.* Severe HF	79%	*a2*	97%
			a3	77%
			a4	78%
			a5	91%
Three class	Guidi *et al.* [6] Mild HF *vs.* Moderate HF *vs.* Severe HF	86%	*b1*	76%
	Guidi *et al.* [7] Mild HF *vs.* Moderate HF *vs.* Severe HF	83%		
	Guidi *et al.* [8] Mild HF *vs.* Moderate HF *vs.* Severe HF	81%	*b2*	87%
	Yang *et al.* [9] Healthy (NYHA I, ACC/AHA A) *vs.* HF-prone group (NYHA I, ACC/AHA B-C) *vs.* HF group (NYHA II-III, ACC/AHA C-D)	74%	*b3*	69%
Four class	n/a		*c1*	67%

REFERENCES

[1] P. Ponikowski, A.A. Voors, S.D. Anker, H. Bueno, J.G.F. Cleland, A.J.S. Coats, *et al.,* "2016 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure", *European Heart Journal* 2015:ehw128.

[2] K. Dickstein, A. Cohen-Solal, G. Filippatos, J.J.V. McMurray, P. Ponikowski, P.A. Poole-Wilson, *et al.,* "ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure 2008 The Task Force for the Diagnosis and Treatment of Acute and Chronic Heart Failure 2008 of the European Society of Cardiology. Developed in collaboration with the Heart Failure Association of the ESC (HFA) and endorsed by the European Society of Intensive Care Medicine (ESICM)", *European Heart Journal*, vol. 29, pp: 2388–442, 2008.

[3] A. Alonso-Betanzos, V. Bolón-Canedo, G.R. Heyndrickx, P.L. Kerkhof, "Exploring Guidelines for Classification of Major Heart Failure Subtypes by Using Machine Learning", *Clin Med Insights Cardiol,* vol. 9, pp:57–71, 2015.

[4] E.E. Tripoliti, T.G. Papadopoulos, G.S. Karanasiou, K.K. Naka,D.I. Fotiadis, "Heart Failure: Diagnosis, Severity Estimation And Prediction Of Adverse Events Through Machine Learning Techniques", *Computational and Structural Biotechnology Journal* (under review).

[5] C.O. Akinyokun, O.U. Obot, F-ME Uzoka, "Application of Neuro-Fuzzy Technology in Medical Diagnosis: Case Study of Heart Failure", in *Proc. World Congress on Medical Physics and Biomedical Engineering*, Munich, Germany, Springer Berlin Heidelberg; pp: 301–4, 2009.

[6] G. Guidi, E. Iadanza, M.C. Pettenati, M. Milli, F. Pavone, G.B. Gentili, "Heart Failure Artificial Intelligence-Based Computer Aided Diagnosis Telecare System", In: Donnelly M, Paggetti C, Nugent C, Mokhtari M, editors. Impact Analysis of Solutions for Chronic Disease Prevention and Management, Springer Berlin Heidelberg; pp: 278–81, 2012.

[7] G. Guidi, M.C. Pettenati, P. Melillo, E. Iadanza, "A machine learning system to improve heart failure patient assistance", *IEEE J Biomed Health Inform*, vol.18, pp:1750–6, 2014.

[8] G. Guidi, L. Pollonini, C.C. Dacso, E. Iadanza, "A multi-layer monitoring system for clinical management of Congestive Heart Failure", *BMC Med Inform DecisMak*, vol.15, Suppl 3:S5, 2015.

[9] G. Yang, Y. Ren, Q. Pan, G. Ning, S. Gong, G. Cai, *et al.*, "A heart failure diagnosis model based on support vector machine", in *Proc. 3rd International Conference on Biomedical Engineering and Informatics* (BMEI), China, vol. 3, pp: 1105–8, 2010.

[10] L. Pecchia, P. Melillo, M. Bracale, "Remote health monitoring of heart failure with data mining via CART method on HRV features", *IEEE Trans Biomed Eng*, vol. 58, pp:800–4, 2011.

[11] P. Melillo, N. De Luca, M. Bracale, L. Pecchia, "Classification tree for risk assessment in patients suffering from congestive heart failure via long-term heart rate variability", *IEEE J Biomed Health Inform*, vol. 17, pp:727–33, 2013.

[12] F. Shahbazi, B. Mohammadzadeh Asl, "Generalized discriminant analysis for congestive heart failure risk assessment based on long-term heart rate variability", *Computer Methods and Programs in Biomedicine*, vol. 122, pp: 191-198, 2015.

[13] N. Chawla, K. Bowyer, L. O. Hall, and P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique", *Journal of Artificial Intelligence Research*, vol. 16, pp: 321-357, 2002.

[14] G.S. Karanasiou, F.G. Kalatzis, E.E. Tripoliti, A. Errachid, M.G. Trivella, R. Fuoco, F. Di Francesco, A. Martínez-García, C. Luís Parra Calderón, J.K. Schubert, W. Miekisch, J. R. Bausells, "A preliminary presentation of a mobile co-operative platform for Heart Failure self-management", in *Proc. 15th IEEE International Conference on Bioinformatics & Bioengineering (BIBE 2015)*. Serbia, 2015.

[15] G.S. Karanasiou, E.E. Tripoliti, T.G. Papadopoulos, F.G. Kalatzis, Y. Goletsis, K.K. Naka, A. Bechlioulis, A. Errachid and D. I. Fotiadis, "Predicting adherence of patients with heart failure through machine learning techniques", *Healthc Technol Lett.*, vol. 3(3), pp:165-170, 2016.

[16] M.A. Hall (1998). Correlation-based Feature Subset Selection for Machine Learning. Ph.D. dissertation, Department of Computer Science, University of Waikato, Hamilton, New Zealand, 1999.

Preprint submitted to 2017 IEEE International Conference on Biomedical and Health Informatics. Received November 29, 2016.