

Prediction of Hospital Readmission in Heart Failure Patients

Bertone Elisa, Vasta Elisa, Veneruso Chiara

1. INTRODUCTION

Heart failure (HF) is a heterogeneous clinical syndrome stemming from cardiac overload and injury that leads to considerable morbidity and mortality. At the same time, elevated costs and use of resources are related to HF management, which impacts on the healthcare system greatly [1]. The annual medical cost for HF care amounts to \$24,383 per patient in the US, more than half destined for hospitalizations driving costs [2]. In Italy, HF is the second leading cause of hospital admission, and it is estimated that 2.4% of the national healthcare budget is allocated for this pathology. Considering that 88% of hospitalized people for heart failure are over 65 years old and that Italy is the second country for number of elderlies, predicting and reducing HF patient admission rates is a national priority [3].

Moreover, unsatisfying patient outcomes may lead to hospital readmission and heavier financial burden to the healthcare system. It was estimated that preventable readmissions can lead to almost \$17 billion annual cost reduction in the US [4]. Since hospital readmittance rate for heart failure is at roughly 20%, unplanned patients' readmission and the subsequent waste of resources lead to a negative financial impact on hospitals, directly affecting their overall performance [5]. Thus, it is extremely important to anticipate this situation before it really happens: if HF patients with high readmission risk could be prioritized, health-care providers could lower their risk of coming back to the hospital.

One strategy to deal with this problem is predicting patients' readmission by applying machine learning (ML) models. Literature analysis revealed various works moving in this direction. Awan et al. tested several machine learning techniques and focused their attention on the importance of using an appropriate metric in the evaluation process. The multi-layer perceptron they proposed as final model reached an AUROC = 0.62, which then revealed to be an overestimation of the true performances of the model since they were dealing with an imbalanced dataset. The area under the precision-recall curve (AUPRC), instead, appeared to be a more appropriate measure, and it was found to be equal to 0.46 [6]. Golas et al. experimented both ML and DL approaches, reaching respectively AUROC values of 0.65 with a gradient boosting model and 0.70 with a deep unified network ad hoc designed in the perspective of overfitting avoidance [7].

The aim of this work is to predict the hospital readmission of heart failure patients 6 months after the discharge. This would allow not only to prepare the resources for future hospitalizations, but it could also give insights about the importance of some clinical variables in order to ensure their recording at the time of the first hospitalization.

2. MATERIALS AND METHODS

2.1 Dataset

The public dataset [16] consists of electronic healthcare records of patients that were admitted at Zigong Fourth People's Hospital (Sichuan, China) from December 2016 to June 2019 with a diagnosis of HF. Information on hospital readmissions were extracted at mandatory follow-up visits at 28 days, 3 months, and 6 months. The dataset included all types of heart failure (acute, chronic, left and/or right), that was defined according to the European Society of Cardiology (ESC) criteria.

The patients admitted for HF in the studied period were 2008 and a total of 164 different parameters are available. They can be divided into wide categories: demographic data, baseline clinical characteristics, comorbidities, laboratory findings, and outcomes. The demographic data included information on age, sex, height, body weight, admission ward, type of admission, occupation, discharge department, visit times. Instead, clinical characteristics involve information on the patient status on the first day of admission, whereas comorbidities give insights on other diseases the patient may have. Laboratory findings include all the physiological parameters tested since the first day of admission; the outcome features contain information on discharge day, death within 28 days, readmission at 28 days, 3 months or 6 months and other similar variables.

The provided variables were grouped into two subsets: numerical and categorical features (furtherly separated between those characterized by an intrinsic order and those which are not). The list of features falling in each set is specified in Appendix A.

The inspection of target's values proportion revealed a significative unbalance: 1235 subjects (0.63%) fell in class 0 (not readmitted), whereas the remaining 773 ones (0.37%) in class 1 (re-hospitalized patients).

2.1.1 Data Preprocessing

Some analyses were performed on the provided dataset to make it suitable to predict the readmission at 6 months. All the outcome variables containing information after the first hospital discharge, specified in Appendix B, were removed, and the variable "*re.admission.within.6.months*" was used as target for the classification task. Further analyses included:

- checking the absence of possible duplicates, looking at the *inpatient number* which univocally identifies each subject;
- looking for the presence of patients that have died within 6 months and removing them. Indeed, including another class of dead patients in the classification problem was not reasonable since they were only the 3.09% of the population under study (62 over 2008). Considering

the improvement of hospital resources' management as the final aim of predicting possible readmissions, dead patients would not affect hospital's finances, so they can be deleted without negative implications. Moreover, deleting them minimizes the risk of introducing bias in the model due to the inclusion of those critical cases. After this deletion, the overall number of patients was reduced to 1946.

2.1.2 Cross Validation

K-fold cross validation was employed in order to evaluate model performance and choose which was the best one for fulfilling the assigned task. Using this method allows to reduce the luck component in the train-test split, which otherwise could severely influence models' performance assessment. In fact, at each iteration, k-1 folds are merged to constitute the training set and the remaining one is used as test set, assuring independence between the sets.

K was chosen equal to 4, searching for a tradeoff between a too high k's value, leading to big computational effort to train k different models, and a too low one, bringing to an insufficient cardinality of the training set resulting from the single split. In this case, 75% of patients (1459) were used each time as training set and 25% as test set (487). In addition, k-folds creation was performed by assuring stratification on the target class.

2.2 Exploratory Data Analysis

An analysis of features distribution was carried out to gain insights, evaluate the necessity of scaling, outlier removal and applying transformation techniques.

All the following observations were conducted separately on numerical and categorical features on the training set resulting from each iteration of k-fold cross validation. The conclusions drawn from this phase were based on the exploratory data analysis conducted on each fold.

Numerical Features From the histograms (Fig. 1), features were found to have different orders of magnitude, as expected. Scaling operation was thus carried out immediately after handling abnormalities, as will be described in section 2.4.3.

Categorical Features To perform a deeper exploration of the extracted features, their distributions were separately plotted for the different target's classes with the aim of visually investigating if each single variable was, alone, able to significantly differentiate whether patients will be readmitted within 6 months.

An example of this representation is shown in Fig. 2.

2.3 Dealing with abnormalities

2.3.1 Missing Values

Since the number of patients was limited, the choice of how to handle missing data was carried by trying to preserve as much as possible the cardinality of the dataset. Before taking any kind of decision, a deep exploration on the predictors was carried out in literature to consider which ones could

be relevant for the intended task. With regard to numerical variables, features with a percentage of NaN values higher than 40% were dropped in each fold, since some of them contained more than a half of missing values. The value of this threshold was chosen after manually inspecting that those features did not seem discriminative for the target and did not fall in the set of good predictors known from literature for most of the folds.

Afterwards, several methods for imputing missing values on the remaining variables were investigated. First, single imputation methods using mean and median values are very fast, but imputing with constant values may cause a significant reduction in dataset variance if the number of missing values is too high. Moreover, if values of certain variables are not missing completely at random (MCAR), imputing with a single value is dangerous because it will introduce bias in the model since their missingness can be related to the target.

Later, more sophisticated iterative multivariate imputers were experimented to overcome those limitations. In this regard, MICE algorithm was applied: a Bayesian Ridge regressor was used as estimator for the iterative imputations. At each iteration, the correlation between the variables with missing values to impute and each of the remaining numerical features is computed, and this value is used by the regressor. The algorithm was tested with different numbers of features to consider (10, 12, 15). To ensure physiological plausibility, it was necessary to constrain the imputed values to be within the minimum and the maximum values assumed in the training set by each feature. Simple and iterative imputation methods were compared by testing a Random Forest classifier on the first fold with a further cross validation procedure. Random forest classifier was chosen here to have a sight on imputation methods' efficacy because this point of the pipeline came before scaling the features, thus it was necessary to choose a ML algorithm which was not too sensitive to features' scale. The relative results of this inspection are quantitatively compared in section 3.

Categorical attributes contained a significantly reduced number of missing values and simple imputation with the most frequent class (mode) among all the training data was applied.

2.3.2 Outliers

Outlier analysis was carried out in parallel with missing data management since applying MICE algorithm in presence of these abnormalities could have brought to a further generation of those vicious observations. The presence of outliers in some numerical features was already visible in the boxplots. However, being in a biomedical context, it is fundamental to distinguish between statistical outliers (e.g. wrong entries), that need to be removed, and pathological outliers reflecting an illness, which instead need to be considered. Therefore, it was not possible to automatically spot and remove them using traditional techniques (e.g. setting a threshold value after the computation of the z-score).

The choice was to filter observations according to their z-score for visualization purpose and then set feature-specific

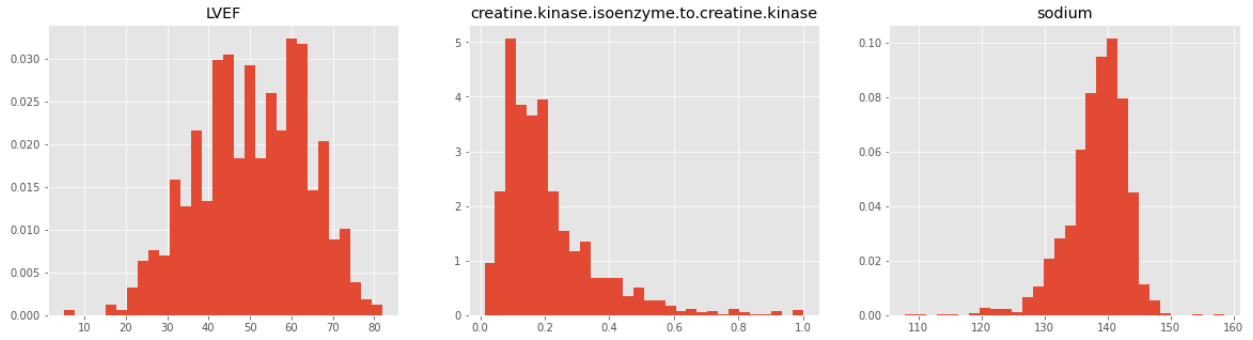


Fig. 1. Histograms of left ventricular ejection fraction, creatine kinase isoenzyme to creatine kinase and sodium features.

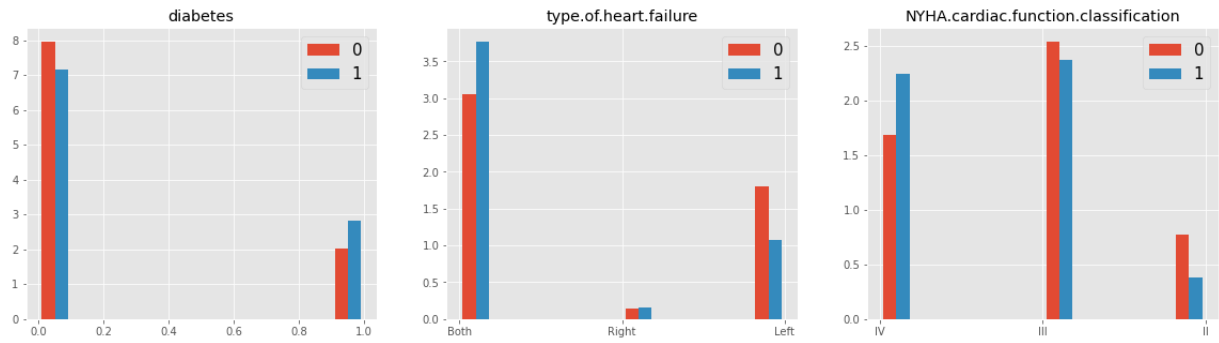


Fig. 2. Histograms of diabetes, type of heart failure and NYHA cardiac function classification features.

thresholds on outliers which were not reflecting a pathological condition. Outlier values were found in weight, height, BMI, systolic and diastolic blood pressure and mean arterial blood pressure. These observations were then treated as missing values and thus imputed as described in section 2.3.1.

2.4 Feature Engineering

2.4.1 Encoding of Categorical Features

Different encoding methods were applied to categorical features based on their type in order to feed the classifier:

1. One-Hot Encoding to variables like gender or type of heart failure.
2. Ordinal Encoding to ordinal categorical variables that have an intrinsic order, like age category, expressed in age range, and Killip grade, originally expressed as a roman number.

The other categorical variables were already in binary form with values of 1 and 0 (e.g. presence of comorbidities) or sorted in a scale of values, like the CCI score, so they did not need to be encoded.

2.4.2 Feature Augmentation

The data analysis phase revealed that 17 numerical features presented the same amount of missing values (51.028%). Under the assumption that those values were obtained after doing a specific exam and that said exam was only performed on patients who seemed to have a more severe condition, a new categorical feature was created. This new variable, called 'exam', is 'True' for patients who received the exam and 'False' for those who did not, meaning the ones who presented a missing value in the aforementioned features.

2.4.3 Feature Scaling

The observation made in section 2.2 about numerical features having different orders of magnitude led to the decision of rescaling these training data using a standardization technique, hence bringing them to follow a normal distribution with 0 mean and unitary standard deviation. This standardization procedure was applied also to categorical variables which were characterized by an intrinsic order. The same statistical parameters computed on the train set were then used to rescale the corresponding test set data.

2.5 Dealing with class imbalance

To avoid having the model biased toward predicting the most frequent class, different possibilities were investigated. Beside classical oversampling, consisting in the repetition of some samples belonging to the minority class to reach the cardinality of the majority one, two other methods were experimented.

The first is the Synthetic Minority Oversampling Technique (SMOTE), which is a data augmentation technique that creates synthetic instances belonging to the minority class. It is the result of a convex combination (i.e. a linear combination in which all coefficients are non-negative and sum up to 1) between an observation x , randomly selected from the class to augment, and another observation y , picked at random as well from the subset of k -nearest neighbors of x .

The other investigated technique preserves the original dataset and consists in introducing class weight during models' training phase. These were computed as the inverse of the frequency of each class label in the training set, in such a way to penalize more the errors that the model performed

when predicting the minority class.

Class weights were found to be the most suitable solution for the designed model.

2.6 Feature Selection

With the aim of removing irrelevant and redundant information to obtain higher efficiency, a dimensionality reduction through feature selection was performed. In this regard, a combination of several methods was investigated.

The first selection of categorical features was performed based on visual inspection, removing features which did not seem to be discriminative for the target, as explained in section 2.2. Then, numerical and ordinal categorical features underwent two filter methods (variance and correlation-based), and later a wrapper selection method (Sequential Backward Selection) was applied to all the remaining categorical and numerical variables.

It is worthy to underline that these methods were applied separately, looking only at the training set of each fold.

Variance based selection The first explored filter method is based on variance analysis, with the idea of removing from the dataset features which were constant. Therefore, features exhibiting insignificant variability (variance lower than 0.01) were removed.

Correlation based selection The second filter method is based on linear correlation analysis, and it aims at removing redundant information. The following criterion was employed: if the linear correlation coefficient is higher than a certain threshold (set equal to 0.85 by visual inspection), the two variables are not independent and therefore it is possible to predict one from the other.

Consequently, it is reasonable to remove the one, among the two, which has a higher average linear correlation coefficient with all other features, keeping instead the one with a higher degree of novelty in terms of information content.

Sequential Backward selection To select a subset of truly relevant features, a backward stepwise selection method was implemented. Starting from the categorical and numerical variables which were kept after the previous steps, this wrapper method iteratively removes one feature based on the cross-validation score of a chosen estimator. Different models were tested as estimators, like KNN, Random Forest, Extratree and, in the end, a DecisionTree was chosen.

The stopping condition for this iterative procedure was selected as the final number of features to keep, chosen by trial and error to be equal to 18.

An alternative option could have been to set a tolerance and stopping the iterative procedure when the improvement in cross-validation score was below that threshold.

2.7 Model Selection

The training set of each k-fold was further subdivided to perform model selection: a train-validation split was carried out with the aim of building a validation set, which was not used for training the model, but to perform an hyperparameter

tuning. This was reached through a grid search procedure with 5-folds cross validation (i.e. at each cycle of cross validation 1167 samples were used for training and 292 for validation) using the f1 score. Thus, it automatically returned the best estimator as the model characterized by the hyperparameters combination that allowed the highest averaged score on the 5 iterations of cross validation.

Several Machine Learning models were tested, after a proper investigation of which were the commonly used algorithms in literature to predict hospital readmission:

- Decision Tree
- K-nearest neighbors (KNN)
- Support Vector Machine (SVM)
- Logistic Regression
- Random Forest
- Extra Tree Classifier
- Adaboost with Decision Trees

2.8 Performance metrics

For each fold, the normalized confusion matrix was used since it allows to gain insights on models' behavior, also taking into account the main clinical objective in this task: avoid underestimating the risk, which means reducing as much as possible the number of readmitted patients (target class 1) predicted as non-readmitted (target class 0). For automatic optimization procedures (the backward feature selection, the comparison of imputation methods and the gridsearch for hyperparameter tuning) the synthetic f1 score metric was chosen since it is less sensitive than accuracy to class unbalance and it resumes recall and precision metrics.

Two different curves were taken into account to evaluate the performances:

- Precision-recall curve and its relative area under the curve (AUPRC) for models developed when preserving the original target class unbalance (i.e. those in which class weights were adopted during training phase).
- Receiving operating curve (ROC) and its relative area under the curve (AUROC) in cases in which SMOTE was adopted to balance the training set.

The overall performance metrics were later computed as the mean of f1 scores and AUPRC obtained for each fold, whereas the normalized confusion matrix was obtained by concatenating the results of the 4 models of relative k-fold splits in terms of true and predicted labels.

3. RESULTS

Figure 3 shows the relative results obtained by the comparison of different imputation methods for the numerical features. Simple imputer with mean strategy and Iterative imputation with 10 features reached the best results in terms of f1 score: in the end, the latter was selected because of its capability to not reduce dataset's variance. Therefore, the final parameter combination for iterative imputer resulting from repeated manual trials is: number of iterations = 10, tolerance of the stopping condition = 10e-3, initial imputation strategy =

TABLE I
Machine learning models with selected hyperparameters and mean performance metrics

Model	Hyperparameters	F1-score train	F1-score	Precision	Recall	AUPRC
Decision Tree	'criterion': 'gini' 'max_depth': 4 'min_samples_leaf': 10 'min_samples_split': 20	0.51	0.49	0.48	0.53	0.46
Logistic Regression	'C': 0.5 'max_iter': 1000	0.53	0.53	0.48	0.56	0.51
KNN	'n_neighbors': 10	0.36	0.36	0.52	0.47	0.48
SVM	'C': 0.5 'shrinking': False 'kernel': 'linear'	0.53	0.53	0.48	0.57	0.51
Random Forest	'criterion': 'entropy' 'max_depth': 4 'min_samples_leaf': 10 'min_samples_split': 4 'n_estimators': 100	0.52	0.50	0.49	0.56	0.50
Extra Tree	'criterion': 'gini' 'max_depth': 10 'min_samples_leaf': 4 'min_samples_split': 7 'n_estimators': 50	0.53	0.50	0.50	0.57	0.53
Adaboost	'learning_rate': 0.1 'n_estimators': 1000	0.45	0.45	0.61	0.48	0.53

mean, imputation order = *ascending* (starting from the feature with fewest missing values), number of features = 10.

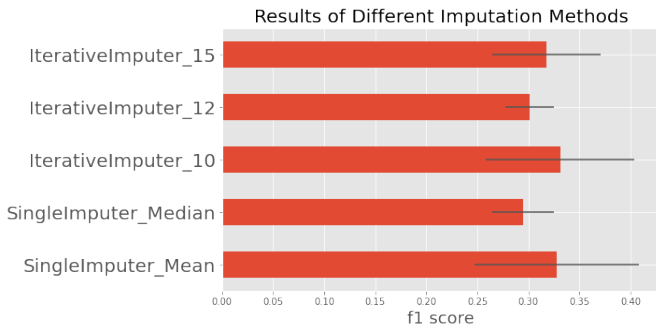


Fig. 3. Comparison among different imputation methods.

Applying the previously described feature selection methods on the training set of each k-fold cross validation led to different subsets of features, while keeping the same cardinality. The variables which were selected in at least one, two, three or all the folds are shown in appendix A.

The results of the tested classifiers, along with the best hyperparameters combination obtained from the grid search procedure, are shown in Table I.

Among the tested models, the Logistic Regression and Support Vector Machine classifiers were the ones which guaranteed the best tradeoff between all these aspects, as can be seen in Fig. 4 and Fig. 5.

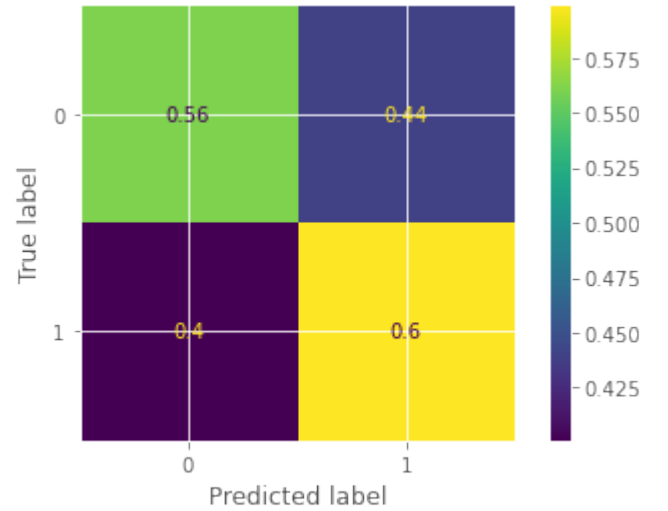


Fig. 4. Confusion matrix of Logistic Regression Classifier.

4. DISCUSSION

Particular attention was put on dimensionality reduction phase, not only in the perspective of improving models' performances and efficacy, but also taking into consideration the clinical implications. In fact, designing a model with a reduced number of predictors implies a lower effort's requirement to clinical people for data collection, as well as a lower probability of having missing data.

With regards to models' prediction performance, Support Vector Machine and Logistic Regression obtained the best

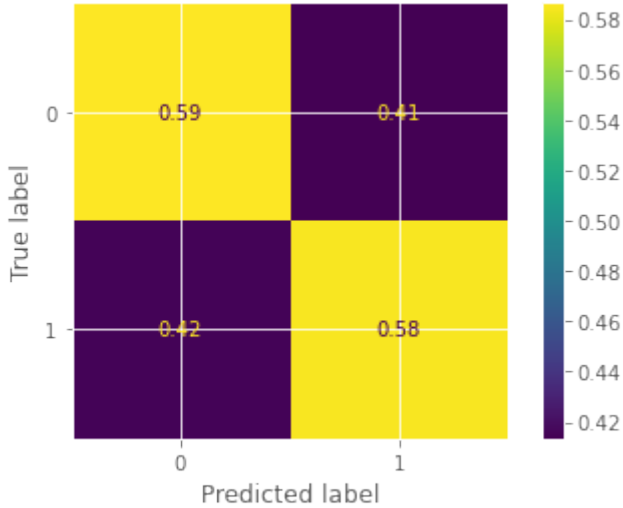


Fig. 5. Confusion matrix of Support Vector Machine classifier.

results in terms of f1 score (0.53), with no overfitting since they got the same value also in the train phase. Indeed, for Logistic Regression, the use of l2 penalty and a proper tuning of the C hyperparameter allowed to avoid overfitting.

Another important metric to consider when choosing the best classifier is the recall, since it is fundamental from a clinical point of view to avoid underestimating the patient's risk and preparing resources for hospitalization by considering the worst-case scenario. Thus, false negative samples must be avoided, and recall must be maximized. SVM and ExtraTree gained 0.57 in recall, whereas Random Forest and Logistic Regression a slightly lower value (0.56). At the same time, considering the AUPRC parameter, the best models were Adaboost and ExtraTree classifiers (0.53), but also Logistic Regression and SVM obtained comparable results (0.51).

It is quite difficult to compare these results to other studies in literature, since most of them performed the prediction on hospital readmission only within 30 days, 3 months or 1 year. At the same time, no studies were carried out to predict readmission on the same dataset of this work, thus different population under study may lead to different prediction performance. For instance, Junar Arciete Landicho et al. tested different classifiers to predict HF patients' readmission within 1 year and they obtained 0.428 of f1 and 0.30 of recall using SVM, testing it on data coming from the Northern Mindanao Medical Center (Philippines) [14]. In another study, a multi-layer perceptron (MLP) was used to predict 30 days readmission for Western Australian HF patients aged above 65 years. It obtained lower results than the models tested in this work in terms of AUPRC (0.46 compared to 0.53) but higher recall (0.70 vs 0.57) [6]. In [15], Hamner and Ellison predicted the rehospitalization within 6 months of patients suffering from heart failure previously admitted to a south-eastern medical center in Alabama (US). Logistic Regression obtained 0.61 of accuracy, but this result is not comparable to our model's performance, since this metric is not suitable for an unbalanced dataset.

Logistic Regression is one of the best models in

most of the considered metrics, but also in terms of interpretability. Especially when working in the medical field, it is fundamental to provide insights to clinicians or doctors to guide their decision: therefore, black box models may be inadequate even with high prediction performance. Conversely, models like logistic regression are considered highly interpretable because it is possible to automatically retrieve the coefficients, representing the importance they assigned to different features for performing predictions, and which were resulting from the log loss minimization. Indeed, it was possible to investigate the relevance of each feature by computing the weights assigned by the model, as shown in Table II. Particularly, it was found that that GCS, sodium, uric acid, exam, discharge day and LVEDD were significant in the prediction among other features.

Referring to the features which were automatically identified as the most important ones in at least half of the iterative folds, several were already known in literature to be significant predictors for hospital readmission. For example, the length of the hospital stay (*dischargeDay*) increased the risk of readmission in most studies [8]. Readmitted patients were found to have lower levels of glomerular filtration rate [9], which in fact was also included by Zamfirescu et al. as one of six independent determinants for heart failure patients' readmission [10]. It was also demonstrated that low knowledge about dietary sodium markedly increased the risk of rehospitalization [11]. Finally, uric acid has been observed to be a good discriminator in [12], even if the p value was thought to be not significant due to small sample size.

Left ventricular end-diastolic diameter (LVEDD) was known to be a significant predictor for risk stratification in sudden cardiac death but, at authors' knowledge, it has never been proved to be relevant for predicting HF patients rehospitalization [13].

The great contribution of this work is thus to sustain the hypothesis of LVEDD for the latter task: this predictor was among the few ones that resulted from the automatic selection procedure of three among four iterations. Moreover, the computation of the weights assigned by the logistic regression model revealed a great importance of this variable. This finding has a great relevance considering the implication that it may have into the real clinical practice: on the original dataset, before the splitting between training and test sets, a percentage of missing values for the LVEDD of 34.32% was observed. This result, instead, suggests to sensitize physicians to prescribe echocardiography exam and clinical people to keep particular attention on this value's recording.

5. CONCLUSIONS

In conclusion, different methods for features selection and imputation of missing values were evaluated, as well as various statistical models to predict hospital readmission within 6 months for patients suffering from heart failure. The results of this study have clinical implications, since LVEDD, discharge day and GCS were some of the most relevant parameters that contribute more to the prediction, suggesting they should be taken into account by clinicians.

TABLE II
Weights assigned by Logistic Regression to features in the first iteration of K-fold cross validation.

Feature	Weight
left.ventricular.end.diastolic.diameter.LV	0.219
uric.acid	0.006
glomerular.filtration.rate	-0.192
coefficient.of.variation.of.red.blood.cell.distribution.width	-0.042
standard.deviation.of.red.blood.cell.distribution.width	0.122
mean.hemoglobin.concentration	-0.056
basophil.count	0.057
eosinophil.count	0.113
D.dimer	-0.228
prothrombin.activity	-0.106
fibrinogen	-0.087
calcium	0.148
sodium	-0.878
creatinine.kinase	-0.005
direct.bilirubin	-0.05
cholesterol	-0.051
dischargeDay	0.198
GCS	-0.551

Even if the results are comparable to literature, the evaluated models have relatively weak ability to predict readmission, suggesting that other nonpatient factors may have a larger role in readmission risk. Readmission risk may be more dependent on improved hospital and post discharge care than mortality risk, so on health care system characteristics. Therefore, future work can include creating a ML classifier to train on a larger dataset, both in terms of number of patients and variables to analyze.

REFERENCES

- [1] Gori M, Marini M, De Maria R, Gonzini L, Gorini M, Cassaniti L, Benvenuto M, Municinò A, Navazio A, Ammirati E, Leonardi G, Pagnoni N, Montagna L, Catalano M, Midi P, Floresta AM, Pulignano G, Maggioni AP, Tavazzi L, Iacoviello M; Italian Network on Heart Failure (IN-HF) Investigators. Age-related changes in clinical characteristics and outcomes of chronic heart failure outpatients in a cardiology setting. A report from the Italian Network on Heart Failure. *Int J Cardiol.* 2022 Jan 1;346:36-44. doi: 10.1016/j.ijcard.2021.11.014. Epub 2021 Nov 16. PMID: 34793855.
- [2] Urbich, M., Globe, G., Pantiri, K. et al. A Systematic Review of Medical Costs Associated with Heart Failure in the USA (2014–2020). *PharmacoEconomics* 38, 1219–1236 (2020). <https://doi.org/10.1007/s40273-020-00952-0>
- [3] Heart Failure Policy Network. 2020. Heart failure policy and practice in Europe: Italy. London: HFPN
- [4] S. Wang, M. E. Elkin and X. Zhu, "Imbalanced Learning for Hospital Readmission Prediction using National Readmission Database," 2020 IEEE International Conference on Knowledge Graph (ICKG), 2020, pp. 116-122, doi: 10.1109/ICKG50248.2020.00026.
- [5] Pishgar, M., Theis, J., Del Rios, M. et al. Prediction of unplanned 30-day readmission for ICU patients with heart failure. *BMC Med Inform Decis Mak* 22, 117 (2022). <https://doi.org/10.1186/s12911-022-01857-y>
- [6] Awan, Saqib Ejaz, et al. "Machine learning-based prediction of heart failure readmission or death: implications of choosing the right model and the right metrics." *ESC heart failure* 6.2 (2019): 428-435
- [7] Golas, S.B., Shibahara, T., Agboola, S. et al. A machine learning model to predict the risk of 30-day readmissions in patients with heart failure: a retrospective analysis of electronic medical records data. *BMC Med Inform Decis Mak* 18, 44 (2018). <https://doi.org/10.1186/s12911-018-0620-z>
- [8] O'Connor M, Murtaugh CM, Shah S, Barrón-Vaya Y, Bowles KH, Peng TR, Zhu CW, Feldman PH. Patient Characteristics Predicting Readmission Among Individuals Hospitalized for Heart Failure. *Med Care Res Rev.* 2016 Feb;73(1):3-40. doi: 10.1177/1077558715595156. Epub 2015 Jul 14. PMID: 26180045; PMCID: PMC4712072.
- [9] Tung YC, Chou SH, Liu KL, Hsieh IC, Wu LS, Lin CP, Wen MS, Chu PH. Worse Prognosis in Heart Failure Patients with 30-Day Readmission. *Acta Cardiol Sin.* 2016 Nov;32(6):698-707. doi: 10.6515/acs20151113a. PMID: 27899857; PMCID: PMC5126448.
- [10] Zamfirescu MB, Ghilencea LN, Popescu MR, Bejan GC, Ghiordanescu IM, Popescu AC, Myerson SG, Dorobanțu M. A Practical Risk Score for Prediction of Early Readmission after a First Episode of Acute Heart Failure with Preserved Ejection Fraction. *Diagnostics (Basel).* 2021 Jan 29;11(2):198. doi:

- 10.3390/diagnostics11020198. PMID: 33572844; PMCID: PMC7911435.
- [11] Usha K. Kollipara, Osman Jaffer, Aamir Amin, Kathleen H. Toto, Lauren L. Nelson, Ruth Schneider, David Markham, Mark H. Drazner, Relation of Lack of Knowledge About Dietary Sodium to Hospital Readmission in Patients With Heart Failure, *The American Journal of Cardiology*, Volume 102, Issue 9, 2008, Pages 1212-1215, ISSN 0002-9149, <https://doi.org/10.1016/j.amjcard.2008.06.047>.
- [12] Urooj fatima, Fahad Ali, et al., Influence of hemoglobin and uric acid level on Re admissions for Heart Failure, *Journal Article*, 2014, *Circulation: Cardiovascular Quality and Outcomes*, A354-A354, 7, 10.1161/circoutcomes.7.suppl_1.354.
- [13] Narayanan K, Reinier K, Teodorescu C, Uy-Evanado A, Aleong R, Chugh H, Nichols GA, Gunson K, London B, Jui J, Chugh SS. Left ventricular diameter and risk stratification for sudden cardiac death. *J Am Heart Assoc*. 2014 Sep 16;3(5):e001193. doi: 10.1161/JAHA.114.001193. PMID: 25227407; PMCID: PMC4323796.
- [14] Junar Arciete Landicho, Vatcharaporn Esichaikul Roy Magdugo Sasil (2021) Comparison of predictive models for hospital readmission of heart failure patients with cost-sensitive approach, *International Journal of Healthcare Management*, 14:4, 1536-1541, DOI: 10.1080/20479700.2020.1797334
- [15] Hamner, J. B., Ellison, K. J. (2005). Predictors of hospital readmission after discharge in patients with congestive heart failure. *Heart Lung*, 34, 231-239.
- [16] Zhang, Z., Cao, L., Chen, R. et al. Electronic healthcare records and external outcome data for hospitalized patients with heart failure. *Sci Data* 8, 46 (2021). <https://doi.org/10.1038/s41597-021-00835-9>
- 'platelet.distribution.width', 'platelet.hematocrit', 'neutrophil.ratio**', 'neutrophil.count*', 'D.dimer**', 'international.normalized.ratio', 'activated.partial.thromboplastin.time*', 'thrombin.time', 'prothrombin.activity*', 'prothrombin.time.ratio', 'fibrinogen*', 'high.sensitivity.troponin**', 'myoglobin', 'carbon.dioxide.binding.capacity*', 'calcium***', 'potassium', 'chloride*', 'sodium**', 'Inorganic.Phosphorus', 'serum.magnesium', 'creatine.kinase.isoenzyme.to.creatine.kinase', 'hydroxybutyrate.dehydrogenase.to.lactate.dehydrogenase*', 'hydroxybutyrate.dehydrogenase', 'glutamic.oxaloacetic.transaminase', 'creatine.kinase*', 'creatine.kinase.isoenzyme', 'lactate.dehydrogenase', 'brain.natriuretic.peptide*', 'high.sensitivity.protein', 'nucleotidase', 'fucosidase', 'albumin*', 'white.globulin.ratio**', 'cholinesterase', 'glutamyltranspeptidase', 'glutamic.pyruvic.transaminase**', 'glutamic.oxaliplatin', 'indirect.bilirubin', 'alkaline.phosphatase', 'globulin', 'direct.bilirubin***', 'total.bilirubin', 'total.bile.acid', 'total.protein*', 'erythrocyte.sedimentation.rate', 'cholesterol*', 'low.density.lipoprotein.cholesterol*', 'triglyceride**', 'high.density.lipoprotein.cholesterol', 'homocysteine', 'apolipoprotein.A', 'apolipoprotein.B', 'lipoprotein', 'pH', 'standard.residual.base', 'standard.bicarbonate', 'partial.pressure.of.carbon.dioxide', 'total.carbon.dioxide', 'methemoglobin', 'hematocrit.blood.gas', 'reduced.hemoglobin', 'potassium.ion', 'chloride.ion', 'sodium.ion', 'glucose.blood.gas', 'lactate', 'measured.residual.base', 'measured.bicarbonate', 'carboxyhemoglobin', 'body.temperature.blood.gas', 'oxygen.saturation', 'partial.oxygen.pressure', 'oxyhemoglobin', 'anion.gap', 'free.calcium', 'total.hemoglobin', 'dischargeDay***'.

Categorical Features

'DestinationDischarge', 'admission.ward', 'admission.way', 'occupation', 'discharge.department', 'gender*', 'type.of.heart.failure', 'type.II.respiratory.failure', 'consciousness', 'respiratory.support', 'oxygen.inhalation', 'outcome.during.hospitalization', 'exam**', 'myocardial.infarction', 'congestive.heart.failure', 'peripheral.vascular.disease', 'cerebrovascular.disease', 'dementia', 'Chronic.obstructive.pulmonary.disease', 'connective.tissue.disease', 'peptic.ulcer.disease', 'diabetes', 'moderate.to.severe.chronic.kidney.disease', 'hemiplegia', 'leukemia', 'malignant.lymphoma', 'solid.tumor', 'liver.disease', 'AIDS', 'acute.renal.failure'.

Ordinal Categorical Features

'NYHA.cardiac.function.classification', 'Killip.grade', 'ageCat', 'visit.times*', 'CCI.score', 'eye.opening', 'verbal.response', 'movement', 'GCS***'

APPENDIX B

Features containing information located in time after the first hospitalization:

're.admission.within.3.months', 're.admission.within.28.days', 'time.of.death..days.from.admission', 're.admission.time..days.from.admission', 'return.to.emergency.department.within.6.months', 'time.to.emergency.department.within.6.months',

APPENDIX A

The list of features falling in each set is listed hereafter. The features which were selected by at least one iteration of the k-fold cross validation are indicated with an '*'; in the same way, features selected by 2, 3 or 4 iterations are indicated, respectively, with 2, 3, or 4 '*'.

Numerical Features

'body.temperature', 'pulse', 'respiration', 'systolic.blood.pressure', 'diastolic.blood.pressure*', 'map', 'weight', 'height*', 'BMI', 'fio2', 'LVEF', 'left.ventricular.end.diastolic.diameter.LV***', 'mitral.valve.EMS', 'mitral.valve.AMS', 'EA', 'tricuspid.valve.return.velocity', 'tricuspid.valve.return.pressure', 'creatinine.enzymatic.method**', 'urea**', 'uric.acid***', 'glomerular.filtration.rate**', 'cystatin*', 'white.blood.cell', 'monocyte.ratio', 'monocyte.count*', 'red.blood.cell', 'coefficient.of.variation.of.red.blood.cell.distribution.width*', 'standard.deviation.of.red.blood.cell.distribution.width*', 'mean.corpuscular.volume**', 'hematocrit', 'lymphocyte.count**', 'mean.hemoglobin.volume', 'mean.hemoglobin.concentration*', 'mean.platelet.volume*', 'basophil.ratio*', 'basophil.count*', 'eosinophil.ratio*', 'eosinophil.count*', 'hemoglobin', 'platelet*',

'death.within.28.days','death.within.3.months',
'death.within.6.months'.