2023 IEEE International Conference on Big Data and Smart Computing (BigComp) | 978-1-6654-7578-5/23/$31.00 ©2023 IEEE | DOI: 10.1109/BigComp57234.2023.00032

One year mortality prediction in heart failure using feature selection and missing value imputation in deep learning

Taehyun Kim
School of Computer Science and Engineering
Pusan National University
Busan, Republic of Korea
rlaxogus505@pusan.ac.kr

Minwook Kim
School of Computer Science and Engineering
Pusan National University
Busan, Republic of Korea
kmiiiaa@pusan.ac.kr

Hye Won Lee
Department of Cardiology
Pusan National University Hospital
Busan, Republic of Korea
a38321401@gmail.com

Giltae Song
School of Computer Science and Engineering, Artificial Intelligence Convergence Research Center
Pusan National University
Busan, Republic of Korea
gsong@pusan.ac.kr

***Abstract*— Heart failure is a disease caused by a deterioration in the function of the heart and a failure to supply the blood properly needed for the body. Follow-up measure with drugs and hospitalization can affect the survival of heart failure patients. Currently, none of the heart failure survival models including theses variables are effective yet. In this paper, we propose a method to effectively predict deaths within a year in patients with heart failure in Korea through preprocessing and deep learning. We used Korea Acute Myocardial Infarction Registry dataset which considers various features of patients with a left ventricular ejection rate 40% or less. Feature importance was measured using four models to find key features related to patients' survival. We conducted several data preprocessing such as missing value imputation. Our machine learning approach showed higher accuracy than existing methods for predicting one year mortality of patients in heart failure.**

Keywords—heart failure, death, KAMIR, missing value, feature selection, machine learning, deep learning

I. Introduction

Heart failure is a disease caused by decrease in the heart's relaxation or contraction function due to abnormalities in the structure or function of the heart, which does not properly supply the blood needed for the body. The normal range of blood ejected from the heart is between 50% and 75%, but heart failure refers to a disease with an ejection rate of less than 50%. Heart failure is classified into two types, depending on the degree of ejection rate. 41-49% is classified as output boundary heart failure (HFpEF) and 40% or less as output reduction heart failure (HFrEF). This study was conducted on HFrEF patients because HFpEF patients are close to the normal range and are difficult to specify as heart failure patients.

The death can be caused by other variables such as drugs taken in the past and follow-up like hospitalization. These external features that cause the death of patients in heart failure are not quite well known. In this paper, we propose a method to predict deaths of patients in heart failure within one year since their heart failure diagnosis. We use the Korea Acute Myocardial Infection Registry (KAMIR) dataset, which includes a total of 614 features, including multiple patient history, follow-up, and sexual diseases. We apply several data pre-processing such as removal of irrelevant features and missing value imputation for the KAMIR dataset. After these pre-processing steps, 231 features for about 2,000 patients are used to predict the one-year mortality with machine learning techniques including TabNet.

II. Related Research

There are some recent major studies of developing machine learning approaches for predicting heart failure patients' mortality [1,2]. [1] used the heart-failure-clinical-records-dataset derived in the UCI machine learning repository with only 13 features for 299 heart failure patients. They selected top two features according to feature importance such as ejection fraction and serum creatinine, and applied several simple classifiers including tree-based methods, artificial neural network, and support vector machines. They measured the model performance using test dataset including only 60 people. This may be too few to measure the actual model performance.

[2] used an old version of the KAMIR data obtained from South Korean patients. They predicted the death of AMI patients using simple classifiers including decision trees, ensembles, logistic regressions, and deepnets. The old version of the KAMIR data consisted of 303 features with 31,149 Korean heart failure patients. [2] selected 95 features for building a prediction model. It may be too slow to train the model.

III. Methods

In this study, we used a dataset of 15,628 patients derived from a new version of the KAMIR dataset. This dataset contained a total of 614 features.

978-1-6654-7578-5/23/$31.00 ©2023 IEEE
DOI 10.1109/BigComp57234.2023.00032

A. Preprocessing

297 features among 614 features were removed by an expert in hospitals, which were regarded as irrelevant features. We also deleted 27 additional features such as death dates, causes of death, and discharge dates, which are related to the predicted target class (i.e. whether to survive or die within a year since the diagnosis of heart failure). Other 38 features associated with the date and time of diagnosis of each disease were also removed because these features in time series are difficult to handle in our tabular data. Other 20 features related to drug capacity were also dropped since they were hard to maintain in categorical types. They were numerical values combined with letters in upper and lower case as well as several special symbols. There were some features of which all values were "Unknown" or "No". They were also removed.

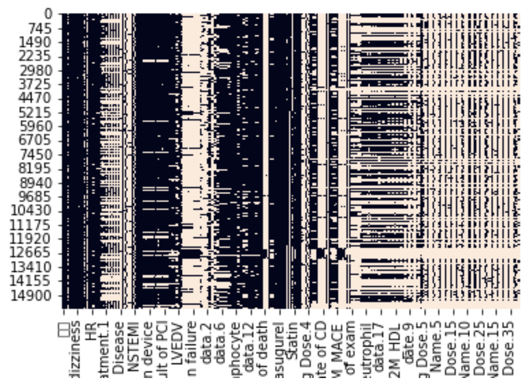


Fig. 1. The distribution of missing values in KAMIR dataset. Yellow elements indicate the missing values.

There are a lot of missing values in the KAMIR data as shown in Figure 1. All missing values of categorical features were treated as "Unknown" for categorical features. For missing values for numerical features, we applied three imputation methods. One method is to treat all missing numeric type features as zero. Another is to fill the missing values with the average of each feature. The other method is to use MissForest [3] to fill the missing values based on Random Forest (RF) [4]. When MissForest were applied, the target value were deleted so that the estimation of the missing values was not affected by the target value.

After this missing value imputation, we selected data of patients with ejection fraction 40% or less, who met the purpose of this study. The final dataset after these preprocessing steps contained 231 features of 1,893 patients: 271 patients were labeled as "dead" and 1,622 survived within a year using the first and second missing imputation methods, and of 2,033 patients: 371 patients were labeled as "dead" and 1,662 "survived" using the third missing value imputation method. We divided this dataset into 64% of the training set, 16% of the validation set, and 20% of the test set.

B. Training

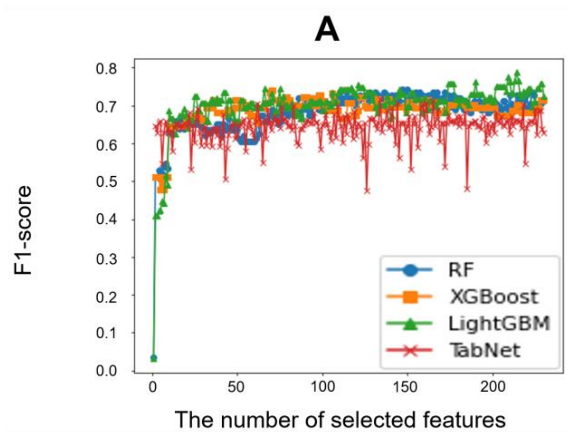
We applied four machine learning approaches: Random Forest (RF), XGBoost [5], LightGBM [6], and TabNet [7] models for predicting one-year mortality of the heart failure patients. We used 5-fold cross validation. TabNet is a deep learning technique specialized for tabular data.

We measured feature importance and optimal feature rankings for all four models. We tuned optimal hyperparameters using grid search for RF, XGBoost, and LightGBM and using optuna library [8] for TabNet.

C. Feature Selection

We measured the importance of features as importance ratios and sorted out them. We illustrate performance changes (F1-score) according to the number of top important features to be selected for building a model (see Figure 2). Figure 2 (a) shows the performance influence for the number of top features when the missing values are processed to be zero. The performance tends to converge after the number of selected features reach a certain number of top features. For example, Until the number of top 147 features is selected in the RF model, the performance keeps increasing. Once it reaches to 147 features, the F1-score of the model converges. In this case, we regarded top 147 features as an optimal number for the RF model. The optimal number of features were determined to be 70 features in XGBoost, 214 features in LightGBM, and 134 features in TabNet.

Figure 2 (b) shows the performance changes for the number of top features when the missing values are assigned to the mean. In this case, 211 features were selected as the optimal for the RF, 224 features for XGBoost, 157 features for LightGBM, and 71 features for TabNet. Figure 2 (c) shows the performance impact for the number of selected features when the missing values are preprocessed using the MissForest library. In this case, 216 features were optimal for the RF, 107 features for XGBoost, 225 features for LightGBM, and 159 features for TabNet.



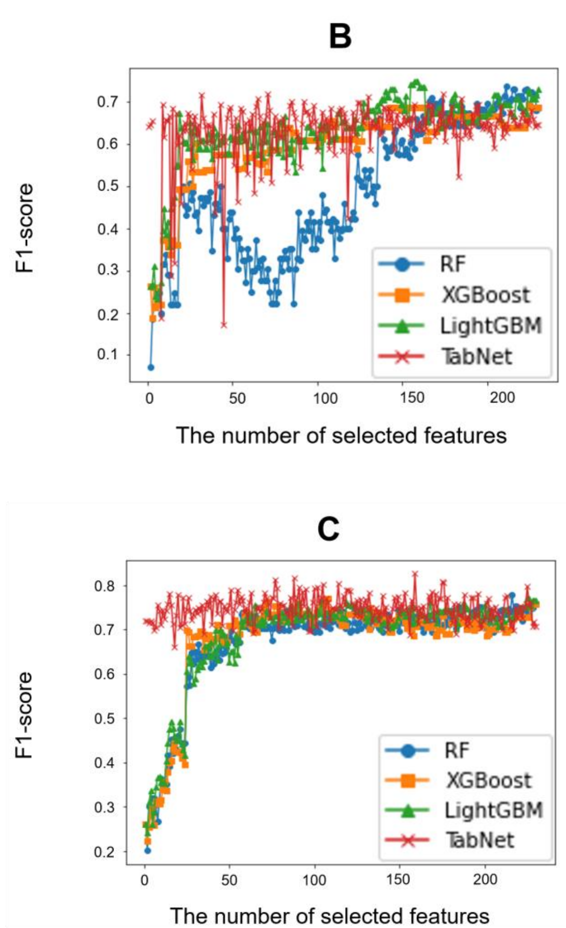


Fig. 2. F1-score changes according to the number of top features selected using four machine learning approaches: RF, XGBoost, LightGBM, and TabNet (a) when missing values were preprocessed to zero, when missing values were preprocessed to the mean, and (c) when missing values were handled using MissForest.

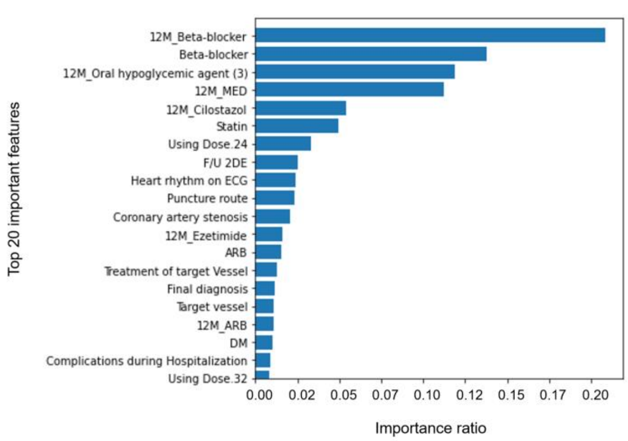


Fig. 3. Top 20 features of the TabNet model and their importance ratios.

IV. RESULT

We measured the performance of four machine learning models: RF, XGBoost, LightGBM, and TabNet with three missing value imputation methods in terms of Recall, Precision, and F1-score.

TABLE I. PERFORMANCE FOR FOUR MACHINE LEARNING MODELS WITH THREE MISSING VALUE IMPUTATION METHOS

Model	Missing Value Imputation	Performance		
		Recall	**Precision**	**F1-score**
RF	0	0.611	**0.942**	0.741
RF	Mean	0.648	0.853	0.736
RF	MissForest	0.783	0.773	0.778
XGBoost	0	0.685	0.804	0.740
XGBoost	Mean	0.611	0.804	0.694
XGBoost	MissForest	0.743	0.797	0.769
LightGBM	0	0.685	0.925	0.787
LightGBM	Mean	0.685	0.804	0.747
LightGBM	MissForest	0.743	0.797	0.769
TabNet	0	0.777	0.666	0.717
TabNet	Mean	0.851	0.621	0.718
TabNet	MissForest	**0.864**	0.790	**0.825**

Recall is the ratio of the number of patients who actually died to the number of patients who predicted as "death", and precision the ratio of the number of patients who were predicted as "death" to the number of patients who actually died. To combine recall and precision, we used F1-score.

In Table 1, while the RF model with missing values to be treated as zero showed the highest precision, the TabNet model with missing value imputation using MissForest showed the best recall and F1-score.

For the Tabnet model with missing value imputation using MissForest, we analyzed top 20 features. These important features include features to indicate whether or not taking 12 months beta-blocker, whether or not taking beta-blocker at the time of first hospitalization, and whether or not taking 12 months oral hypoglycemic agent as illustrated in Figure 3. These 20 features need to be examined more thoroughly to understand their clinical relevance.

TABLE II. PERFORMANCE COMPARISON WITH EXISTING STUDIES

	Performance		
	Recall	**Precision**	**F1-score**
[1]	0.541		0.754
[2]	0.536	**0.851**	0.658
TabNet	**0.864**	0.790	**0.825**

We compared our TabNet model with two major existing models [1,2] (see Table 2) in terms of recall, precision, and F1-score. While [2] showed the highest precision, our TabNet model outperformed according to recall and F1-score. In this TabNet model, we selected top 159 features when trained the model.

V. CONCLUSION

Predicting mortality of heart failure patients is critical in developing clinical treatments and diagnosis in cardiovascular clinics. We used a new version of the KAMIR data from 15,628 patients with 614 features and applied four machine learning methods including TabNet that is a deep learning framework for tabular data. We also handled the missing values of the KAMIR data using three imputation methods. After preprocessing steps including the missing value imputations, 231 features with about 2,000 patients were extracted to build a model. According to F1-score, the Table model with missing value imputation using MissForest and top 159 features showed the best performance. This outperforms comparing to two major existing methods.

The important features of the TabNet model include whether or not taking 12 months beta-blocker, whether or not taking beta-blocker at the time of first hospitalization, and whether or not taking 12 months oral hypoglycemic agent. It is interesting to show that these features are more important than age in this study. Using all 159 features would be impossible to consider for heart failure patients care in clinics. These optimal top features to be selected in clinics for the mortality prediction need to be thoroughly investigated by clinical experts.

REFERENCES

[1] Chicco, D., & Jurman, G. (2020). Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone. *BMC medical informatics and decision making*, *20*(1), 1-16.

[2] Lee, H. C., Park, J. S., Choe, J. C., Ahn, J. H., Lee, H. W., Oh, J. H., ... & Korea Working Group on Myocardial Infarction (KorMI) Investigators. (2020). Prediction of 1-year mortality from acute myocardial infarction using machine learning. *The American Journal of Cardiology*, *133*, 23-31.

[3] Stekhoven, D. J., & Bühlmann, P. (2012). MissForest—non-parametric missing value imputation for mixed-type data. *Bioinformatics*, *28*(1), 112-118.

[4] Liaw A, & Wiener M (2002). “Classification and Regression by randomForest.” *R News*, **2**(3), 18-22.

[5] Chen, T., & Guestrin, C. (2016, August). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 785-794).

[6] Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., ... & Liu, T. Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems*, *30*.

[7] Arik, S. Ö., & Pfister, T. (2021, May). Tabnet: Attentive interpretable tabular learning. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 35, No. 8, pp. 6679-6687).

[8] Akiba, T., Sano, S., Yanase, T., Ohta, T., & Koyama, M. (2019, July). Optuna: A next-generation hyperparameter optimization framework. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining* (pp. 2623-2631).