# Predicting heart failure prognosis using deep learning based on FT-transformer

Geunho Kim
*Medical AI Research Team*
*Chungbuk National University Hospital*
Chungcheongbuk-do, Rep. of Korea
krmsgh1101@naver.com

Minuk Yang
*Medical AI Research Team*
*Chungbuk National University Hospital*
Chungcheongbuk-do, Rep. of Korea
yhu0409@naver.com
(0000-0002-2141-7484)

Geun-Hyeong Kim
*Medical AI Research Team*
*Chungbuk National University Hospital*
Chungcheongbuk-do, Rep. of Korea
kgh5408@nate.com

Seong-Hwan Eom
*Medical AI Research Team*
*Chungbuk National University Hospital*
Chungcheongbuk-do, Rep. of Korea
potatokiller1@naver.com

Tae-Soo Lee
*Department of Biomedical Engineering*
*Chungbuk National University Hospital*
Chungcheongbuk-do, Rep. of Korea
tslee@cbnuh.ac.kr

Seung Park
*Department of Biomedical Engineering*
*Chungbuk National University Hospital*
Chungcheongbuk-do, Rep. of Korea
spark.cbnuh@gmail.com

***Abstract*— Although heart failure (HF) diagnosis and treatment techniques have advanced, more than 50% of HF patients are readmitted. Readmission worsens the life quality of patients due to economic and psychological burdens. Therefore, readmission prediction for patients is important to prevent unnecessary readmissions. We used a feature tokenizer transformer (FT-transformer) to predict readmission by embedding all features and analyzing via transformer encoder. Our experiment with 615 HF patients outperformed conventional machine learning models, achieving an area under the curve of 0.7434 within 28 days, 0.7063 within 3 months, and 0.7039 within 6 months. FT-transformer can potentially improve patient outcomes by enabling early interventions to prevent readmissions.**

***Keywords— Heart failure, Deep learning, FT-transformer, Tabular data***

## I. Introduction

Heart failure (HF) is a medical condition in which the heart is unable to pump enough blood to meet the body's needs. Hospitalization and discharge rates of HF patients increased by 233% from 3 million (1997) to 7 million (2014) [1]. Furthermore, patients hospitalized for HF high risk of readmission because there is no complete cure for HF [2]. Although HF diagnosis and treatment techniques have advanced, approximately 64.3 million people worldwide are still diagnosed with HF [3, 4]. Some studies reported that the treatment cost of HF accounts for a significant portion of the medical budget, and readmissions by HF are caused by premature discharges and misdiagnosis by doctors [5-7]. In addition, 17.6% of patients discharged with HF were readmitted within 30 days, and approximately 50% were readmitted within 1 year [3, 8]. These non-selective readmissions severely deteriorated the quality of life in HF patients [8, 9].

Readmission prevention can improve patient's quality of life and reduce the mental and economic burden of not only patients but also their families. However, readmission prediction is still challenging because it is complex entanglements with the patient's health status, especially chronic disease [8]. Single clinical indicators have been used as HF predictors, but there were no powerful features [10]. Recently, several studies utilized machine learning to predict the readmission of HF patients. Shiyu Chen et al. [11] compared 6 ML models (logistic regression, classification and regression trees, extreme gradient boosting, naive Bayes, support vector machine, and random forest) in predicting within 6 months readmission using 10 prediction features from electronic health record data of 2,002 HF patients. Their study showed that logistic regression performed that logistic regression achieved the best performance with an AUC of 0.634. Awan, Saqib Ejaz, et al. [12] employed a database of 10,757 HF patients to predict readmission or death within 30 days. In addition, considering the imbalance problem of the database, they set the minority class weight to three times the weight of the majority class to produce a model with better generalization. Their study used a multilayer perceptron with a performance of 0.64 AUC. Sarijaloo, FarnazBabaie, et al. described that period of vulnerability for HF patients are likely longer, with the spike in event rates occurring out 90 days post-discharge and plateauing thereafter [13]. Therefore, they used an EHR and echocardiographic report database of 3,189 HF patients that included patient characteristics, vital signs, treatment history, etc to predict readmission within 90 days. Furthermore, they compared 6 models (LR, lasso regression, gradient boosting machine, RF, SVM, Combined lasso regression, and LR). As a result, Combined lasso regression and LR model achieve outperformance (AUC: 0.76).

However, limitations remain in these studies: First, they showed low predictive accuracy using machine learning models, which are an old fashion analysis method. Second, there was no study that comprehensively analyzed considering significant factors such as left ventricular ejection fraction (LVEF) and heart rate.

Aforementioned limitations, we predicted readmission of HF patients using a feature tokenizer transformer (FT-transformer) which achieved superior performance on tabular data analysis. FT-transformer converts all features (categories and numbers) to embeddings and yields high-dimension features

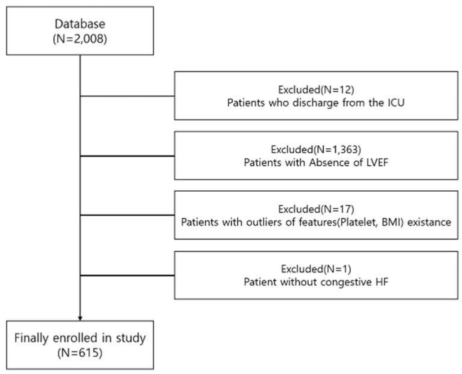


Figure 1. A detailed description of the selection patient process of the study. ICU : Intensive care unit; LVEF : Left ventricular ejection fraction; BMI : Body mass index; HF : Heart failure.

by calculating similarity among features. Specifically, to compare the FT-transformer performance with conventional ML models and clinical indicator, we performed experiments on a public dataset containing 615 HF patients in predicting the probability of readmission within 28 days, 3 months, and 6 months.

## II. Method

### A. Database

The patient-specific, freely accessible database used in this study was obtained from the PhysioNet data portal (https://physionet.org/content/heart-failure-zigong/1.3/) [14]. The database consists of 2,008 patients data admitted for HF collected from the Fourth People's Hospital in China from December 2016 to June 2019. HF was defined according to the European society of cardiology (ESC) criteria [15].

The database contains 166 features, including patient demographic (gender, age, weight, body mass index (BMI), etc), disease states (diabetes, renal failure, etc), laboratory information (uric acid, hemoglobin, urea, etc), and Echocardiographic (LVEF, a ratio of the velocity of mitral valve E wave and A wave (E/A), etc).

### B. Feature selection

Two clinical indicators were referred for feature selection: First, heart failure survival score (HFSS) as an instrument to improve risk stratification in patients with HF [16], which uses seven features such as heart rate, mean blood pressure, and LVEF; Second, Seattle heart failure model (SHFM) as used to predict the prognosis of HF [17], which uses about 20 features consisting of clinical characteristics, drug treatment status, mechanical treatment, and laboratory data [18]. We selected 16 features, including BMI, LVEF, pulse, platelet, uric acid, hemoglobin, sodium, systolic blood pressure, creatine kinase, age, type of HF, gender, Killip grade, new york heart association (NYHA) class, myocardial infarction, Charlson comorbidity index (CCI) score among 166 features by referring to two indicators.

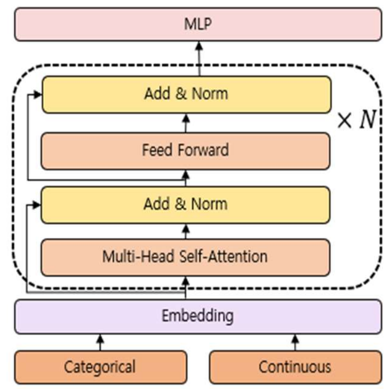


Figure 2. Architecture of our model (FT-transformer). MLP : multi-layer perceptron.

Table 1. Hyperparameters of various models.

| Model | Hyper parameter | parameter (28 days / 3 months / 6 months) |
|---|---|---|
| FT-Transformer | Embedding dimension | 256 / 256 / 256 |
| | Number of heads | 8 / 8 / 8 |
| | Number of layers | 3 / 3 / 3 |
| | Learning rate | 1.0e-3 / 1.0e-3 / 1.0e-3 |
| | Optimizer | Adam / Adam / Adam |
| Logistic Regression | C | 0.001 / 0.0001 / 1 |
| RandomForest | Max depth | 4 / 4 / 10 |
| | Min samples leaf | 16 / 8 / 16 |
| | Min samples split | 6 / 2 / 10 |
| | Number of estimators | 100 / 50 / 50 |
| SVM | C | 0.001 / 0.001 / 0.65 |
| XGBoost | Alpha | 6.86 / 6.864 / 6.74 |
| | Gamma | 0.08 / 0.03 / 0.1 |
| | Max depth | 5 / 6 / 7 |
| | Min child weight | 7.0 / 7.5229 / 7.0 |
| | Colsample bytree | 0.6 / 0.5884 / 0.3 |
| | Eta | 0.1 / 0.298 / 0.1 |

### C. Study population

As shown in Fig. 1, we performed the patient exclusion process to predict the readmission probability of HF patients: First, patients who were discharged from the intensive care unit (ICU) were excluded because follow-up was impossible and the prognosis was uncertain. Second, patients without LVEF

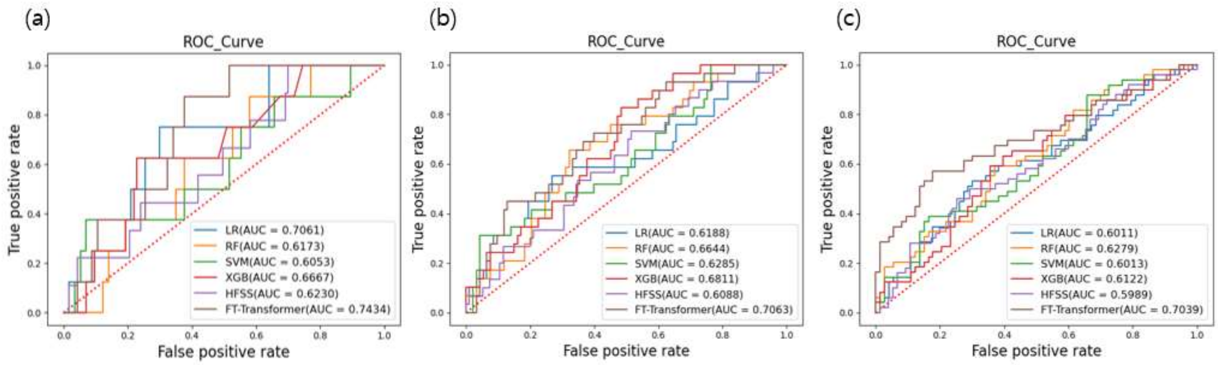


Figure 3. Receiver operating curves (ROC) and areas under the curves (AUCs) for 6 models. (a) readmission within 28 days, (b) readmission within 3 months, (c) readmission within 6 months.

Table 2. Performance of 6 prediction models for within 28 days readmission.

| Model | | Performance | | | |
|---|---|---|---|---|---|
| | | Sensitivity | Specificity | F1-score | AUC |
| Clinical indicator | HFSS | 0.5299 | 0.5556 | 0.6776 | 0.623 |
| Machine leaarning | Logistic Regression | 0.6667 | 0.7500 | 0.7917 | 0.7061 |
| | RandomForest | 0.5439 | 0.6250 | 0.6927 | 0.6173 |
| | SVM | 0.6053 | 0.5000 | 0.738 | 0.6053 |
| | XGBoost | 0.6140 | 0.7500 | 0.7527 | 0.6667 |
| Deep learning | FT-transformer | 0.6228 | 0.8750 | 0.7634 | 0.7434 |

Table 3. Performance of 6 prediction models for within 3 months readmission

| Model | | Performance | | | |
|---|---|---|---|---|---|
| | | Sensitivity | Specificity | F1-score | AUC |
| Clinical indicator | HFSS | 0.6105 | 0.5333 | 0.6946 | 0.6088 |
| Machine leaarning | Logistic Regression | 0.6129 | 0.5517 | 0.6994 | 0.6188 |
| | RandomForest | 0.6344 | 0.5862 | 0.7195 | 0.6644 |
| | SVM | 0.6559 | 0.5172 | 0.7262 | 0.6285 |
| | XGBoost | 0.6452 | 0.5862 | 0.7273 | 0.6811 |
| Deep learning | FT-transformer | 0.6344 | 0.6897 | 0.7329 | 0.7063 |

Table 4. Performance of 6 prediction models for within 6 months readmission

| Model | | Performance | | | |
|---|---|---|---|---|---|
| | | Sensitivity | Specificity | F1-score | AUC |
| Clinical indicator | HFSS | 0.5733 | 0.5400 | 0.6099 | 0.5989 |
| Machine leaarning | Logistic Regression | 0.6027 | 0.5918 | 0.6423 | 0.6011 |
| | RandomForest | 0.6849 | 0.5714 | 0.6944 | 0.6279 |
| | SVM | 0.6400 | 0.5306 | 0.6575 | 0.6013 |
| | XGBoost | 0.6438 | 0.5918 | 0.6714 | 0.6122 |
| Deep learning | FT-transformer | 0.6986 | 0.6122 | 0.7133 | 0.7039 |

affecting HF were excluded [19]. Third, patients with an abnormal platelet count (>= 300) or BMI (< 13) were excluded. Fourth, patients without congestive heart failure were excluded. Finally, we selected 615 among 2,008 patients.

### *D. Data preprocessing*

The database of 615 HF patients contains missing values. Missing values were filled using the KNN method that handles missing values through the mean or median of k nearest values; K was set to 9 [11]. In addition, a label encoder which converts categorical data to numerical data was applied to categorical data. Consequently, a total of 615 study participants with 16 features were used for analysis after preprocessing and utilized the readmission period (within 28 days, 3 months, and 6 months) as the target feature. Note that we performed classification tasks for each of the three readmission periods in this study.

### *E. Readmission prediction model*

In this study, we introduced the FT-transformer suitable for analysis of all features (categorical and numerical variables) to predict readmission of HF patients. FT-transformer is a DL-based model that achieved superior prediction performance in tabular data analysis. As shown in Fig. 2, It briefly demonstrates the architecture of FT-transform. FT-transformer consists of a feature tokenizer that embeds input data and a transformer encoder [20]. FT-transformer applies embedding layers to both numerical and categorical data. On the contrary, the DL-based tab-transformer applies an embedding layer only to categorical data [21, 22]. The features obtained through the transformer encoder are used as the input value of the multi-layer perceptron (MLP) to predict the readmission of HF patients.

### *F. Quantitative evaluation*

To evaluate the readmission prediction model, we employed four statistical metrics: sensitivity, specificity, F1-score, and area under the receiver operating characteristic curve (AUROC or AUC). These are formulated as the following:

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{1}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \tag{2}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{3}$$

$$F_1 = 2 \times \frac{\text{Precision} \times \text{Sensitivity}}{\text{Precision} + \text{Sensitivity}} \tag{4}$$

where TP, FP, TN, and FN represent the number of true positives, false positives, true negatives, and false negatives. The statistical values were calculated as the confusion matrix for the test dataset.

### *G. Model comparison*

To predict the 28-day, 3-month, and 6-month readmission rates, we divided a database of 615 patients with 16 features at a ratio of 8:2 and used 8 ratios for training, 2 ratios for testing database (493 for training and 122 for the testing database). In addition, The FT-transformer model was compared with four ML models (Logistic regression, RandomForest, SVM, XGBoost) which are commonly used in the conventional tabular classification and clinical indicator of HF (HFSS). Table 1 shows the hyperparameters of various models used in this study. We used 256 embedding and MLP dimensions, 3 layers, 1.0e-3 of the learning rate, Adam optimizer, and binary cross-entropy as hyperparameters for our model. In addition, to find optimal hyperparameters of ML models, we utilized GridSearchCV.

## III. Result and Discussion

Data preprocessing and model implementation were held in a Python 3.8 environment using TensorFlow 2. All experiments were processed on a desktop computer with an Intel Core i7-12700 2.10 GHz CPU, 32 GB RAM, and 24 GB NVIDIA GeForce RTX 3090 Ti GPU.

A database of 615 HF patients was used in FT-transformer for readmission prediction. The FT-transformer applied embedding layers to all features, contrary to tab-transformer models applying only categorical features, as mentioned in section II. To evaluate the performance of the FT-transformer, we compared the FT-transformer with 4 ML models and 1 clinical indicator. The ROC curves for each model by readmission periods are shown in Fig. 3. Tables 2, 3, and 4 present the overall sensitivity, specificity, F1-score on the optimal threshold obtained by the ROC curve, and AUC. In addition, we summarized the hyperparameters of each model used in the experiment in Table 1.

The experimental results demonstrated a significant difference in performance between the FT-transformer model and conventional ML models. As shown in Fig. 3, the AUC of all models(FT-transformer, 4 ML models, 1 clinical indicator) exhibited, and the FT-transformer model outperformed in overall periods compared with conventional ML models and clinical indicator. Specifically, it can be seen that the AUC of the FT-transformer is the widest in all periods. In addition, Tables 2, 3, and 4 summarize performance for 3 periods (28 days, 3 months, 6 months) of readmission predictions of clinical indicator, ML, and FT-transformer. Specifically, SVM recorded poor performance with 0.6053 of AUC for readmission prediction within 28 days in Table 2. Clinical indicator recorded poor performance with 0.6088 and 0.5989 of AUC for readmission prediction within 3 and 6 months, respectively in Tables 2 and 3. Compared to FT-transformer based on DL calculating by a non-linear equation, the clinical indicator was calculated by a linear equation that classifies targets through a linear model, it is difficult to analyze complex data of non-linear relationships. In addition, ML models were calculated by patterns between features, becomes difficult to predict the result when data of other patterns is added. On the contrary, FT-transformer was calculated through the similarity between features. Due to this difference, FT-transformer achieved outperformance in Tables 2, 3, and 4. According to the AUC of readmission 0.7434 (within 28 days), 0.7063 (within 3 months), and 0.7039 (within 6 months), the short prediction period causes a high readmission prediction rate. As the prediction period increases, changes in the patient's characteristics, such as health status, vital signs, etc, accordingly can lead to such results.

However, this study has several limitations. First, the study conducted with small databases can be difficult to generalize. Therefore, it is necessary to revalidate in a large-scale database. Finally, the target (readmission) of the database is limited to a specific period. In the future, we plan to expand the study to a model that predicts more HF prognosis by utilizing a database of mortality and different lengths of readmission. The number of HF patients deaths was very low at 9 in the database used in this study, and only readmission was used as a target.

## IV. Conclusion

HF is a serious medical condition that affects a large number of people worldwide. Despite advances in diagnosis and treatment for HF, readmission rates for HF patients remain high. These readmissions cause significant negative impacts on patients' quality of life by economic burdens such as high costs of healthcare. Predicting readmission is a complex challenge due to the many factors involved, especially chronic diseases.

In this study, we selected 16 features of the 615 HF patients database using clinical indicators (HFSS, SHFM) to predict readmission rates. Label encoder was applied to categorical data and the KNN method was used to fill missing values for continuous features. We predicted the probability of readmission periods within 28 days, 3 months, and 6 months of HF patients using FT-transformer, respectively, and compared its performance with 4 conventional machine learning models and clinical indicator. Compared to other models, our model achieved outperformance. The study's results suggest that the FT-transformer is a promising approach for predicting the readmission probability of HF patients.

## ACKNOWLEDGEMENTS

This research was supported by a grant of the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (grant number: HI21C1074070021).

## REFERENCES

[1] E. J. Benjamin *et al.*, "Heart disease and stroke statistics—2019 update: a report from the American Heart Association," *Circulation,* vol. 139, no. 10, pp. e56-e528, 2019.

[2] D. i. C. W. t. A. C. o. C. Physicians *et al.*, "ACC/AHA 2005 guideline update for the diagnosis and management of chronic heart failure in the adult—summary article: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Writing Committee to Update the 2001 Guidelines for the Evaluation and Management of Heart Failure)," *Journal of the American College of Cardiology,* vol. 46, no. 6, pp. 1116-1143, 2005.

[3] A. Groenewegen, F. H. Rutten, A. Mosterd, and A. W. Hoes, "Epidemiology of heart failure," *European journal of heart failure,* vol. 22, no. 8, pp. 1342-1356, 2020.

[4] N. L. Bragazzi *et al.*, "Burden of heart failure and underlying causes in 195 countries and territories from 1990 to 2017," *European Journal of Preventive Cardiology,* vol. 28, no. 15, pp. 1682-1690, 2021.

[5] K. Dharmarajan *et al.*, "Diagnoses and timing of 30-day readmissions after hospitalization for heart failure, acute myocardial infarction, or pneumonia," *Jama,* vol. 309, no. 4, pp. 355-363, 2013.

[6] C. Annema, M.-L. Luttik, and T. Jaarsma, "Reasons for readmission in heart failure: perspectives of patients, caregivers, cardiologists, and heart failure nurses," *Heart & Lung,* vol. 38, no. 5, pp. 427-434, 2009.

[7] R. P. Kocher and E. Y. Adashi, "Hospital readmissions and the Affordable Care Act: paying for coordinated quality care," *Jama,* vol. 306, no. 16, pp. 1794-1795, 2011.

[8] C. Xiao, T. Ma, A. B. Dieng, D. M. Blei, and F. Wang, "Readmission prediction via deep contextual embedding of clinical concepts," *PloS one,* vol. 13, no. 4, p. e0195024, 2018.

[9] A. Kimmoun *et al.*, "Temporal trends in mortality and readmission after acute heart failure: a systematic review and meta‐regression in the past four decades," *European Journal of Heart Failure,* vol. 23, no. 3, pp. 420-431, 2021.

[10] J. J. A. d. Santos, J. E. A. Plewka, and P. R. S. Brofman, "Quality of life and clinical indicators in heart failure: a multivariate analysis," *Arquivos brasileiros de cardiologia,* vol. 93, pp. 159-166, 2009.

[11] S. Chen *et al.*, "Predicting Six-Month Re-Admission Risk in Heart Failure Patients Using Multiple Machine Learning Methods: A Study Based on the Chinese Heart Failure Population Database," *Journal of Clinical Medicine,* vol. 12, no. 3, p. 870, 2023.

[12] S. E. Awan, M. Bennamoun, F. Sohel, F. M. Sanfilippo, and G. Dwivedi, "Machine learning‐based prediction of heart failure readmission or death: implications of choosing the right model and the right metrics," *ESC heart failure,* vol. 6, no. 2, pp. 428-435, 2019.

[13] F. Sarijaloo, J. Park, X. Zhong, and A. Wokhlu, "Predicting 90 day acute heart failure readmission and death using machine learning‐supported decision analysis," *Clinical cardiology,* vol. 44, no. 2, pp. 230-237, 2021.

[14] Z. Zhang *et al.*, "Hospitalized patients with heart failure: integrating electronic healthcare records and external outcome data," in *PhysioNet*, vol. 101, no. 23), 2020, pp. e215-e220.

[15] E. J. Benjamin *et al.*, "Heart disease and stroke statistics—2017 update: a report from the American Heart Association," *circulation,* vol. 135, no. 10, pp. e146-e603, 2017.

[16] C. Zugck *et al.*, "Risk stratification in middle‐aged patients with congestive heart failure: prospective comparison of the Heart Failure Survival Score (HFSS) and a simplified two‐variable model," *European journal of heart failure,* vol. 3, no. 5, pp. 577-585, 2001.

[17] A. P. Kalogeropoulos *et al.*, "Utility of the Seattle Heart Failure Model in patients with advanced heart failure," *Journal of the American College of Cardiology,* vol. 53, no. 4, pp. 334-342, 2009.

[18] A. Goda, P. Williams, D. Mancini, and L. H. Lund, "Selecting patients for heart transplantation: comparison of the Heart Failure Survival Score (HFSS) and the Seattle heart failure model (SHFM)," *The Journal of heart and lung transplantation,* vol. 30, no. 11, pp. 1236-1243, 2011.

[19] M. Toma *et al.*, "The relationship between left ventricular ejection fraction and mortality in patients with acute heart failure: insights from the ASCEND‐HF T rial," *European journal of heart failure,* vol. 16, no. 3, pp. 334-341, 2014.

[20] A. Vaswani *et al.*, "Attention is all you need," *Advances in neural information processing systems,* vol. 30, 2017.

[21] X. Huang, A. Khetan, M. Cvitkovic, and Z. Karnin, "Tabtransformer: Tabular data modeling using contextual embeddings," *arXiv preprint arXiv:2012.06678,* 2020.

[22] Y. Gorishniy, I. Rubachev, V. Khrulkov, and A. Babenko, "Revisiting deep learning models for tabular data," *Advances in Neural Information Processing Systems,* vol. 34, pp. 18932-18943, 2021.