

Optimization of the Classic Transfer-Stacking Model Migration Algorithm: A Way to Solve Time-Varying Performance Degradation of Acute Kidney Injury Clinical Prediction Model

Yunfei Xue

Department of Computer Science, Jinan University, Guangzhou, China

Email: xueyunfly@163.com

How to cite this paper: Xue, Y.F. (2021) Optimization of the Classic Transfer-Stacking Model Migration Algorithm: A Way to Solve Time-Varying Performance Degradation of Acute Kidney Injury Clinical Prediction Model. *Journal of Biosciences and Medicines*, 9, 14-28.

<https://doi.org/10.4236/jbm.2021.94002>

Received: February 20, 2021

Accepted: April 4, 2021

Published: April 7, 2021

Copyright © 2021 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Acute Kidney Injury (AKI) is one of the most common acute and critical illnesses in general wards and intensive care units. Its high morbidity and high fatality rate have become a major global public health problem. There are often serious lags in clinical diagnosis of AKI. Early diagnosis and timely intervention and effective care become critical. The use of electronic medical record data to build an AKI risk prediction model has been proven to help prevent the occurrence of AKI. However, in actual clinical applications, the distribution of historical data and new data will continue to vary over time, resulting in a significant decrease in the performance of the model. How to solve the problem of model performance degradation over time will be a core challenge for the long-term use of predictive models in clinical applications. Aiming at the above problems, this paper studies the classic Transfer-Stacking model migration algorithm. Aiming at the lack of this algorithm, such as the loss of a large amount of feature information of the target domain and poor fit when integrating the model of the target domain, the Accumulate-Transfer-Stacking algorithm is proposed to improve it. Improvements include: 1) Optimize the input vector and model integration algorithm of Transfer-Stacking's target domain model. 2) Optimize Transfer-Stacking from a single-source domain model to a multi-source domain model. The experimental results show that for the improved algorithm proposed in this paper when the data is sufficient and insufficient, the average AUC value of the model on the data of subsequent years is 0.89 and 0.87, and the average F1 Score value is 0.45 and 0.36. Moreover, this method is significantly better than the unimproved Transfer-Stacking algorithm and baseline method, and can effectively overcome the problem of data distribution heterogeneity caused by time factors.

Keywords

Acute Kidney Injury, Electronic Medical Record, Risk Prediction, Transfer Learning

1. Introduction

There is currently no specific treatment for AKI, and renal replacement therapy (*i.e.*, kidney transplantation or kidney dialysis) is required in severe cases. If the risk of acute kidney injury can be assessed in advance and intervention and management can be carried out, it will have a better prognosis than treatment after the occurrence of AKI [1]. However, on the one hand, the causes of acute kidney injury are complicated, and on the other hand, the diagnosis method has serious lag. When the diagnosis is made, patients often have missed the optimal treatment period. Acute kidney injury is often missed or delayed in clinical diagnosis, and missed diagnosis has been proven to be an independent risk factor for patient death [1]. Authoritative literature shows that the missed diagnosis rate of acute kidney injury in hospitalized patients in the United States is estimated to be 25% [2]. The missed diagnosis rate of acute kidney injury in hospitalized patients in my country is as high as 74.2%, and 17.6% of confirmed cases are delayed diagnosis [3]. Reducing the missed diagnosis rate of acute kidney injury, early diagnosis of acute kidney injury and early prediction of the risk of acute kidney injury are all of great significance.

Electronic Medical Record (EMR) records comprehensive data during the patient's health and medical process, including a large amount of historical information such as the patient's basic condition, laboratory measurement index values, vital signs, medical history, medication, treatment records, etc. [4]. The information it contains is closely related to the patient's health and treatment. Analyzing electronic medical record data and constructing intelligent disease risk prediction models have gradually become one of the emerging research hotspots in the interdisciplinary research field of medical information and data mining.

Today, most hospitals are equipped with electronic medical record systems and have accumulated a large amount of electronic medical record data suitable for computer analysis, which can be used to establish acute kidney injury risk prediction models [3] [4] [5] [6]. The predictive model helps doctors identify high-risk patients with acute kidney injury in advance and reduces the missed diagnosis rate, making it possible to intervene and care for patients with acute kidney injury in advance. The acute kidney injury risk prediction model based on electronic medical records can significantly reduce the harm caused by the high incidence, high mortality and high disability rate of acute kidney injury [5] [7] [8] [9].

The key issue in the field of acute kidney injury prediction is that the inci-

dence of acute kidney injury may be affected by a variety of diseases and treatments, and the pathogenesis is diverse, and the main risk mechanisms faced by different patients are not the same. Therefore, when the distribution of patients, treatment methods, hospital conditions, etc. change over time, the effect of the original model is likely to decrease significantly [10] [11] [12] [13] [14]. However, using new data to retrain the model may face problems such as high data collection and cleaning costs, as well as insufficient data [15] [16] [17]. The research focus of transfer learning is to deal with the heterogeneity between new and old data, and to formulate appropriate transfer learning strategies based on the heterogeneity of data [15]. At present, there is still a lack of research on transfer learning in the field of acute kidney injury electronic medical record analysis, and whether the existing transfer learning methods can adapt to the heterogeneity between the new and old acute kidney injury electronic medical records data is lack of sufficient verification.

Based on this key issue, we studied the transfer-stacking [18], a classic model transfer learning method. The model-based transfer learning method assumes that the model of the source domain and the model of the target domain share the prior distribution of some parameters or hyperparameters. It realizes model migration by directly migrating the model parameters shared between the two domains from the model of the source domain to the model of the target domain. The Transfer-Stacking is mainly divided into two stages for model training: 1) In the training set processing stage of the source domain, N different machine learning algorithms are trained using source domain data to obtain a set of N source domain models. 2) In the data processing stage of the target domain, the N source domain models are used to predict the target data, and the prediction result is used as the input of the logistic regression algorithm to train the final prediction model. The study found that Transfer-Stacking is difficult to adapt to the electronic medical record data migration tasks in different years. The main reasons are: 1) The algorithm has the problem of losing a lot of data characteristic information of the target domain. 2) The algorithm has a poor fit when integrating the model of the target domain. In response to the above problems, we improved the Transfer-Stacking algorithm and proposed the Accumulate-Transfer-Stacking algorithm. Experimental results show that the improved algorithm is significantly better than the original Transfer-Stacking algorithm, and effectively solves the problem of model performance degradation.

2. Materials and Methods

The experimental data used in this study is derived from the electronic medical record data of inpatients in general wards admitted to the hospital from 2010 to 2017 collected by the cooperative hospital. In order to meet the needs of the experiment, this study removed the electronic medical record data of the following three types of patients: 1) Patients whose hospital stay is less than 24 hours. Because the prediction window length of this study is 24 hours, the length of hos-

pital stay must be greater than the length of the prediction window. 2) It is impossible to judge whether patients suffered from acute kidney injury during hospitalization, that is, patients whose serum creatinine concentration (SCr) recorded during hospitalization was less than twice. Because it is necessary to measure the serum creatinine concentration at least twice to determine whether the patient suffered from acute kidney injury during hospitalization. 3) Patients with moderate or severe renal dysfunction at the time of admission, if the serum creatinine concentration measured within 24 hours of admission is higher than 1.3 mg/dl or the estimated glomerular filtration rate (eGFR) is lower than 60 ml/min/1.73m². After screening by the above conditions, we obtained a total of 141,696 electronic medical records that can be used in this study. **Figure 1** shows the detailed electronic medical record screening process.

In this study, according to the Common Data Model (CDM) [19] [20] standard data collection format, the patient's demographic information (Demographic), vital signs (Vital), laboratory measurement indicators (Lab Test), complications (Comorbidity), type of surgery (Procedure), and medication (Medication) in six categories. The detailed data collection format of each sample is shown in **Table 1**.

2.1. Transfer-Stacking Algorithm

Transfer-Stacking is a commonly used method in the field of model-based transfer methods. Transfer-Stacking is improved from the traditional Stacking modeling framework, and the model training is carried out in two stages [18] [21] [22] [23]. Stacking algorithm is an example of integrated learning algorithm, which is a hierarchical model integration framework [18] [24]. Taking two layers as an example, the first layer is composed of multiple base models, whose input is the original training set, and the second layer model uses the output of the first layer base model as a feature to construct a new training data

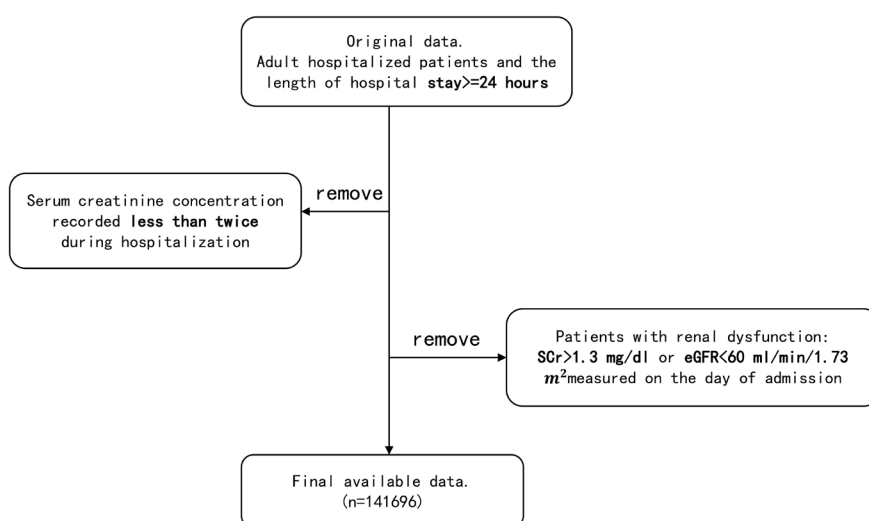


Figure 1. The screening process of electronic medical records.

Table 1. CDM standard data collection format.

Feature category	Feature number	details
Demographic	4	AGE, HISPANIC, RACE, SEX
Vital	8	HEIGHT, WEIGHT, BODY MASS INDEX, SMOKING, TOBACCO, TOBACCO TYPE, SYSTOLIC BLOOD PRESSURE, DIASTOLIC BLOOD PRESSURE
Lab Test	812	Measurement values of 812 laboratory measurement indicators
Comorbidity	280	Complications based on CDM classification and coding standards
Procedure	15,335	Types of surgery based on CDM classification and coding standards
Medication	11,539	Drugs used before and during hospitalization

matrix for retraining, so as to get a complete Stacking model. The detailed description of the Stacking integration method is shown in **Algorithm 1**.

As shown in **Figure 2**, in the training set processing stage of the source domain, N different machine learning algorithms are trained using source domain data to obtain a set of N source domain models. In the data processing stage of the target domain, the N source domain models are used to predict the target data, and the prediction result is used as the input of the logistic regression algorithm to train the final prediction model.

2.2. Improvement Based on Transfer-Stacking Algorithm

Transfer-Stacking has the following problems in the application scenarios of this research: 1) In the target domain training set processing stage, only the source domain model is used to predict the output of the target domain data as the training feature of the final model, and the original features of the target domain are completely discarded. This may lose a lot of useful information in the original features. 2) The final integrated model of Transfer-Stacking is a logistic regression model with poor fit [25]. 3) The past method is aimed at the static target domain, so the classification model of the integrated source domain in Transfer-Stacking is fixed and unique. However, when dealing with the problem of data distribution differences caused by time changes, the data distribution differences between the source domain and the target domain will continue to change over time, which will cause the original source model performance to gradually decline. Aiming at the shortcomings of Transfer-Stacking, this research proposes the Accumulate-Transfer-Stacking algorithm. Its key technical improvements include:

1) Change the Transfer-Stacking model from a single-source domain form to a multi-source domain form. Multiple source domains are used to train multiple source domain models, and the source domain model's prediction results of target domain data are spliced with the original target domain features as the input of the final target domain model. The detailed generation process of the input vector of the target domain model is shown in **Figure 3**.

Algorithm 1. Stacking ensemble.

input: training data set $D = \{(x_i, y_i)\}_{i=1}^m$
output: Integrated classifier H

1. Learning base classifier
 For $t = 1, 2, \dots, T$
 Learn the base classifier h_t on the data set D
2. Construct a new data set D_h
 For $i = 1, 2, \dots, m$
 $D_h = \{(x'_i, y_i)\}$, in $x'_i = \{h_1(x_i), \dots, h_T(x_i)\}, i = 1, 2, \dots, m$
3. Learn a meta-classifier, H on the data set D_h
4. Return H

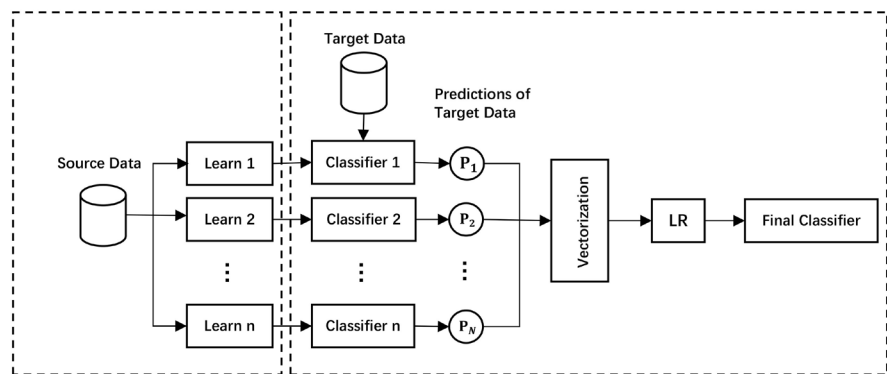


Figure 2. Transfer-stacking algorithm flowchart.

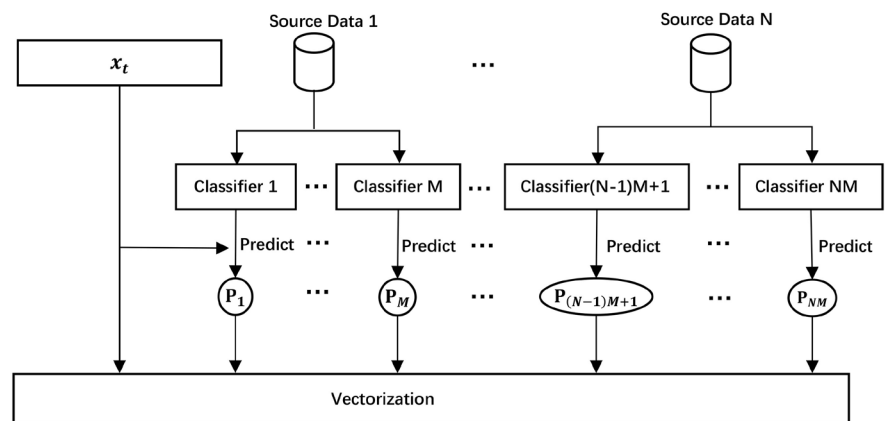


Figure 3. The input vector generation process of the target domain model.

2) Improve the final target domain model modeling strategy. Compared with logistic regression, LightGBM is a classification model with high fit [26]. Accumulate-Transfer -Stacking replaces the target domain model with the LightGBM model.

The complete process of the Accumulate-Transfer-Stacking algorithm is shown in **Figure 4**.

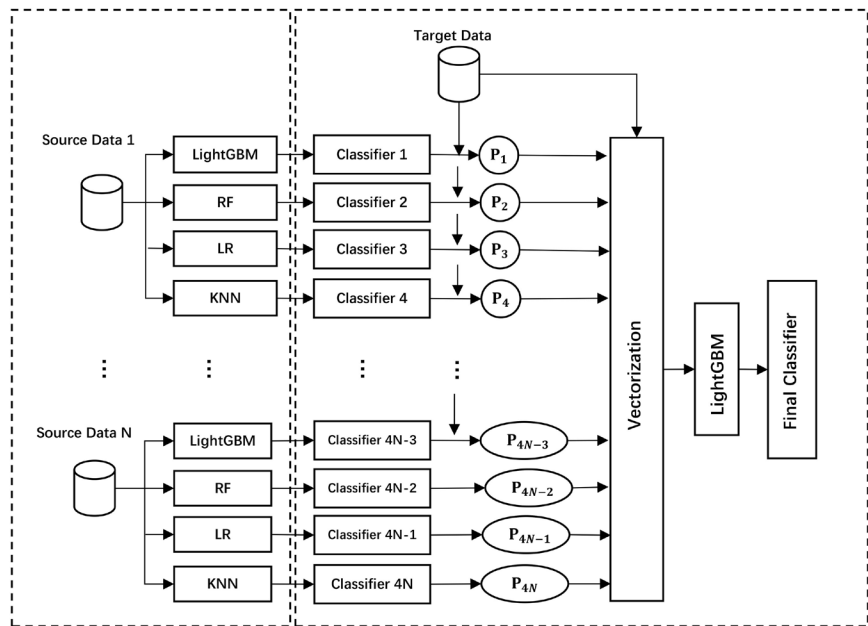


Figure 4. Accumulate-transfer-stacking algorithm flowchart.

1) The input and output of the model. Input N source domain data sets and a target domain data set; the output model is the source domain model set and the final target domain model LightGBM.

2) The model training phase of the source domain. Four machine learning algorithms, LightGBM [26], Random Forest [27], Logistic Regression [28], and K-Nearest Neighbors [29] are trained on the data set of each source domain. A model set of the source domain composed of $4N$ models can be obtained.

3) The training phase of the target domain model. The $4N$ source domain model is used to predict the target domain data, and the output result and the input feature are spliced into a new feature vector. The new feature vector is used as the input of the LightGBM model. After training, the final LightGBM model is obtained.

3. Results

3.1. Evaluation Metrics

We used the area under the receiver operator curve (AUC) and balanced F Score (F1 Score) to compare the overall prediction performance, with the latter known to be more robust to imbalanced datasets. In addition, True positive (TP), True negative (TN), False positive (FP), and False negative (FN) were used to evaluate the results, as defined in **Table 2**.

1) Precision

Precision represents the proportion of samples whose category is actually positive among the samples predicted to be positive by the model. The calculation formula is:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (1)$$

Table 2. Definition of the TP, TN, FP and FN.

Category	Predicted	
	Predicted target area	Non-predicted target area
Actual target area	True Positive (TP)	False Negative (FN)
Non-actual target area	False Positive (FP)	True Negative (TN)

2) Recall

Recall represents the proportion of samples that are actually positive by the model. The calculation formula is:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (2)$$

3) F1 Score

F1 Score is the harmonic average of precision and recall. The calculation formula is:

$$\text{F1 Score} = \frac{2\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} = \frac{2\text{TP}}{2\text{TP} + \text{FP} + \text{FN}} \quad (3)$$

The true meaning of AUC refers to the probability that a positive sample and a negative sample are randomly given, and the probability of the positive sample output by the classifier is greater than the probability of the negative sample output by the classifier, which can be used as an evaluation index of the average performance of the classifier. The calculation formula is:

$$\text{AUC} = \frac{\sum_{i \in \text{positive}} \text{rank}_i - \frac{M \times (M + 1)}{2}}{M \times N} \quad (4)$$

M is the number of positive samples, N is the number of negative samples, rank_i is the order of the probability of positive samples from high to low. It can be seen that the larger the AUC value, the better the classification result and the more reliable the performance of model prediction.

3.2. Analysis of Experimental Results of Model Performance Degradation

This experiment explores how the performance of the AKI prediction model developed by five commonly used machine learning algorithms changes over time. This experiment uses 2010 data to train five common machine learning models, and uses five-fold cross-validation for internal performance testing. These five models include: Logistic Regression (LR), Decision Tree (DT), Random Forest (RF), K-Nearest Neighbor (KNN) and LightGBM. In order to test their model performance in the time dimension; This experiment uses 7 years (2011 to 2017) of test data to test the performance of these models year by year to verify whether these models can maintain stable performance during long-term use.

In the process of year-by-year performance testing of the model, the change of

AUC over time is shown in **Figure 5**. Different models have different sensitivity to time changes. The AUC of LightGBM continues to decline over time. The performance trend of RF and DT is similar to LightGBM. Compared with LR and KNN, LightGBM, RF and DT require regular calibration or retraining of models. Regularly updating the model is essential to maintain the accuracy of these three models. These findings have important implications for the long-term application of acute kidney injury models in clinical decision-making.

The AUC and F1 Score verified by each model are shown in **Table 3**. The AUC range of each model is [0.601, 0.834], and the F1 Score range is [0.163, 0.505]. Whether it is AUC or F1 Score, the LightGBM model has achieved the best predictive performance. Its AUC and F1 Score are 0.829 and 0.486, respectively. KNN has the worst AUC performance, with an AUC of only 0.612. LR's F1 Score has the worst performance, and its F1 Score is only 0.197.

3.3. Analysis of Algorithm Validity Experiment Results

Figure 6 shows the performance of Accumulate-Transfer-Stacking and Transfer-Stacking in the case of sufficient and insufficient data in the target domain. Baseline is a source domain model trained purely based on 2010 data. Comparing the results of Transfer-Stacking and the baseline, it can be seen that the AUC

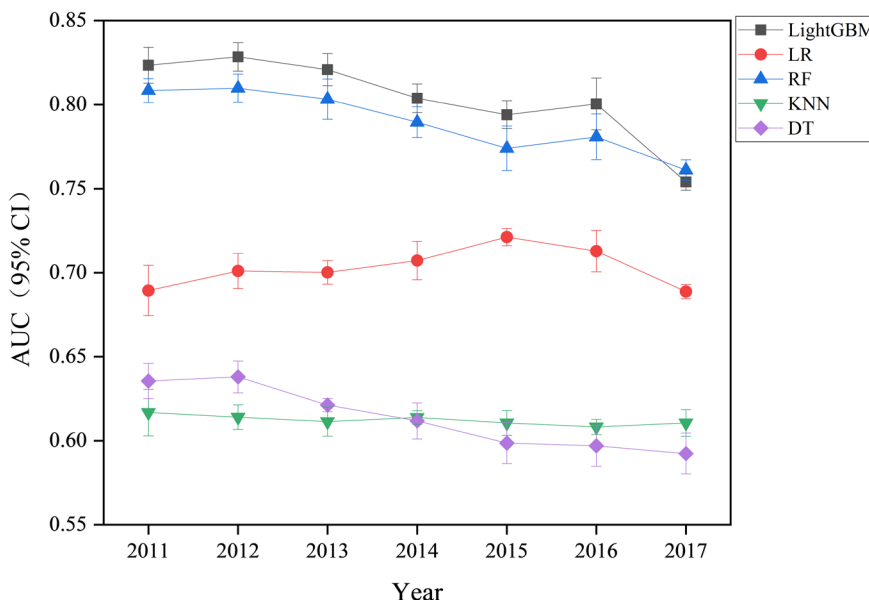


Figure 5. Changes in AUC of each model over time.

Table 3. Internally verified performance of each model (95% CI).

	LightGBM	LR	RF	KNN	DT
AUC	0.829 [0.824, 0.834]	0.704 [0.689, 0.718]	0.813 [0.803, 0.822]	0.612 [0.601, 0.623]	0.634 [0.621, 0.648]
F1 Score	0.486 [0.467, 0.505]	0.197 [0.163, 0.231]	0.397 [0.379, 0.415]	0.206 [0.193, 0.219]	0.391 [0.369, 0.413]

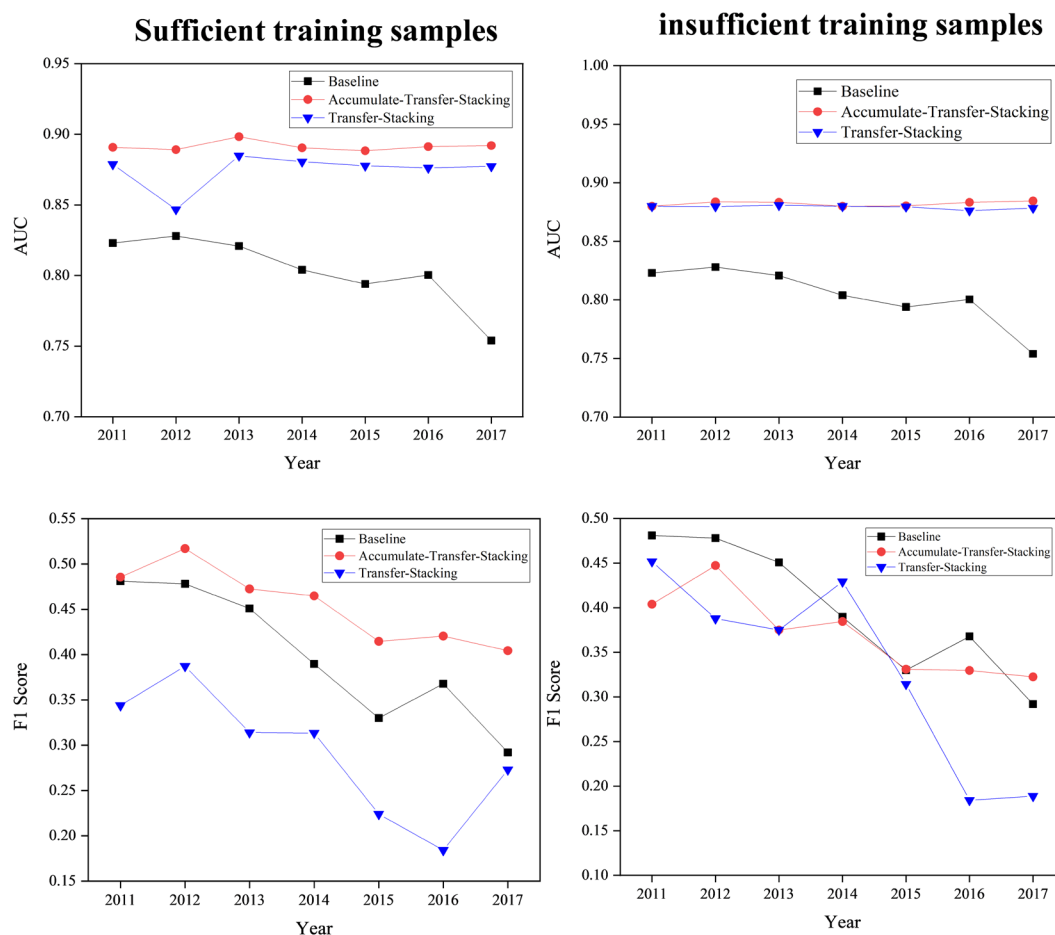


Figure 6. Accumulate-transfer-stacking improvement effect comparison.

of Transfer-Stacking is significantly better than the results of the baseline regardless of whether the data is sufficient or insufficient. This shows that the model can effectively use the knowledge learned from the source domain when making predictions in the target domain. However, the F1 score of Transfer-Stacking is not as good as the results of the baseline, especially in the years when the data is insufficient and the data heterogeneity is large, the F1 score of Transfer-Stacking drops significantly. This is because Transfer-Stacking does not use the original features of the target domain, and the combiner uses a relatively simple logistic regression model, which makes it unable to effectively learn the information about the data distribution change. Then pay attention to the Accumulate-Transfer-Stacking proposed in this article. When the data is sufficient, AUC is significantly better than Transfer-Stacking and Baseline. When the data is insufficient, Accumulate-Stacking's F1 score is significantly better than Transfer-Stacking, and there is no significant decline in the year when the data heterogeneity is greatest. The results show that Accumulate-Transfer-Stacking uses LightGBM as the combiner and the optimization strategy of splicing the original features of the target domain, which overcomes the problems of the original Transfer-Stacking and achieves better results.

Comprehensive comparison of the performance of Accumulate-Transfer-Stacking, Transfer-Stacking and Baseline. Regardless of the scenario where the target domain training sample size is sufficient or insufficient, the AUC and F1 Score of Accumulate-Transfer-Stacking are better than those of Transfer-Stacking, which proves that the improvement of Accumulate-Transfer-Stacking is indeed effective. However, in the scenario where the sample size of the target domain training set is insufficient, the F1 Score of Accumulate-Transfer-Stacking and Transfer-Stacking are prone to negative transfer.

3.4. Odds Ratio Analysis of Disease Risk Factors

In order to analyze the mechanism of the impact of important characteristics on acute kidney injury and the degree of association, this paper conducts Odds Ratio Analysis on the top 100 demographic information and drug-related characteristics in the most important feature set. This study calculated the Odds Ratio (OR) and the Confidence Interval (CI) between related features and acute kidney injury. OR is used to measure the relationship between exposure to specific characteristics and risk of disease [30]. If the OR value is less than 1, it means that the characteristic exposure has a protective effect on the disease; if the OR value is greater than 1, it means that the characteristic exposure is a risk factor for the disease. The relationship between disease and characteristics has four situations as shown in **Table 4**.

OR, as an estimate of relative risk, is the most commonly used analysis method in medical record control studies, and its calculation formula is as follows:

$$\text{Odds Ratio} = \frac{\text{Ratio of disease in exposure group}}{\text{Ratio of disease in the non-exposed group}} = \frac{a/c}{b/d} = \frac{ad}{bc} \quad (5)$$

The formula for calculating the 95% confidence interval is as follows:

$$95\% \text{CI of Odds Ratio} = e^{\ln(\text{Odds Ratio}) \pm 1.96 \left(\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d} \right)^{0.5}} \quad (6)$$

This paper analyzes demographic information and medication is at risk odds ratio under different exposure conditions. Due to the large number of drugs, **Table 5** only shows the results with 95% significance. By observing the table below, it can be concluded that the risk of acute kidney injury in patients is increasing with age. Especially when the patient reaches 56 years old, the risk of acute kidney injury will increase significantly. Indians and African Americans have a higher risk of acute kidney injury than other races. Among them, in addition to the PULMOZYME, Cefazolin and NARCAN, which can reduce the risk of disease, other drugs will significantly increase the risk of acute kidney injury

Table 4. Correspondence between diseases and characteristics.

	Exposure	Non-exposed
AKI	a	b
Non-AKI	c	d

Table 5. Odds Ratio analysis of disease risk under different exposures.

Feature	Odds Ratio	95%CI Lower	95%CI Upper
AGE			
18 - 25	0.717	0.667	0.771
26 - 35	0.694	0.655	0.737
36 - 45	0.823	0.781	0.866
46 - 55	0.895	0.869	0.932
56 - 65	1.123	1.086	1.163
>65	1.265	1.226	1.304
RACE			
American Indian	1.157	0.917	1.460
Asian	0.958	0.821	1.118
African American	1.083	1.039	1.130
Native Hawaiian	0.993	0.575	1.716
White	0.911	0.880	0.942
Multiple race	0.838	0.594	1.182
Other	0.849	0.716	1.006
SEX			
male	1.332	1.293	1.373
female	0.750	0.728	0.773
MEDICATION			
FUROSEMIDE (LASIX) BOLUS FOR CONTINUOUS INFUSION	12.058	1.093	133.000
PIPERACILLIN-TAZOBACTAM-DEXTRS 4.5 GRAM/100 ML IV PGBK	1.778	1.16	2.710
PULMOZYME 1 MG/ML IN SOLN	0.187	0.088	0.397
CEFAZOLIN IV	0.935	0.463	1.888
NARCAN 0.4 MG/ML IJ SOLN	0.445	0.279	0.709
PRINIVIL 20 MG PO TAB	1.184	1.053	1.331
MAGNESIUM SULFATE IN D5W 1% IV SOLN	1.615	1.527	1.709
CARVEDILOL 3.125 MG PO TAB	1.340	1.238	1.450
VECURONIUM BROMIDE 10 MG IV SOLR	1.696	1.246	2.308
SPIRONOLACTONE 100 MG PO TAB	1.401	1.245	1.578
INSULIN GLARGINE 100 UNIT/ML SC SOLN	1.655	1.392	1.969

in patients, including: Furosemide (LASIX) Bolus, Piperacillin, PRINIVIL, Magnesium Sulfate, Carvedilol, VECURONIUM Bromide, Spironolactone and Insulin Glargine.

4. Discussion

Problems in this research and possible future improvements:

Consider more prediction windows. One day (24 h) is the most common prediction window length in the existing research literature, so one day is also selected for the prediction window of this study. However, with changes in clinical application scenarios, prediction windows of other lengths may be required. The longer the prediction window, the earlier the doctor can be warned, and the longer it will take for the doctor to make clinical decisions. Existing studies have shown that the performance of models with different prediction window lengths is different. When the acute kidney injury prediction model has a longer prediction window, it is a feasible research direction to explore the change of model performance over time and establish a model with stable performance.

5. Conclusions

This study analyzes the gradual decline in performance of five AKI prediction models over time, and studies the effect of using transfer learning strategies to deal with this problem. This paper studies the modeling effect of the classic Transfer-Stacking model migration algorithm, and designs an improvement strategy for its shortcomings in this research scenario. The experimental results show that the Accumulate-Transfer-Stacking algorithm proposed in this paper is significantly better than the original Transfer-Stacking algorithm when the data is sufficient and insufficient, and it can effectively overcome the problem of data distribution heterogeneity caused by time factors. In terms of mining important risk factors, this paper conducts an odds ratio analysis of important features and finds that 11 drugs will significantly affect the risk of acute kidney injury in patients.

This research belongs to the interdisciplinary research of computer science and medical informatics. The research results of this paper not only enrich the theory of machine learning algorithms, but also study the key issues in the field of acute kidney injury. Therefore, this research not only has strong theoretical research significance, but also has important practical application value.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

- [1] Khwaja, A. (2012) KDIGO Clinical Practice Guidelines for Acute Kidney Injury. *Nephron Clinical Practice*, **120**, c179-c184. <https://doi.org/10.1159/000339789>
- [2] Wilson, F.P., Bansal, A.D., Jasti, S.K., *et al.* (2013) The Impact of Documentation of Severe Acute Kidney Injury on Mortality. *Clinical Nephrology*, **80**, 417-425. <https://doi.org/10.5414/CN108072>
- [3] Wilson, F.P., Shashaty, M., Testani, J., *et al.* (2015) Automated, Electronic Alerts for Acute Kidney Injury: A Single-Blind, Parallel-Group, Randomised Controlled Trial. *The Lancet*, **385**, 1966-1974. [https://doi.org/10.1016/S0140-6736\(15\)60266-5](https://doi.org/10.1016/S0140-6736(15)60266-5)
- [4] Bagshaw, S.M., Goldstein, S.L., Ronco, C., *et al.* (2016) Acute Kidney Injury in the

- Era of Big Data: The 15th Consensus Conference of the Acute Dialysis Quality Initiative (ADQI). *Canadian Journal of Kidney Health and Disease*, **3**, 5. <https://doi.org/10.1186/s40697-016-0103-z>
- [5] Wu, L., Hu, Y., Liu, X., *et al.* (2018) Feature Ranking in Predictive Models for Hospital-Acquired Acute Kidney Injury. *Scientific Reports*, **8**, Article No. 17298. <https://doi.org/10.1038/s41598-018-35487-0>
- [6] He, J., Hu, Y., Zhang, X., *et al.* (2019) Multi-Perspective Predictive Modeling for Acute Kidney Injury in General Hospital Populations Using Electronic Medical Records. *JAMIA Open*, **2**, 115-122. <https://doi.org/10.1093/jamiaopen/ooy043>
- [7] Cronin, R.M., VanHouten, J.P., Siew, E.D., *et al.* (2015) National Veterans Health Administration Inpatient Risk Stratification Models for Hospital-Acquired Acute Kidney Injury. *Journal of the American Medical Informatics Association*, **22**, 1054-1071. <https://doi.org/10.1093/jamia/ocv051>
- [8] Koynier, J.L., Adhikari, R., Edelson, D.P., *et al.* (2016) Development of a Multicenter Ward-Based AKI Prediction Model. *Clinical Journal of the American Society of Nephrology*, **11**, 1935-1943. <https://doi.org/10.2215/CJN.00280116>
- [9] Goldstein, B.A., Navar, A.M., Pencina, M.J., *et al.* (2017) Opportunities and Challenges in Developing Risk Prediction Models with Electronic Health Records Data: A Systematic Review. *Journal of the American Medical Informatics Association*, **24**, 198-208. <https://doi.org/10.1093/jamia/ocw042>
- [10] Davis, S.E., Lasko, T.A., Chen, G., *et al.* (2017) Calibration Drift in Regression and Machine Learning Models for Acute Kidney Injury. *Journal of the American Medical Informatics Association*, **24**, 1052-1061. <https://doi.org/10.1093/jamia/ocx030>
- [11] Chen, P.M. and Hemmen, T.M. (2020) Evolving Healthcare Delivery in Neurology during the Coronavirus Disease 2019 (COVID-19) Pandemic. *Frontiers in Neurology*, **11**, 578. <https://doi.org/10.3389/fneur.2020.00578>
- [12] Mann, D.M., Chen, J., Chunara, R., *et al.* (2020) COVID-19 Transforms Health Care through Telemedicine: Evidence from the Field. *Journal of the American Medical Informatics Association*, **27**, 1132-1135. <https://doi.org/10.1093/jamia/ocaa072>
- [13] Vergouwe, Y., Moons, K.G.M. and Steyerberg, E.W. (2010) External Validity of Risk Models: Use of Benchmark Values to Disentangle a Case-Mix Effect from Incorrect Coefficients. *American Journal of Epidemiology*, **172**, 971-980. <https://doi.org/10.1093/aje/kwq223>
- [14] Hickey, G.L., Grant, S.W., Murphy, G.J., *et al.* (2013) Dynamic Trends in Cardiac Surgery: Why the Logistic EuroSCORE Is No Longer Suitable for Contemporary Cardiac Surgery and Implications for Future Risk Models. *European Journal of Cardio-Thoracic Surgery*, **43**, 1146-1152. <https://doi.org/10.1093/ejcts/ezs584>
- [15] Weiss, K., Khoshgoftaar, T.M. and Wang, D.D. (2016) A Survey of Transfer Learning. *Journal of Big Data*, **3**, 1-40. <https://doi.org/10.1186/s40537-016-0043-6>
- [16] Pan, S.J. and Yang, Q. (2009) A Survey on Transfer Learning. *IEEE Transactions on Knowledge and Data Engineering*, **22**, 1345-1359. <https://doi.org/10.1109/TKDE.2009.191>
- [17] Cheplygina, V., Pena, I.P., Pedersen, J.H., *et al.* (2017) Transfer Learning for Multi-center Classification of Chronic Obstructive Pulmonary Disease. *IEEE Journal of Biomedical and Health Informatics*, **22**, 1486-1496. <https://doi.org/10.1109/JBHI.2017.2769800>
- [18] Sikora, R. (2015) A Modified Stacking Ensemble Machine Learning Algorithm Using Genetic Algorithms. In: *Handbook of Research on Organizational Transformations through Big Data Analytics*, IGI Global, Hershey, 43-53.

- <https://doi.org/10.4018/978-1-4666-7272-7.ch004>
- [19] Qualls, L.G., Phillips, T.A., Hammill, B.G., *et al.* (2018) Evaluating Foundational Data Quality in the National Patient-Centered Clinical Research Network (PCORnet®). *EGEMS*, **6**, 3. <https://doi.org/10.5334/egems.199>
- [20] Hripcsak, G., Duke, J.D., Shah, N.H., *et al.* (2015) Observational Health Data Sciences and Informatics (OHDSI): Opportunities for Observational Researchers. *Studies in Health Technology and Informatics*, **216**, 574-578.
- [21] Pardoe, D. and Stone, P. (2010) Boosting for Regression Transfer. *Proceedings of the 27th International Conference on Machine Learning*, Haifa, 21-24 June 2010, 863-870.
- [22] Deo, R.V., Chandra, R. and Sharma, A. (2017) Stacked Transfer Learning for Tropical Cyclone Intensity Prediction.
- [23] Wang, H., Jin, Y., Yang, C., *et al.* (2020) Transfer Stacking from Low- to High-Fidelity: A Surrogate-Assisted Bi-Fidelity Evolutionary Algorithm. *Applied Soft Computing*, **92**, Article ID: 106276. <https://doi.org/10.1016/j.asoc.2020.106276>
- [24] Yi, H.C., You, Z.H., Wang, M.N., *et al.* (2020) RPI-SE: A Stacking Ensemble Learning Framework for ncRNA-Protein Interactions Prediction Using Sequence Information. *BMC Bioinformatics*, **21**, Article No. 60. <https://doi.org/10.1186/s12859-020-3406-0>
- [25] Kleinbaum, D.G., Dietz, K., Gail, M., *et al.* (2002) Logistic Regression. Springer-Verlag, New York, 43-53.
- [26] Ke, G., Meng, Q., Finley, T., *et al.* (2017) Lightgbm: A Highly Efficient Gradient Boosting Decision Tree. *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems*, Long Beach, 4-9 December 2017, 3146-3154.
- [27] Liaw, A. and Wiener, M. (2002) Classification and Regression by Random Forest. *R News*, **2**, 18-22.
- [28] Steyerberg, E.W., Borsboom, G.J.J.M., van Houwelingen, H.C., *et al.* (2004) Validation and Updating of Predictive Logistic Regression Models: A Study on Sample Size and Shrinkage. *Statistics in Medicine*, **23**, 2567-2586. <https://doi.org/10.1002/sim.1844>
- [29] Peterson, L.E. (2009) K-Nearest Neighbor. *Scholarpedia*, **4**, 1883. <https://doi.org/10.4249/scholarpedia.1883>
- [30] Cummings, P. (2009) The Relative Merits of Risk Ratios and Odds Ratios. *Archives of Pediatrics & Adolescent Medicine*, **163**, 438-445. <https://doi.org/10.1001/archpediatrics.2009.31>