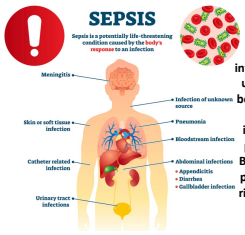
Graphical Abstract

A Temporal GRU with Ensemble Learning for Early Detection of Sepsis (TGEL-EDS)

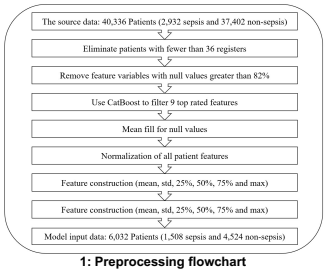
Quanli Pei, Yulong Chen, Jimmy Huang

Purpose

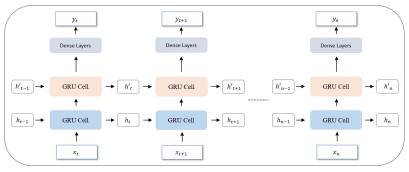


Early prediction of sepsis infection can help intensive care unit(ICU) physicians make the best clinical decisions to reduce their patients' risk of sepsis infection. The purpose of this paper is to develop Stacking-Based Temporal GRU model to predict whether a patient is at risk of sepsis infection within a certain time frame.

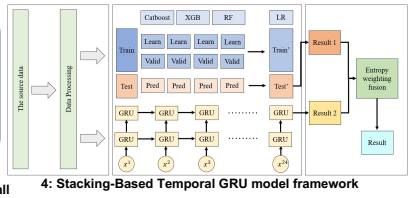
Method



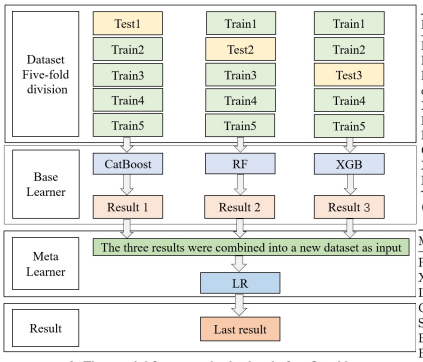
1: Preprocessing flowchart



3: Two-tier GRU model framework (the second level of the overall prediction model)



4: Stacking-Based Temporal GRU model framework



2: The model framework obtained after Stacking ensemble learning method

Result

Mean-filled sample data (Scoring results of Stacking fusion with different base models and LR combinations)

Base model	AUC	F1	Accuracy
RF+Catboost	0.884	0.791	0.793
RF+XGB	0.894	0.761	0.737
RF+LGB	0.861	0.681	0.818
catboost+XGB	0.897	0.681	0.731
XGB+LGB	0.881	0.742	0.815
RF+Catboost+LGB	0.875	0.701	0.762
RF+XGB+LGB	0.886	0.721	0.782
CatBoost+XGB+LGB	0.888	0.736	0.795
XGB+LGB+CatBoost+RF	0.891	0.733	0.796
RF+CatBoost+XGB	**0.898**	**0.751**	**0.831**

Comparison of the evaluation results of various models applied in the early prediction of sepsis

Models	AUC	F1	Accuracy
RF	0.871	0.713	0.773
XGB	0.887	0.744	0.813
LGB	0.833	0.663	0.701
CatBoost	0.874	0.694	0.796
Stacking(best)	0.898	0.742	0.815
Bi-GRU[40]	0.888	0.719	0.771
Bi-LSTM[41]	0.884	0.705	0.742
KANNEDS[10]	0.894	0.731	0.798
TGEL-EDS	**0.908**	**0.752**	**0.819**

A Temporal GRU with Ensemble Learning for Early Detection of Sepsis (TGEL-EDS)

Quanli Pei[a], Yulong Chen[a], Jimmy Huang[a]

[a]*Information Retrieval and Knowledge Management Research Lab, York University, Toronto, On, Canada*

Abstract

Early prediction of sepsis infection can help intensive care unit (ICU) physicians make the best clinical decisions to reduce their patients' risk of sepsis infection. The purpose of this paper is to develop Stacking-Based Temporal GRU model to predict whether a patient is at risk of sepsis infection within a certain time frame. In this study, we used the input data for the stacking ensemble learning method to filter the best combination of base models, and four common machine learning models were chosen for the fusion of base models. Finally, we fused the models obtained by stacking ensemble learning method with the two-layer GRU model by entropy weighting method to obtain our model, and compared it with the mainstream machine learning and deep learning models. Three evaluation metrics were chosen: Area under the curve (AUC), F1 score (F1) and Accuracy score (Accuracy). The experimental results are divided into two parts, the first part is stacking ensemble learning method, we choose the best combination of base model is extreme gradient boosting (XGB), random forest (RF) and categoricalboosting (CatBoost), the metamodel we choose is logistic regression(LR). This combination of model scores 0.898, 0.751 and 0.831 under three evaluation metrics, the best performance among all combinations. Another part is the performance comparison of early sepsis prediction models, comparing the selected mainstream machine learning and deep learning models, our model has the best scores in all three evaluation metrics, in order of 0.908, 0.752 and 0.819.In Conclusion, our model developed has good early predictive performance for sepsis. It has the potential to help ICU physicians to reduce the risk of patients contracting sepsis by early pharmacological treatment and clinical interventions for patients who are likely to contract sepsis.

Keywords: Intensive care unit, Sepsis, Stacking ensemble learning, Deep

learning, Early prediction model

1. Background

Sepsis is a dangerous medical condition caused by infections that can be difficult to identify, treat, and research effectively due to its significant heterogeneity[1, 2]. It affects millions of people worldwide every year and is responsible for over 6 million deaths annually. Mortality rates vary depending on the region but can be sztaggering, such as reaching 41% in Europe and 28.3% in the United States[3]. Unfortunately, sepsis is also the primary cause of in-hospital deaths in the US and results in an estimated 24 billion in costs annually[4].Despite advances in treating sepsis, mortality rates have seen little improvement and continue to rise. However, early prediction, detection, and management of sepsis can significantly improve a patient's chances of survival and help reduce its incidence[5]. Therefore, it is crucial for healthcare providers to remain vigilant in identifying and treating sepsis to minimize its impact on public health[6, 7].

In clinical settings, several hospitals use a scoring system called Sequential Organ Failure Assessment (SOFA) to assess the severity of sepsis. SOFA is recommended by The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3)[2], which provides guidelines for the diagnosis and management of sepsis.Another scoring system proposed for identifying patients with suspected infection and high risk of death is QuickSOFA. QuickSOFA is a rapid bedside score that includes three variables: altered mental status, low blood pressure, and high respiratory rate.The sepsis-3 requires healthcare providers to monitor a patient's vital signs (VS), such as heart rate, oxygen saturation, arterial pressure, and respiratory rate, along with laboratory test results, to calculate the sepsis score. These vital signs are considered characteristic variables that play a significant role in diagnosing and managing sepsis.Patient monitoring is essential for the prevention and treatment of sepsis, and the use of scales such as SOFA and QuickSOFA can aid in the clinical assessment and identification of high-risk cases. In addition, monitoring patient variables also promotes research on to improve its prevention and treatment strategies[8, 9].

In recent years, there has been a significant interest in predicting sepsis early using patient's vital sign (VS) data and other characteristic data. In this context, Marcio and his colleagues[10] have developed a kinematic model

that calculates the location of sepsis spatially to obtain kinematic features (KF) of the patient. VS and KF are then used together as input variables for neural network model prediction. Their findings reveal that when VS and KF are used together, the accuracy of sepsis prediction is higher compared to using VS data alone. This suggests that the incorporation of the kinematic features provides valuable additional information for early prediction of sepsis in patients.

Manaktala and colleagues[11] have created a unique system that utilizes change management, electronic monitoring and algorithms in combination. This novel approach is designed to offer accurate decision support to users. The system has demonstrated excellent sensitivity and specificity, suggesting it's highly effective.

Sidney et al.[12] conducted a study on pediatric patients at a medical center using electronic health record data to develop machine learning algorithms for predicting sepsis severity. their findings showed that the algorithm was effective in detecting and preventing sepsis severity, with significant effects due to two specific indicators - PELOD-2 and SiRS. Guilan[13] conducted another study comparing the performance of different prediction models for sepsis using machine learning algorithms such as LASSO, RF, GBM, and traditional methods like LR. They found that the machine learning models outperformed the traditional simplified acute physiology score (SAPS) model, suggesting that these advanced models could help improve the prognoses of sepsis patients in icus.

Xiang et al.[14] presented an ML model that is time-phased for real-time sepsis prediction in ICU settings. On the other hand, Akram et al.[15] employed Support Vector Machines (SVM) to segregate patients with sepsis across five hospitals. Furthermore, it was found that there were significant differences in physiological indicators in patients with sepsis over time. Based on these insights, artificial intelligence models can be developed to predict the occurrence of sepsis in patients using a time series approach.

Lucas et al.[16] and Michael et al.[17] conducted research on the utilization of machine learning methods in forecasting sepsis. they employed the GRADE and PRISMA guidelines to derive an extensive collection of literary models.the outcome of their studies highlighted a wide-spread adoption of machine learning technology for early prediction of sepsis. machine learning-based models were observed to be highly accurate in predicting episodes of sepsis, thus holding promise for clinical application.

Hye et al.[18] achieved better performance by implementing Long Short-

Term Memory (LSTM) deep learning models, enabling temporal pattern learning from time-series data. Dongdong et al.[19] proposed a deep learning model with temporal encoding to predict sepsis, utilizing patients' electronic health records (EHR).The deep learning models deployed in these studies were found to provide highly accurate predictions, and the transparency and clinical interpretability of their outcomes are considered critical for medical applications. The conclusion drawn is that deep learning models can substantially improve the accuracy of sepsis prediction, particularly when time-series data is analyzed or when EHR is used. These models offer greater transparency and clinical interpretability, allowing clinicians to readily interpret and apply the results.

Tunc et al.[20] utilized a regression-based analysis to review data on seven key vital signs of patients in ICU beds. Following this initial step, the researchers employed the use of a specific algorithm known as the Deep SOFA (Sequential Organ Failure Assessment) Sepsis Prediction Algorithm (DSPa). The purpose behind utilizing the DSPa was to predict Sofa Scores in patients who were diagnosed with sepsis. The rationale for making such a prediction was to accurately gauge the severity level of organ failure that is present within patients who have been diagnosed with sepsis.

Gavin et al.[21] provided a statistically significant improvement to machine learning modeling applications by proposing a new deep learning application for effective prediction of sepsis six hours prior to onset, including a boosted cascading training methodology and adjustable margin hinge loss function.

It is easy to see that machine learning and deep learning are playing a role in the medical field that cannot be ignored[22–27].They have shown superior performance in predicting sepsis compared to traditional scoring criteria. utilizing methods to process time series and vital sign variables of patients can further improve the performance of these models. While deep learning models may not perform well on small data sets, machine learning models may not be as proficient at learning variations over time series as deep learning models. to tackle this problem, the authors developed a unique sepsis prediction model. they used a stacking-based temporal GRU approach that integrates both machine learning and deep learning models at different levels. At the first level, the model uses various algorithms such as RF, CatBoost and XGBoost fused by an integrated learning method. at the second level, a double-layer GRU is used. the two levels are then blended by an entropy weighting method to obtain the final sepsis prediction model. We validated our model and found

that our model outperforms other models that solely rely on machine learning or deep learning. in addition, our model also had better results than another neural network model that was introduced with kinematic features (KF)[10]. All in all, We have created a novel sepsis prediction model that merges both machine learning and deep learning approaches to handle different types of data sets effectively. our model provides higher accuracy compared to other models and has the potential to improve sepsis diagnosis and treatment.

2. Methods

2.1 Data source

We used data from The PhysioNet/Computing in Cardiology Challenge 2019 as our original dataset[28–30]. The dataset source data for 40,336 patients, saved separately in individual text files, with each line in the individual text file representing the status of each variable for the patient at the current hour. Available patient variables include demographics, vital signs (VS), and laboratory values, for a total of more than 40 characteristic variables, with each patient having a different time series length.

2.2 Data preprocessing

Since each patient in the original dataset may or may not be ill in the entire time series, we first defined the SepsisLabel variable (For sepsis patients: 1, For non-sepsis patients: 0) to be non-sepsis patients if all of them are 0 in the entire time series, and sepsis patients whenever 1 occurs, and redefined a Label variable (For sepsis patients: 1, For non-sepsis patients: 0) indicates whether the patient is a patient in the whole sequence. The number of patients was obtained as 2932 and the number of non-patients as 37404. Secondly, we only kept the data of patients with more than 36 hours of recording and removed the characteristic variables with null values greater than 82% in these patient data. The final data showed a total number of patients of 24,071, of which the number of patients was 15,08 and the number of non-patients was 22,563. Considering that even after removing the feature variables with missing values greater than 82%, there may still exist variables that are not very helpful for model prediction, we used the CatBoost model[31] to filter the remaining variables and took the variables with the top 9 scores as predictor variables, and the variable names and their values are introduced as shown in Table 1.

Table 1: Predictor variables remaining after feature variable screening

Varible	Unit	Positive	Negative
Number	#	1508	3016
MAP	mm Hg	80.64 ± 59.46	82.23 ± 62.23
Temp	°C	36.98 ± 4.41	36.97 ± 16.07
O2Sat	%	97.46 ± 66.46	97.25 ± 77.25
SBP	mm Hg	121.79 ± 147.20	123.35 ± 174.64
HR	beats/min	86.84 ± 124.15	84.62 ± 195.37
DBP	mm Hg	62.08 ± 185.91	63.67 ± 234.32
Resp	beats/min	19.13 ± 51.86	18.60 ± 81.39
HospAdmTime	h	-75.48 ± 76.27	-53.52 ± 5313.33
Age	years	62.35 ± 37.64	62.39 ± 37.60

After the feature filtering was performed, the null values under each variable obtained from the filtering were filled by means of mean filling. The mean value was selected as the mean value of all patient data, which is more effective compared to using the mean value of individual patient data, and the data were normalized after mean filling, and the sample data obtained are shown in Table 2.

Table 2: Mean-filled sample data (selected from a single patient for three consecutive hours)

MAP	Temp	O2Sat	SBP	HR	DBP	Resp	HospAdmTime	Age
0.02	0.70	1.00	0.44	0.27	0.23	0.30	0.99	0.51
0.22	0.70	0.97	0.32	0.24	0.26	0.23	0.99	0.51
0.24	0.70	1.00	0.36	0.28	0.28	0.29	0.99	0.51

Next, we used feature construction to expand the dataset by first obtaining the mean (mean), standard deviation (std), 25th percentile, 50th percentile, 75th percentile, and maximum (max) corresponding to each patient's nine feature variables, which were added to the dataset as new features, and the sample data obtained after feature construction are shown in Table 3.

Finally, to ensure a uniform distribution of positive and negative samples (positive samples are patients and negative samples are non-patients), we randomly selected 4524 from 22,563 non-sepsis patients as negative samples

Table 3: Sample data obtained after feature construction (three patients at hour 1 after MAP variable expansion)

MAP	_mean	_std	_25%	_50%	_75%	_max
0.02	0.36	0.36	0.33	0.37	0.38	0.17
0.22	0.34	0.24	0.35	0.34	0.30	0.15
0.24	0.58	0.38	0.52	0.60	0.57	0.28

to achieve a patient to non-patient ratio of 1:3, and the entire data preprocessing process is shown in Figure 1.

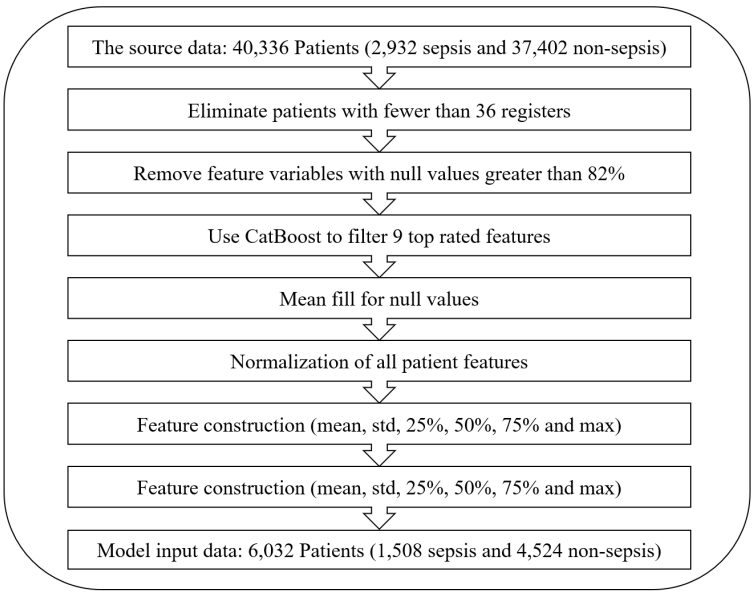


Figure 1: Preprocessing flowchart

2.3 Stacking method

The stacking ensemble learning method mixes many machine learning model types through some sort of fusion, utilizing the benefits of various machine learning models to analyze data from various angles. As a result, the first layer's base learner must choose a variety of models in addition to high-performing ones. To repair the faults of the base learner in the first layer and get the best predicting, the meta-learner in the second layer should

choose models with strong generalization potential. CatBoost is an innovative algorithm for processing classification features that solves the problem of prediction bias that can result from gradient enhancement algorithms[31]. The random forest (RF) model, which can successfully avoid overfitting and is immune to interference and parallel processing, is a model integrated by bagging[32, 33]; Extreme gradient boosting (XGB) integrated by boosting with a regularization term to prevent overfitting, with high computational efficiency and multi-threaded operation[34]; The lightgbm model (LGB) is a Histogram-based decision number algorithm that allows the information gain from large gradient samples to be amplified while retaining large gradient samples[35]; Logistic regression (LR) is a widely used classification algorithm that is widely accepted in the field of statistics[36]. In summary, we consider the combination of CatBoost, RF, XGB, and LGB in different combinations as the base learner of the Stacking ensemble learning method, and LR models as a meta-learning model, using a five-fold cross-validation approach for Stacking integration, and the integrated model is called the first-level model of the overall prediction model. We created trials to individually forecast the various combinations and assessed the prediction outcomes (the evaluation metrics we used were Accuracy, F1 and AUC in Evaluation). The finally selected optimal base model combination is RF, CatBoost and XGB, the model framework is shown in Figure 2.

2.4 Two-layer GRU model

The Gated Recurrent Unit (GRU) was proposed by Junyoung et al, and used in sequence studies[37]. It shares the same gating unit as the LSTM to minimize information loss during the propagation of long sequences (long-term) through the gating unit, but it lacks a recurrent memory unit, making it less complex than the LSTM. In a different research, Cho demonstrated that GRU outperforms LSTM with a sufficient dataset and is quicker than LSTM[38].In this paper, we have used the two-layer GRU model as a second level for the overall prediction model and we can see how the two-layer GRU model works when we input x_t in Figure 3.

2.5 Stacking-Based Temporal GRU model

The model obtained by the Stacking integration method after filtering the base model is used as the first level model of the sepsis early prediction model, and the two-layer GRU model is used as the second level model of the sepsis prediction model, and the two levels of models are combined to form

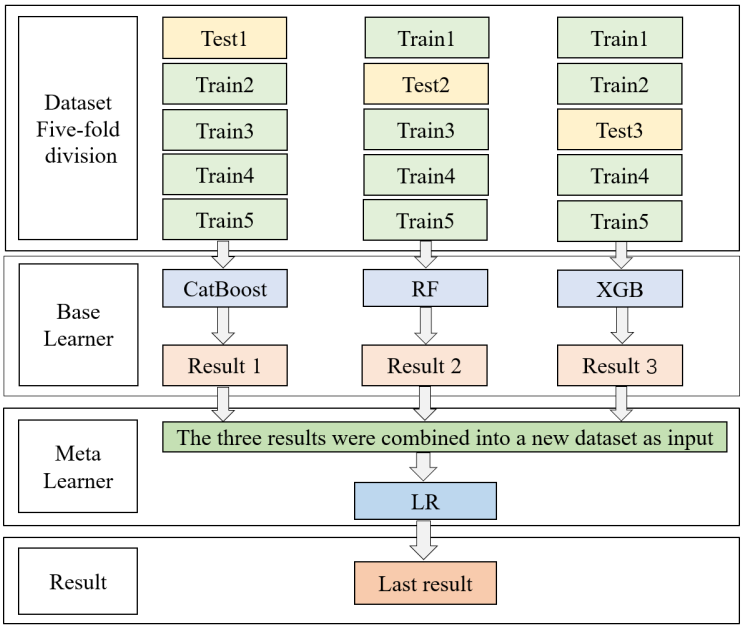


Figure 2: The model framework obtained after Stacking ensemble learning method

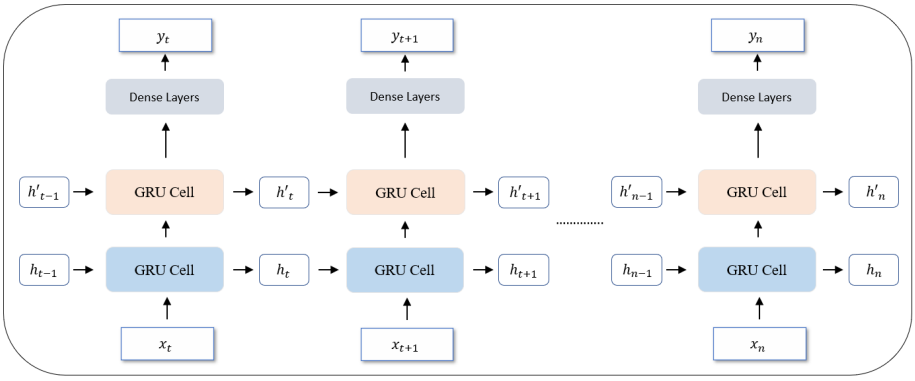


Figure 3: Two-tier GRU model framework (the second level of the overall prediction model)

our overall prediction model. The overall model framework is shown in Figure 4. The weights for the entropy weighting fusion are determined according to the entropy weighting method[39]. The mathematical expressions can be

denoted as

$$x'_{ij} = \frac{|x_{ij} - min(x_j)|}{max(x_j) - min(x_j)} \tag{1}$$

$$e_j = -\frac{1}{ln(n)} \sum_{i=1}^{n} \frac{x'_{ij}}{\sum_{i=1}^{n} x'_{ij}} ln(\frac{x'_{ij}}{\sum_{i=1}^{m} x'_{ij}}), j = 1, ..., m \tag{2}$$

$$\omega_j = \frac{1 - e_j}{\sum_{j=1}^{m}(1 - e_j)}, j = 1, ..., m \tag{3}$$

$$s_j = \frac{\sum_{j=1}^{m} \omega_j x_{ij}}{\sum_{j=1}^{m} \omega_j}, i = 1, ...n \tag{4}$$

x_{ij} refers to the ith prediction result of the jth model, and x'_{ij} is the result obtained after normalizing x_{ij}. i ranges from 1 n and j ranges from 1 $m.e_j$ represents the information entropy of the jth model, ω_j represents the weight of the jth model, and s_j represents the result obtained after entropy weighting fusion.

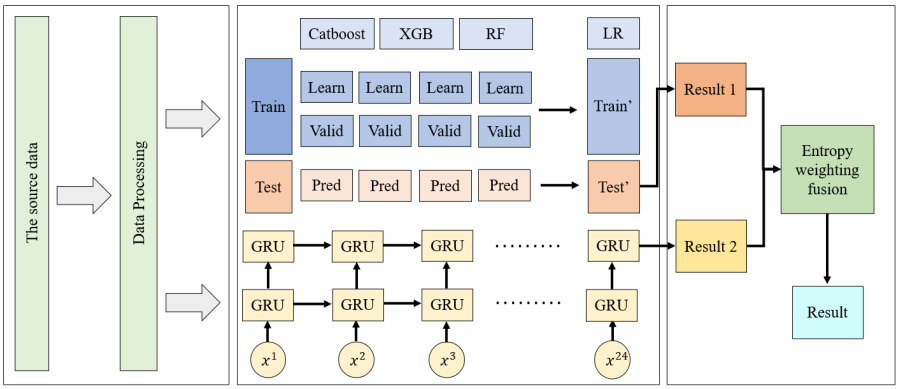


Figure 4: Stacking-Based Temporal GRU model framework (Train: original training data, Test: original test data, Learn: training set divided in five-fold cross-validation, Valid: validation set divided in five-fold cross-validation, Pred: prediction result of the training set, Train': new training data obtained by combining the prediction result of the base model and the original training data)

2.6 Evaluation

This article selects Accuracy,F1 and AUC to measure the performance of the different models. They are denoted by

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{5}$$

$$F_1 = 2 \cdot \frac{precision \cdot recall}{precision + recall} \tag{6}$$

Where: TP means predicted 1, actual 1, predicted correctly. FP means predicted 1, actual 0, predicted incorrectly. FN means predicted 0, actual 1, predicted incorrectly. TN means predicted 0, actual 0, predicted correctly. $precision = TP/(TP + FP)$, $recall = TP/(TP + FN)$. $FPR = TP/(TP + FN)$, $TPR = TP/(FP + TN)$. AUC refers to the area below the ROC curve consisting of FPR as the $x - axis$ and TPR as the $y - axis$.

3. Results

3.1 Filter the base model in Stacking ensemble method

We use the four machine learning models XGB, LGB, RF, and CatBoost as the base models for the Stacking ensemble method, and then take LR as the meta model. To determine the ideal hyperparameter values for each of the four machine learning models, we employed the grid parameter search technique offered by Scikit-Learn. We can see from the findings in Table1 that the base model of XGB, CatBoost and RF combined with LR as the meta-model beats the other base and meta model combinations. As a result, we decided to employ the model created utilizing the aforementioned integration techniques as the first level model for our overall prediction model.

Table 4: Mean-filled sample data (Scoring results of Stacking fusion with different base models and LR combinations)

Base model	AUC	F1	Accuracy
RF+Catboost	0.884	0.791	0.793
RF+XGB	0.894	0.761	0.737
RF+LGB	0.861	0.681	0.818
Catboost+XGB	0.897	0.681	0.731
XGB+LGB	0.881	0.742	0.815
RF+Catboost+LGB	0.875	0.701	0.762
RF+XGB+LGB	0.886	0.721	0.782
CatBoost+XGB+LGB	0.888	0.736	0.795
XGB+LGB+CatBoost+RF	0.891	0.733	0.796
RF+CatBoost+XGB	**0.898**	**0.751**	**0.831**

3.2 Performance of different models on the sepsis early prediction problem

We compare the evaluation results of the Stacking-Based Temporal GRU model constructed with XGB, LGB, the first-level model constructed in this paper through the Stacking ensemble method, Bi-GRU and Bi-LSTM for the sepsis early prediction problem. It is important to note that the two-layer LSTM and two-layer GRU models we selected are both two-layer unidirectional models since, according to our tests, they outperform all other two-layer models. In addition, the parameter values for the XGB model, the LGB model, and the crucial parameter settings for the GRU and LSTM are displayed in Table 5.

Table 5: Mean-filled sample data (Important hyperparameter settings for GRU and LSTM)

Hyperparameter	LSTM	GRU
Sequence Length	36	36
Epochs	100	100
Epochs	100	100

The final evaluation findings for these various models on the sepsis prediction are provided in Table 6. In terms of AUC, Accuracy and F1 scores, our Stacking-Based Temporal GRU model performs better than the other models, notably in the F1 and Accuracy assessment metrics, which are noticeably higher than the other models, reaching 0.752 and 0.819 respectively. This shows that our model performs well when used to predict sepsis.

4. Discussion

Sepsis patient data were first extracted from the PhysioNet Computing in Cardiology Challenge 2019 dataset in this study. Then, We apply a series of pre-processing operations to the source data to obtain the input data for the early prediction model. Finally, we propose the Stacking-Based Temporal GRU model and compare the performance with other mainstream models under three evaluation metrics.

In the data preprocessing section, firstly, in consideration of the fact that different time series measured for each patient may have an effect on the effectiveness of the deep learning model, we selected the first 36 hours of data for each patient and removed those patients with less than 36 hours of

Table 6: Comparison of the evaluation results of various models applied in the early prediction of sepsis

Models	AUC	F1	Accuracy
RF	0.871	0.713	0.773
XGB	0.887	0.744	0.813
LGB	0.833	0.663	0.701
CatBoost	0.874	0.694	0.796
Stacking(best)	0.898	0.742	0.815
Bi-GRU[40]	0.888	0.719	0.771
Bi-LSTM[41]	0.884	0.705	0.742
KANNEDS[10]	0.894	0.731	0.798
TGEL-EDS	**0.908**	**0.752**	**0.819**

time series length. Secondly, we removed the feature variables with null values accounting for more than 82% since too many null values in the feature variables of some patients would affect the classification effect of the model. Thirdly, we performed feature variable screening by the Catboost model, and finally kept the top 9 feature variables with the highest model scores to avoid the interference of unnecessary variables on the model performance, and the null values are filled using the mean value. Besides, to improve the model effect further, we performed feature construction and data normalization operations. Specifically, feature construction is variable expansion using the mean, standard deviation, 25% quantile, 50% quantile, 75% quantile, and maximum value of each feature variable. Finally, in order to make the positive and negative samples more balanced, we randomly selected some samples from the negative samples to make the proportion of positive and negative samples 1:3.

The stacking ensemble learning method[42] tends to perform better by using different machine learning models to learn different features of the overall data, and after fusion. We used different combinations of LGB, XGB, RF and CatBoost models as the base model for the stacking ensemble learning method. After a series of experiments we finally selected a combination of base models suitable for early prediction of sepsis, which are RF, XGB and CatBoost. Stacking fusion of RF, XGB and CatBoost as the base model and LR as the meta-model constitutes the first level model of our early sepsis prediction model. Considering that the prediction performance of the GRU

model may be higher than that of the LSTM model under the larger data size, we used the two-layer GRU model as the second-level model of our early sepsis prediction model, and the final experimental results also confirmed the suitability of the two-layer GRU model as the second-level model. In the end, the two-level model was fused by entropy weighting method to form our early sepsis prediction model. After constructing the Stacking-Based Temporal GRU model, we conducted extensive comparison experiments on our processed dataset using RF, XGB, CatBoost, Bi-GRU, Bi-LSTM model and KANNEDS methods. Our early sepsis prediction model was found to perform significantly better than other models in three metrics, Accuracy, F1 and AUC, and proved the accuracy and completeness of our model.

The study focused on developing a model for predicting sepsis using clinical data from patients in certain medical centers. By utilizing this data, the researchers aimed to create a Stacking-Based Temporal GRU model of anticipating the onset of sepsis in patients. However, it is essential to note that the dataset used had limitations as it only pertained to some medical center patients. This could affect the model's validity when applied to other datasets or domains, which may require further evaluation.During data preprocessing, some characteristic variables of the patients were discarded, potentially impacting the results. The model was also limited in its ability to predict sepsis, as it was only based on the baseline data from the first 36 hours of a patient's stay in the ICU rather than dynamically considering data throughout the patient's ICU experience.

5. Conclusion

This study contributes to clinical areas with a Stacking-Based Temporal GRU model for early detection of sepsis in ICU. The prediction model has the ability to help ICU physicians predict which patients are likely to contract sepsis so that they can take medication and clinical interventions to reduce their patients' risk of infection. Since all variables used for model construction are regularly collected by the information systems of three separate hospital ICU in this study, it is feasible to ensure that we use the Stacking-Based Temporal GRU model to help ICU physicians with clinical interventions. We plan to ensemble the Stacking-Based Temporal GRU model into existing ICU medical systems for real-time prediction of whether a patient will contract sepsis to provide clinical decision support. Our future research will focus on evaluating the general applicability of our early sepsis prediction model,

conducting experiments on data from additional medical centers to confirm the generalizability and validity of the model.

6. Acknowledgements

We would like to express our sincere gratitude to the Natural Sciences and Engineering Research Council (NSERC) of Canada for the support of this work.

References

[1] Gyawali, Ramakrishna, Sepsis: The evolution in definition, pathophysiology, and management, SAGE Open Medicine 7 (2019) 205031211983504. doi:10.1177/2050312119835043.

[2] M. Singer, C. Deutschman, C. Seymour, M. Shankar-Hari, D. Annane, M. Bauer, R. Bellomo, G. Bernard, J.-D. Chiche, C. Coopersmith, R. Hotchkiss, M. Levy, J. Marshall, G. Martin, S. Opal, G. Rubenfeld, T. Poll, J.-L. Vincent, D. Angus, The third international consensus definitions for sepsis and septic shock (sepsis-3), JAMA 315 (2016) 801. doi:10.1001/jama.2016.0287.

[3] M. Levy, A. Artigas, G. Phillips, A. Rhodes, R. Beale, T. Osborn, J.-L. Vincent, S. Townsend, S. Lemeshow, R. dellinger, Outcomes of the surviving sepsis campaign in intensive care units in the usa and europe: A prospective cohort study, The Lancet infectious diseases 12 (2012). doi:10.1016/S1473-3099(12)70239-6.

[4] K. Rudd, S. Johnson, K. Agesa, K. Shackelford, D. Tsoi, D. Kievlan, D. Colombara, K. Ikuta, N. Kissoon, S. Finfer, C. Fleischmann, F. Machado, K. Reinhart, K. Rowan, C. Seymour, S. Watson, E. West, M. D. F. Marinho de Souza, S. Hay, M. Naghavi, Global, regional, and national sepsis incidence and mortality, 1990-2017: analysis for the global burden of disease study, The Lancet 395 (2020) 200–211. doi:10.1016/s0140-6736(19)32989-7.

[5] K.-M. Kaukonen, M. Bailey, S. Suzuki, D. Pilcher, R. Bellomo, Mortality related to severe sepsis and septic shock among critically iii patients in australia and new zealand, 2000-2012, JAMA : the journal of the American Medical Association 311 (2014). doi:10.1001/jama.2014.2637.

[6] R. Zheng, Y. Zhang, Z. Rong, W. Huang, X. Fu, Surviving sepsis campaign: international guidelines for management of sepsis and septic shock 2021, interpretation and expectation, Zhonghua wei zhong bing ji jiu yi xue 33 (2021) 1159–1164. doi:10.3760/cma.j.cn121430-20211009-01442.

[7] G. Westphal, A. Gonçalves, M. Caldeira-Filho, E. Silva, R. Salomão, W. Bernardo, F. Machado, Guidelines for treatment of severe sepsis/septic shock –tissue perfusion assessment, RBTI 23 (2011) 6–12. doi:10.1590/S0103-507X2011000100003.

[8] M. Cruz, C. Cavalcante, S. Barretto, Using opc and hl7 standards to incorporate an industrial big data historian in a health it environment, Journal of Medical Systems 42 (2018) 122. doi:10.1007/s10916-018-0979-5.

[9] J. Rousselot, J.-D. Decotignie, Wireless communication systems for continuous multiparameter health monitoring, 2009, pp. 480 – 484. doi:10.1109/ICUWB.2009.5288747.

[10] M. Cruz, N. Ono, M. Huang, A. Amin, S. Kanaya, C. Cavalcante, Kinematics approach with neural networks for early detection of sepsis (kanneds), BMC Medical Informatics and Decision Making 21 (2021). doi:10.1186/s12911-021-01529-3.

[11] S. Manaktala, S. Claypool, Evaluating the impact of a computerized surveillance algorithm and decision support system on sepsis mortality, Journal of the American Medical Informatics Association 24 (2016) ocw056. doi:10.1093/jamia/ocw056.

[12] S. Le, J. Hoffman, C. Barton, J. Fitzgerald, A. Allen, E. Pellegrini, J. Calvert, R. Das, Pediatric severe sepsis prediction using machine learning, Frontiers in Pediatrics 7 (2019). doi:10.3389/fped.2019.00413.

[13] G. Kong, K. Lin, Y. Hu, Using machine learning methods to predict in-hospital mortality of sepsis patients in the icu, BMC medical informatics and decision making 20 (2020) 251. doi:10.1186/s12911-020-01271-2.

[14] X. Li, X. Xu, F. Xie, X. Xu, Y. Sun, X. Liu, X. Jia, Y. Kang, L. Xie, F. Wang, et al., A time-phased machine learning model for real-time

prediction of sepsis in critical care, Critical Care Medicine 48 (2020) e884–e888.

[15] A. Mohammed, F. Wyk, L. Chinthala, A. Khojandi, R. Davis, C. Coopersmith, R. Kamaleswaran, Temporal differential expression of physiomarkers predicts sepsis in critically ill adults, Shock (Augusta, Ga.) Publish Ahead of Print (2020). doi:10.1097/SHK.0000000000001670.

[16] L. M. Fleuren, T. L. Klausch, C. L. Zwager, L. J. Schoonmade, T. Guo, L. F. Roggeveen, E. L. Swart, A. R. Girbes, P. Thoral, A. Ercole, et al., Machine learning for the prediction of sepsis: a systematic review and meta-analysis of diagnostic test accuracy, Intensive care medicine 46 (2020) 383–400.

[17] M. Moor, B. Rieck, M. Horn, C. R. Jutzeler, K. Borgwardt, Early prediction of sepsis in the icu using machine learning: a systematic review, Frontiers in medicine 8 (2021) 607952.

[18] H. J. Kam, H. Y. Kim, Learning representations for the early detection of sepsis with deep neural networks, Computers in biology and medicine 89 (2017) 248–255.

[19] D. Zhang, C. Yin, K. Hunold Buck, X. Jiang, J. Caterino, P. Zhang, An interpretable deep learning model for early prediction of sepsis in the emergency department, 2020. doi:10.1101/2020.09.21.20198895.

[20] B. Y. Al-Mualemi, L. Lu, A deep learning-based sepsis estimation scheme, Ieee Access 9 (2020) 5442–5452.

[21] G. Tsang, X. Xie, Deep learning based sepsis intervention: the modelling and prediction of severe sepsis onset, in: 2020 25th international conference on pattern recognition (ICPR), IEEE, 2021, pp. 8671–8678.

[22] S. P. Shashikumar, C. S. Josef, A. Sharma, S. Nemati, Deepaise – an interpretable and recurrent neural survival model for early prediction of sepsis, Artificial Intelligence in Medicine 113 (2021) 102036. URL: https://www.sciencedirect.com/science/article/pii/S0933365721000294. doi:https://doi.org/10.1016/j.artmed.2021.102036.

[23] A. Alexiadis, Deep multiphysics: Coupling discrete multiphysics with machine learning to attain self-learning in-silico models replicating human physiology, Artificial intelligence in medicine 98 (2019) 27–34.

[24] I. Sharifi, S. Goudarzi, M. B. Khodabakhshi, A novel dynamical approach in continuous cuffless blood pressure estimation based on ecg and ppg signals, Artificial intelligence in medicine 97 (2019) 143–151.

[25] M. Halper, H. Gu, Y. Perl, C. Ochs, Abstraction networks for terminologies: supporting management of "big knowledge", Artificial intelligence in medicine 64 (2015) 1–16.

[26] Y. Kazemi, S. A. Mirroshandel, A novel method for predicting kidney stone type using ensemble learning, Artificial intelligence in medicine 84 (2018) 117–126.

[27] H. Ayadi, M. Torjmen-Khemakhem, J. Huang, Term dependency extraction using rule-based bayesian network for medical image retrieval, Artificial Intelligence in Medicine (2023) 102551.

[28] M. A. Reyna, C. Josef, S. Seyedi, R. Jeter, S. P. Shashikumar, M. B. Westover, A. Sharma, S. Nemati, G. D. Clifford, Early prediction of sepsis from clinical data: the physionet/computing in cardiology challenge 2019, in: 2019 Computing in Cardiology (CinC), IEEE, 2019, pp. Page–1.

[29] J. C. J. R. S. S. M. B. W. M. B. S. A. N. S. . C. G. D. Reyna, M., Early prediction of sepsis from clinical data: The physionet/computing in cardiology challenge 2019 (version 1.0.0), PhysioNet (2019). doi:10.13026/v64v-d857.

[30] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, H. E. Stanley, Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals, circulation 101 (2000) e215–e220.

[31] L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, A. Gulin, Catboost: unbiased boosting with categorical features, Advances in neural information processing systems 31 (2018).

[32] L. Breiman, Bagging predictors, Machine learning 24 (1996) 123–140.

[33] L. Breiman, Random forests, Machine learning 45 (2001) 5–32.

[34] T. Chen, C. Guestrin, Xgboost: A scalable tree boosting system, in: Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, 2016, pp. 785–794.

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

[36] J. S. Cramer, The origins of logistic regression (2002).

[37] J. Chung, C. Gulcehre, K. Cho, Y. Bengio, Empirical evaluation of gated recurrent neural networks on sequence modeling, arXiv preprint arXiv:1412.3555 (2014).

[38] K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, Y. Bengio, Learning phrase representations using rnn encoder-decoder for statistical machine translation, arXiv preprint arXiv:1406.1078 (2014).

[39] J. Zhao, G. Ji, Y. Tian, Y. Chen, Z. Wang, Environmental vulnerability assessment for mainland china based on entropy method, Ecological Indicators 91 (2018) 410–422.

[40] R. Zhou, X. Hu, B. Yuan, Q. Xu, Lithology classification system for well logging based on bidirectional gated recurrent unit, in: 2021 4th International Conference on Artificial Intelligence and Big Data (ICAIBD), IEEE, 2021, pp. 599–603.

[41] Z. Huang, W. Xu, K. Yu, Bidirectional lstm-crf models for sequence tagging, arXiv preprint arXiv:1508.01991 (2015).

[42] X. Dong, Z. Yu, W. Cao, Y. Shi, Q. Ma, A survey on ensemble learning, Frontiers of Computer Science 14 (2020) 241–258.