

Research Article

Early Prediction of Organ Failures in Patients with Acute Pancreatitis Using Text Mining

Jiawei Luo,¹ Lan Lan^{ID},¹ Dujiang Yang,² Shixin Huang,³ Mengjiao Li,¹ Jin Yin,¹ Juan Xiao^{ID},⁴ and Xiaobo Zhou⁵

¹West China Biomedical Big Data Center, West China Hospital/West China School of Medicine, Sichuan University, Chengdu 610041, China

²Department of Gastrointestinal Surgery, West China Hospital/West China School of Medicine, Sichuan University, Chengdu 610041, China

³School of Communication and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, China

⁴Department of Cardiovascular Medicine, The Affiliated Hospital of Southwest Medical University, Luzhou 646000, China

⁵School of Biomedical Informatics, University of Texas Health Science Center at Houston, Houston 77030, USA

Correspondence should be addressed to Lan Lan; lanl@scu.edu.cn and Juan Xiao; 879214513@qq.com

Received 8 December 2020; Revised 12 January 2021; Accepted 27 March 2021; Published 12 May 2021

Academic Editor: Pengwei Wang

Copyright © 2021 Jiawei Luo et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

It is of great significance to establish an assessment model for organ failures in the early stage of admission in acute pancreatitis (AP). And the clinical notes are underutilized. To predict organ failures for AP patients using early clinical notes in hospital, early text features obtained from the pretrained Chinese Bidirectional Encoder Representations from Transformers model and attention-based LSTM were combined with early structured features (laboratory tests, vital signs, and demographic characteristics) to predict organ failures (respiratory, cardiovascular, and renal) in 12,748 AP inpatients in West China Hospital, Sichuan University, from 2008 to 2018. The text plus structured features fusion model was used to predict organ failures, compared to the baseline model with only structured features. The performance of the model with text features added is superior to the model that only includes structured features.

1. Introduction

Organ failure is a serious complication of patients with acute pancreatitis (AP). Acute renal failure is one of the most common causes of death in patients with severe AP [1]. Acute renal failure in the setting of AP has been shown to have a 10-fold increase in mortality in a study of 563 patients [2]. Approximately, one-third of severe pancreatitis patients develop acute lung injury and acute respiratory distress syndrome that account for 60% of all deaths within the first week [3]. Patients with AP associated with congestive heart failure (CHF) have significantly higher mortality in comparison with those without CHF [4]. Therefore, it is of great significance to establish an assessment model for organ failures in the early stage of admission in AP.

Three organ systems should be assessed to define organ failure: respiratory, cardiovascular, and renal based on the classification of AP—2012 [5]. In this study, organ failures included respiratory failure, circulatory failure, and renal failure. Previous studies [6, 7] used a single indicator or a linear model of several indicators for predicting the risk of organ failure within 12 hours of admission. Their sample size was small, and their generalization ability was relatively weak. Additionally, their methods did not make full use of the patient's clinical manifestations and medical history.

The electronic medical record (EMR) of the hospital information system contains structured and unstructured data, which does not include bio-omics data [8–10]. Integration with clinical narrative would be highly useful because rich information is buried in unstructured text. It is of

great value for the mining and utilization of unstructured data. For clinical notes (an example in West China Hospital, Sichuan University, shown in Figure S1), the clinician and patient's history of present illness (HPI) may contain information that is not extracted from structured data, such as the onset of AP, recurrence time of AP symptoms, the description of early symptoms before admission, and clinician' speculations about certain complications or conditions.

Therefore, we will make full use of clinical notes, not just extracting some of them manually as features. We will use AP patient's HPI, early laboratory tests, early vital signs, and demographic data to assess the risk of organ failures between 72 hours after admission and discharge. Pretrained Chinese Bidirectional Encoder Representations from Transformers (BERT) model [11] will be used to transform the sentence into a feature matrix, and the characteristics of the sentence will be captured by an attention-based long short-term memory (LSTM) [12] so that a unified representation of the text information can be obtained for all patients.

2. Materials and Methods

2.1. Comparison to Recent Works on Medical Records. Nguyen et al. [13] developed a deep neural net based on convolutional neural networks named Deepr for representing EMR and predicting unplanned readmission. Deepr sequences the EMR into a "sentence," or equivalently, a sequence of "words." Each word represents a discrete object or event such as diagnosis, procedure, or any derived object such as time-interval or hospital transfer. Their "word" is different from our "word." Our word is extracted from the text sentence of clinical notes, rather than combining discrete objects together.

Soysal et al. [14] developed a toolkit CLAMP for building customized clinical natural language processing pipelines. Although the CLAMP default pipeline achieved good performance on named entity recognition and concept encoding, it has limitations in handling other languages such as Chinese.

Krishnan and Kamath [15] used unstructured physician notes modeled using hybrid word embeddings to generate quality features which were used to train and build a deep neural network model to predict disease groups based on ICD code. Because there are too many types of diseases involved, it is not possible to obtain important characteristics specific to a certain disease.

Yuwono et al. [16] proposed a novel neural network named convolutional residual recurrent neural network that learns to diagnose acute appendicitis based on doctors' free-text emergency department notes and optional real-valued features (from the structured fields) without any feature engineering. The effectiveness of their method in the diagnosis or prediction of other diseases or adverse outcomes needs to be further explored.

Akbilgic et al. [17] used a text mining approach (hybrid model) on preoperative notes to obtain a text-based risk score to predict death within 30 days of surgery in children, which significantly improved the performance of C-statistic from 0.76 to 0.92 when text-based risk scores were included in addition to structured data.

Therefore, this study fully considers the disease characteristics of AP, collects and analyzes relevant features as much as possible, and uses the attention mechanism to output important words, which leaves room for the treatment of AP. We propose two strategies, including the baseline model of only structured features and text plus structured features fusion model.

2.2. Our Methods. To implement this task (Figure 1), we will do the following:

- (1) Extract text information of HPI, laboratory tests, vital signs, and demographic data within the first 72 h of admission
- (2) Clean HPI, including but not limited to, punctuation marks and meaningless symbol, then make word segmentation
- (3) By using the pretrained BERT model for Chinese, convert words into word vectors, which are then converted into uniform representation feature vectors by attention-based LSTM
- (4) Combine structured features (laboratory tests, vital signs, and demographic data) and unstructured features (HPI) and feed them into softmax function to predict organ failures in AP
- (5) Use minimized cross entropy to estimate the parameters

The input of the model is $\{X_i, S_i, Y_i\}$, where i is the index of the sample, X is the structural variables such as vital signs and demographic, S is the text description, and Y is the occurrence of a specific target event, such as organ failure. S_i can be expressed as a string of Chinese words. For each Chinese word, the corresponding Chinese word is first converted into the form of one-hot coding according to the dictionary, and then each word is converted into a word vector by using the word embedding matrix of the pretrained BERT model, as shown in formula (1). Then, the word vector of each word is input into the attention-based LSTM network according to the order of sentences. The last hidden layer is taken as the input of the next layer, which is recorded as $h_u^{(i)}$. The structured information X_i is mapped to the hidden layer vector $h_X^{(i)}$ through the neural network, and then $h_X^{(i)}$ and $h_u^{(i)}$ are spliced to get the vector $h_f^{(i)}$. Finally, $h_f^{(i)}$ is used as the input of FC_k , and $Y_k^{(i)}$ is used as the output for the specific target event. The formula of attention-based LSTM is defined as follows. Based on the existing standard

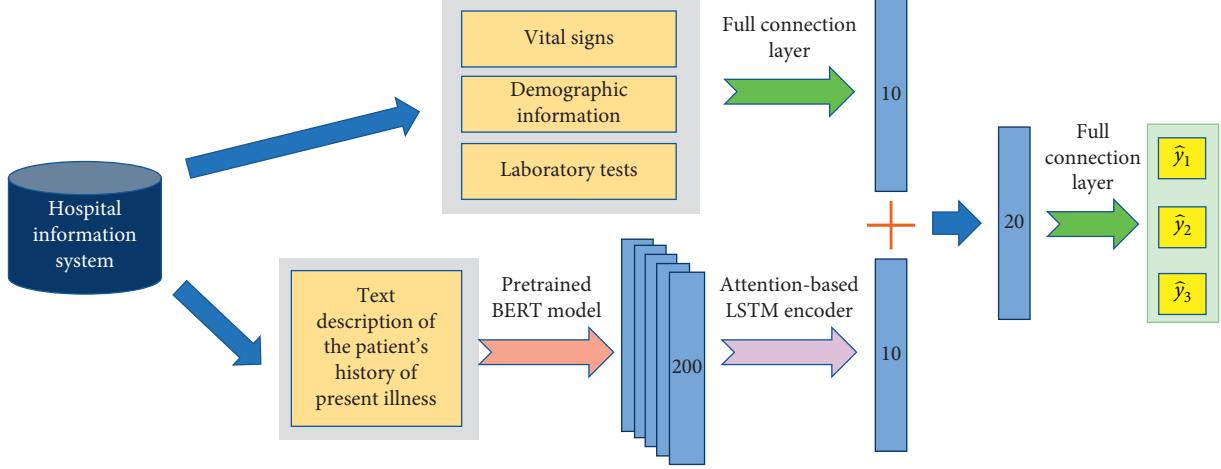


FIGURE 1: Schematic diagram of prediction organ failures in AP using text mining.

LSTM network, when the last hidden layer outputs, attention weight is added, which is recorded as $e_m^{(i)}$. $h_m^{(i)}$ represents hidden layer vector mapped from the word vector of a word:

$$h_X^{(i)} = \text{FC}(X_i), \quad (1)$$

$$\begin{aligned} u_m^{(i)} &= W * \text{onehot}(w_m^{(i)}), \\ h_u^{(i)} &= \text{ATTLSTM}([u_1^{(i)}, u_2^{(i)}, \dots, u_m^{(i)}]), \\ h_f^{(i)} &= \text{concat}([h_X^{(i)}, h_u^{(i)})], \\ Y_k^{(i)} &= \text{FC}_k(h_f^{(i)}), \\ h_m^{(i)} &= \text{LSTM}(u_m^{(i)}), \\ h_u^{(i)} &= h_{\text{last}}^{(i)} = \sigma \left(\sum_{m=1}^M e_m^{(i)} * h_m^{(i)} \right), \\ e_m^{(i)} &= \frac{h_{\text{last}}^{(i)} \odot h_m^{(i)}}{\sum_{m=1}^M (h_{\text{last}}^{(i)} \odot h_m^{(i)})}. \end{aligned} \quad (2)$$

Different from other ways of using the pretrained BERT model, we did not directly use the prediction task of organ failure as a subtask of the BERT model but used the embedding matrix of the BERT model to convert the sparse word vector represented by one-hot encoding to a dense real vector of degree 200. Therefore, in this study, a sentence is a matrix, which can also be regarded as a sequence of vectors. Then, we input this vector sequence into the attention-based long short-term memory model in turn and take out the final hidden layer with a length of 10 neurons as the feature representation of the text information.

2.3. Data. The patients diagnosed with AP based on ICD code from the hospital information system of West China Hospital, Sichuan University, from 2008 to 2018 were included. There were 15,813 AP patients included initially. We extracted AP patients' demographic data, laboratory tests, and vital signs within 72 hours of admission. HPI of AP patients also was extracted. Respiratory failure was defined as the partial pressure of oxygen in blood gas analysis was

less than 60 mmHg or the use of a ventilator. Circulatory failure was defined as diastolic blood pressure was less than 60 mmHg or systolic blood pressure was less than 90 mmHg and the use of vasoactive drugs. Kidney failure was defined as creatinine was greater than 177 umol/L. The research protocol was approved by the ethics review board of West China Hospital, Sichuan University, and the need for informed consent was waived owing to the retrospective nature of the study.

2.4. Data Preprocessing. The features with a missing rate of more than 20% in the structured data and patients without a complete description of HPI were deleted. A total of 12,748 patients were used for prediction finally. We used mice package of R software to perform multiple linear interpolation on structured data. Before and after linear interpolation (see Figure S2), there is no statistical difference in the distribution of the data ($P > 0.05$).

2.5. Word Embedding. Word embedding can be represented by a sparse one-hot vector or a dense real vector. Word vectors represented sparsely not only require large storage space and computing resources but are also difficult to reflect the correlation between words. Therefore, since dense vectors solve the above two problems at the same time, dense real vectors are more commonly used. In the classification and prediction tasks of natural language processing, that is, the issue of sequence to sequence, the quality of the word vector often has a great impact on the prediction ability of the model. There are currently a large number of word vectors that have been trained based on continuous bag-of-words (CBOW) and skip-gram models [18]. This study adopted an analogical reasoning task on Chinese [19]. In order to ensure the accuracy of Chinese medical word segmentation, Tsinghua Open Chinese Lexicon [20] was applied.

2.6. Training and Testing. The data then was further divided into a training dataset and testing dataset according to the

TABLE 1: Baseline structured features and organ failures of AP patients.

Characteristics	
Demographic characteristics	
Age, y , mean (SD)	47.58 (15.01)
Male, n (%)	7.764 (60.9)
Early vital signs	
Respiratory rate, mg (μl)/(h·g), mean (SD)	20.46 (2.22)
Pulse rate/min, mean (SD)	85.70 (15.19)
Early laboratory tests	
Serum	
AST/ALT, mean (SD)	1.24 (1.17)
Cholesterol, mmol/L, mean (SD)	4.57 (2.31)
High-density lipoprotein, mmol/L, mean (SD)	0.87 (0.42)
Low-density lipoprotein, mmol/L, mean (SD)	1.91 (1.00)
Amylase, IU/L, mean (SD)	304.87 (555.37)
Carbon dioxide combining power, mmol/L, mean (SD)	22.20 (3.58)
Calcium, mmol/L, mean (SD)	2.06 (0.24)
Glutamyl transpeptidase, IU/L, mean (SD)	139.90 (214.64)
Albumin, g/L, mean (SD)	35.71 (5.98)
Albumin/globulin, mean (SD)	1.47 (0.36)
White blood cell count, $10^9/\text{L}$, mean (SD)	10.96 (5.13)
Alanine aminotransferase, IU/L, mean (SD)	62.57 (113.28)
Whole blood	
RBC distribution width CV, %, mean (SD)	14.20 (1.48)
RBC distribution width SD, fL, mean (SD)	46.94 (5.27)
Monocytes, %, mean (SD)	5.50 (2.19)
Monocytes, $10^9/\text{L}$, mean (SD)	0.57 (0.31)
Organ failures	
Respiratory failure, n (%)	1.806 (14.2)
Circulatory failure, n (%)	2.752 (21.6)
Renal failure, n (%)	579 (4.5)

SD: standard deviation.

proportions of 70% and 30%. Because of class imbalance and the importance of positive identification in medicine, we used $1 - r_i$ as the weight of the i -th class to make up for the problems caused by a class imbalance in the cross-entropy loss function [21]. We used a pretrained word vector network with a word vector length of 60. The batch size was set to 500 and the learning rate was set to 0.005. The number of hidden layer neurons of LSTM was set to 200. Models were optimized using a gradient descent approach. The performance of training was an average of 1000 epochs. In our model, mixed features including structured features and text features were as input and combined attention mechanism, compared to the model that used only structured features as input. PyTorch framework was adopted to implement the experiment on a Dell T640 GPU server.

3. Results and Discussion

There were 12,748 AP patients with average age of 47.58 years and 60.9% male in this study. Respiratory failure, circulatory failure, and renal failure accounted for 14.2%, 21.6%, and 4.5%, respectively. Early vital signs and early laboratory tests are shown in Table 1.

Table 2 contains the results of four models: model 1, we use structured features as input to predict three organ failures, respectively; model 2, we use text features as input to predict three organ failures, respectively; model 3, we use mixed

features including structured features and text features as input combined with the attention mechanism to predict three organ failures, respectively; model 4, we use mixed features as input and combine with the attention mechanism to predict three organ failures together as multitask. From the training dataset perspective, model 2 performed best through the prediction of organ failures. From the testing dataset perspective, the accuracies of model 3 and model 4 were higher than model 1 and model 2 to predict respiratory failure, circulatory failure, and renal failure, which shows that adding text features can effectively help improve the accuracy of predicting organ failures in AP.

In the classification task, a good classifier should have a good effect on the judgment of each class, so when evaluating the effect of the classifier, the recall and specificity should be considered comprehensively. The specificity of model 3 is the highest to predict respiratory failure and circulatory failure. It shows that model 3 can effectively learn some characteristics of AP patients without organ failures. Although the accuracy, precision, and specificity of renal failure are the highest in model 1, the recall of model 1 is lower than that of model 3 and model 4. The highest recall of respiratory failure comes from model 3. Based on the comprehensive evaluation using the performance matrix, the performance of the model with text features added is superior to the model that only includes structured features or text features. The top 30 important Chinese word

TABLE 2: The performance of early predicting organ failures in AP using our proposed model.

Model	Performance	Respiratory failure	Circulatory failure	Renal failure
Model 1				
Training	Accuracy	0.727 (0.033)	0.668 (0.034)	0.786 (0.048)
	Recall	0.762 (0.023)	0.680 (0.036)	0.758 (0.033)
	Precision	0.446 (0.039)	0.374 (0.040)	0.141 (0.039)
	Specificity	0.776 (0.044)	0.727 (0.026)	0.832 (0.015)
Testing	Accuracy	0.646 (0.027)	0.684 (0.022)	0.829 (0.046)
	Recall	0.649 (0.034)	0.664 (0.023)	0.681 (0.036)
	Precision	0.338 (0.013)	0.374 (0.013)	0.169 (0.040)
	Specificity	0.692 (0.047)	0.718 (0.027)	0.842 (0.020)
Model 2				
Training	Accuracy	0.744 (0.039)	0.820 (0.013)	0.871 (0.007)
	Recall	0.866 (0.011)	0.769 (0.035)	0.905 (0.036)
	Precision	0.482 (0.020)	0.584 (0.021)	0.264 (0.031)
	Specificity	0.789 (0.025)	0.844 (0.034)	0.892 (0.036)
Testing	Accuracy	0.659 (0.030)	0.752 (0.033)	0.823 (0.042)
	Recall	0.619 (0.025)	0.723 (0.020)	0.519 (0.022)
	Precision	0.342 (0.048)	0.457 (0.008)	0.136 (0.029)
	Specificity	0.683 (0.022)	0.783 (0.050)	0.746 (0.020)
Model 3				
Training	Accuracy	0.659 (0.008)	0.820 (0.029)	0.803 (0.027)
	Recall	0.850 (0.022)	0.784 (0.035)	0.894 (0.013)
	Precision	0.402 (0.014)	0.582 (0.034)	0.194 (0.019)
	Specificity	0.788 (0.045)	0.844 (0.015)	0.892 (0.021)
Testing	Accuracy	0.687 (0.019)	0.771 (0.035)	0.789 (0.024)
	Recall	0.664 (0.014)	0.732 (0.031)	0.708 (0.017)
	Precision	0.377 (0.012)	0.485 (0.010)	0.143 (0.043)
	Specificity	0.727 (0.021)	0.808 (0.046)	0.831 (0.035)
Model 4				
Training	Accuracy	0.772 (0.047)	0.784 (0.036)	0.838 (0.032)
	Recall	0.738 (0.044)	0.792 (0.044)	0.825 (0.024)
	Precision	0.371 (0.017)	0.480 (0.016)	0.214 (0.011)
	Specificity	0.814 (0.043)	0.831 (0.018)	0.873 (0.019)
Testing	Accuracy	0.689 (0.030)	0.687 (0.011)	0.655 (0.044)
	Recall	0.543 (0.010)	0.688 (0.016)	0.854 (0.026)
	Precision	0.228 (0.033)	0.374 (0.026)	0.105 (0.041)
	Specificity	0.649 (0.038)	0.720 (0.014)	0.768 (0.023)

Model 1: input of structured features only. Model 2: input of text features only. Model 3: input of mixed features combined with attention mechanism. Model 4: input of mixed features and use of attention mechanism with multitask. Standard deviation is in the bracket.

segmentations were obtained through the attention mechanism from model 4, which is shown in Table S1.

Compared to the model of Krishnan and Kamath [15] (Table 3), our proposed model performs better than it.

Mentula et al. [6] used 351 AP patients' some laboratory tests and APACHE II score within 12 h of admission to predict organ failure through logistic regression. Their results showed 0.82 of recall to predict respiratory failure and renal failure without testing, which is lower than the predicted performance on the training dataset of respiratory failure and renal failure with text features added in this study (0.850 and 0.894, resp.). Khanna et al. [22] used 72 AP patients' various scores and a few laboratory tests to predict organ failure with a maximum recall of 1 using procalcitonin and a minimum recall of 0.652 using CT severity index. It is difficult to believe that only one score or a single indicator can be used to achieve such a good prediction effect. Their findings need to be further verified.

Koutroumpakis et al. [23] used 1,612 AP patients' three laboratory tests and APACHE II score to predict

persistent organ failure. The best recall was 0.684 using admission APACHE II score, and the lowest was 0.249 using admission creatinine. Although these results seem reasonable, they have not been tested and are lower than the results of the training dataset after adding text features in this study. Hyland et al. [24] developed a machine learning method to predict circular failure in ICU patients. Although the predicted performance of AUC of 0.94 was obtained, the data used was routinely collected structured data. In addition, ICU data is monitored in real time, so the onset and duration of organ failure can be observed and a time series model can be used. However, in routine inpatients, laboratory tests and vital signs are monitored irregularly. This is why this study did not use earlier data and time series models to predict persistent organ failure.

In addition to predicting single organ failure, we also used predictive multitasking to output the results of three organ failures simultaneously after adding text features. In the

TABLE 3: The performance of early predicting organ failures in AP using the model of Krishnan and Kamath [15].

Model	Performance	Respiratory failure	Circulatory failure	Renal failure
Model 1				
Training	Accuracy	0.585 (0.046)	0.526 (0.047)	0.640 (0.067)
	Recall	0.624 (0.033)	0.537 (0.050)	0.616 (0.046)
	Precision	0.352 (0.054)	0.280 (0.056)	0.111 (0.055)
	Specificity	0.630 (0.062)	0.588 (0.036)	0.696 (0.021)
	Accuracy	0.506 (0.038)	0.546 (0.032)	0.682 (0.064)
	Recall	0.507 (0.048)	0.526 (0.033)	0.538 (0.050)
Testing	Precision	0.253 (0.019)	0.288 (0.019)	0.139 (0.056)
	Specificity	0.546 (0.065)	0.578 (0.038)	0.705 (0.028)
Model 2				
Training	Accuracy	0.600 (0.055)	0.685 (0.018)	0.738 (0.009)
	Recall	0.732 (0.016)	0.626 (0.049)	0.762 (0.050)
	Precision	0.395 (0.028)	0.447 (0.029)	0.237 (0.044)
	Specificity	0.650 (0.036)	0.702 (0.047)	0.749 (0.050)
	Accuracy	0.518 (0.042)	0.610 (0.046)	0.678 (0.058)
	Recall	0.480 (0.035)	0.586 (0.028)	0.381 (0.031)
Testing	Precision	0.246 (0.067)	0.374 (0.010)	0.110 (0.040)
	Specificity	0.545 (0.032)	0.636 (0.070)	0.609 (0.028)
Model 3				
Training	Accuracy	0.525 (0.012)	0.680 (0.040)	0.663 (0.038)
	Recall	0.712 (0.031)	0.641 (0.050)	0.758 (0.018)
	Precision	0.316 (0.021)	0.440 (0.048)	0.172 (0.026)
	Specificity	0.643 (0.063)	0.708 (0.021)	0.754 (0.030)
	Accuracy	0.550 (0.026)	0.628 (0.050)	0.650 (0.034)
	Recall	0.529 (0.020)	0.591 (0.044)	0.572 (0.024)
Testing	Precision	0.292 (0.016)	0.401 (0.014)	0.112 (0.061)
	Specificity	0.589 (0.030)	0.662 (0.064)	0.688 (0.050)
Model 4				
Training	Accuracy	0.626 (0.065)	0.641 (0.050)	0.696 (0.046)
	Recall	0.593 (0.062)	0.646 (0.062)	0.686 (0.034)
	Precision	0.285 (0.024)	0.394 (0.023)	0.194 (0.015)
	Specificity	0.668 (0.061)	0.695 (0.025)	0.736 (0.026)
	Accuracy	0.548 (0.042)	0.552 (0.015)	0.509 (0.061)
	Recall	0.409 (0.014)	0.551 (0.023)	0.714 (0.036)
Testing	Precision	0.201 (0.047)	0.284 (0.036)	0.075 (0.057)
	Specificity	0.506 (0.053)	0.584 (0.021)	0.630 (0.033)

Model 1: input of structured features only. Model 2: input of text features only. Model 3: input of mixed features combined with attention mechanism. Model 4: input of mixed features and use of attention mechanism with multitask. Standard deviation is in the bracket.

multitask prediction model, the loss function of each task can be regarded as the constraint of other tasks. The prediction results of the three organ failures can be obtained in a shorter time, which may meet different clinical needs. We do not want the model to be limited to the learning of the target task but can adapt to multiple task scenarios, which can greatly increase the functional capability of the model (generalization).

Since it is a single-center study, patients only come from West China Hospital, Sichuan University, which is a large general hospital with 4300 beds in China. Doctors in West China Hospital, Sichuan University, may describe the patient's current medical history in more detail than doctors in other hospitals. Therefore, the addition of text information will increase the predictive ability of the model. However, in the same way, when used in other hospitals, the text information of the current medical history may be different from the text information of this study, so it should be careful when using the proposed model.

4. Conclusions

We performed single-task and multitask prediction of organ failures in AP by the joint representation of structured features and text features. According to our best knowledge, this is the first time to use clinical notes to predict organ failures in AP. Our methods achieve superior accuracy compared to traditional techniques and uncover the underlying structure of the disease and intervention space.

Data Availability

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

Jiawei Luo and Lan Lan contributed equally.

Acknowledgments

This work was supported by the Postdoctoral Research Project, West China Hospital, Sichuan University (no. 2019HXBH039), the 135 Project for Disciplines of Excellence, West China Hospital, Sichuan University (no. ZYJC18010), and the Center of Excellence-International Collaboration Initiative Grant (no. 139170052).

Supplementary Materials

The supplementary materials contain Figures S1 and S2 and Table S1. (*Supplementary Materials*)

References

- ## Authors' Contributions

Jiawei Luo and Lan Lan contributed equally.

Acknowledgments

This work was supported by the Postdoctoral Research Project, West China Hospital, Sichuan University (no. 2019HXBH039), the 1 3 5 Project for Disciplines of Excellence, West China Hospital, Sichuan University (no. ZYJC18010), and the Center of Excellence-International Collaboration Initiative Grant (no. 139170052).

Supplementary Materials

The supplementary materials contain Figures S1 and S2 and Table S1. (*Supplementary Materials*)

References

 - [1] H. Li, Z. Qian, Z. Liu, X. Liu, X. Han, and H. Kang, "Risk factors and outcome of acute renal failure in patients with severe acute pancreatitis," *Journal Of Critical Care*, vol. 25, no. 2, pp. 225–229, 2010.
 - [2] P. Kes, Ž. VuČIĆeviĆ, I. RatkoviĆ-GusiĆ, A. Fotivec, and P. Kes, "Acute renal failure complicating severe acute pancreatitis," *Renal Failure*, vol. 18, no. 4, pp. 621–628, 1996.
 - [3] C. J. Shields, D. C. Winter, and H. P. Redmond, "Lung injury in acute pancreatitis: mechanisms, prevention, and therapy," *Current Opinion in Critical Care*, vol. 8, no. 2, pp. 158–163, 2002.
 - [4] H. Mehta, I. Shah, M. Pahuja et al., "Outcomes of acute pancreatitis in patients with heart failure: insights from the nationwide inpatient sample," *Journal of Cardiac Failure*, vol. 25, no. 8, pp. S57–S58, 2019.
 - [5] P. A. Banks, T. L. Bollen, C. Dervenis et al., "Classification of acute pancreatitis--2012: revision of the Atlanta classification and definitions by international consensus," *Gut*, vol. 62, no. 1, pp. 102–111, 2013.
 - [6] P. Mentula, M. L. Kylänpää, E. Kemppainen et al., "Early prediction of organ failure by combined markers in patients with acute pancreatitis," *British Journal of Surgery*, vol. 92, no. 1, pp. 68–75, 2005.
 - [7] M. Sporek, P. Dumnicka, A. Gala-Bladzinska et al., "Angiopoietin-2 is an early indicator of acute pancreatic-renal syndrome in patients with acute pancreatitis," *Mediators of Inflammation*, vol. 2016, 7 pages, 2016.
 - [8] C. Liu, J. Chyr, W. Zhao et al., "Genome-wide association and mechanistic studies indicate that immune response contributes to alzheimer's disease development," *Frontiers in Genetics*, vol. 9, p. 410, 2018.
 - [9] Z. Ji, D. Wu, W. Zhao et al., "Systemic modeling myeloma-osteoclast interactions under normoxic/hypoxic condition using a novel computational approach," *Scientific Reports*, vol. 5, no. 1, Article ID 13291, 2015.
 - [10] Z. Ji, W. Zhao, H. K. Lin, and X. Zhou, "Systematically understanding the immunity leading to CRPC progression," *PLoS Computational Biology*, vol. 15, no. 9, Article ID e1007344, 2019.
 - [11] Y. Cui, W. Che, T. Liu et al., "Pre-training with whole word masking for Chinese BERT," 2019, <https://arxiv.org/abs/1906.08101>.
 - [12] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," 2016, <https://arxiv.org/abs/1409.0473>.
 - [13] P. Nguyen, T. Tran, N. Wickramasinghe, and S. Venkatesh, "\$\mathcal{M}^{Deepr}\$: a convolutional net for medical records," *IEEE Journal of Biomedical and Health Informatics*, vol. 21, no. 1, pp. 22–30, 2017.
 - [14] E. Soysal, J. Wang, M. Jiang et al., "Clamp-a toolkit for efficiently building customized clinical natural language processing pipelines," *Journal of The American Medical Informatics Association*, vol. 25, no. 3, pp. 331–336, 2018.
 - [15] G. S. Krishnan and S. Kamath, "Hybrid text feature modeling for disease group prediction using unstructured physician notes," 2019, <https://arxiv.org/abs/1911.11657>.
 - [16] S. Yuwono, H. Ng, and K. Ngiam, "Learning from the experience of doctors: automated diagnosis of appendicitis based on clinical notes," in *Proceedings of the 18th BioNLP Workshop and Shared Task*, pp. 11–19, Florence, Italy, August 2019.
 - [17] O. Akbilgic, R. Homayouni, K. Heinrich, M. Langham, and R. Davis, "Unstructured text in EMR improves prediction of death after surgery in children," *Informatics*, vol. 6, pp. 1–11, 2019.
 - [18] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," 2013, <https://arxiv.org/abs/1301.3781>.
 - [19] S. Li, Z. Zhao, R. Hu, W. Li, T. Liu, and X. Du, "Analogical reasoning on Chinese morphological and semantic relations," 2018, <https://arxiv.org/abs/1805.06504>.
 - [20] A. K. Khanna, S. Meher, S. Prakash et al., "Comparison of ranson, glasgow, moss, sirs, bisap, Apache-II, ctsi scores, IL-6, crp, and procalcitonin in predicting severity, organ failure, pancreatic necrosis, and mortality in acute pancreatitis," *HPB Surgery*, vol. 2013, 10 pages, 2013.
 - [21] E. Koutroumpakis, B. U. Wu, O. J. Bakker et al., "Admission hematocrit and rise in blood urea nitrogen at 24 h outperform other laboratory markers in predicting persistent organ failure and pancreatic necrosis in acute pancreatitis: a post hoc analysis of three large prospective databases," *American Journal of Gastroenterology*, vol. 110, no. 12, pp. 1707–1716, 2015.
 - [22] S. L. Hyland, M. Faltys, M. Hüser et al., "Early prediction of circulatory failure in the intensive care unit using machine learning," *Nature Medicine*, vol. 26, no. 3, pp. 364–373, 2020.
 - [23] S. Han, Y. Zhang, Y. Ma et al., "THUOCL: Tsinghua open Chinese lexicon," *Journal of Chinese Linguistics*, vol. 2020, 5 pages, 2020.
 - [24] Y. S. Aurelio, G. M. de Almeida, C. L. de Castro, and A. P. Braga, "Learning from imbalanced data sets with weighted cross-entropy function," *Neural Processing Letters*, vol. 50, no. 2, pp. 1937–1949, 2019.