A Novel Method of Heart Failure Prediction Based on DPCNN-XGBOOST Model

Yuwen Chen[1, 2, 3, *], Xiaolin Qin[1, 3], Lige Zhang[1, 3] and Bin Yi[4]

Abstract: The occurrence of perioperative heart failure will affect the quality of medical services and threaten the safety of patients. Existing methods depend on the judgment of doctors, the results are affected by many factors such as doctors' knowledge and experience. The accuracy is difficult to guarantee and has a serious lag. In this paper, a mixture prediction model is proposed for perioperative adverse events of heart failure, which combined with the advantages of the Deep Pyramid Convolutional Neural Networks (DPCNN) and Extreme Gradient Boosting (XGBOOST). The DPCNN was used to automatically extract features from patient's diagnostic texts, and the text features were integrated with the preoperative examination and intraoperative monitoring values of patients, then the XGBOOST algorithm was used to construct the prediction model of heart failure. An experimental comparison was conducted on the model based on the data of patients with heart failure in southwest hospital from 2014 to 2018. The results showed that the DPCNN-XGBOOST model improved the predictive sensitivity of the model by 3% and 31% compared with the text-based DPCNN Model and the numeric-based XGBOOST Model.

Keywords: Deep pyramid convolutional neural networks, extreme gradient boosting, heart failure prediction.

1 Introduction

Heart failure has been considered as one of the deadliest human diseases worldwide, and the accurate prediction of this risk would be vital for heart failure prevention and treatment. However, the simple early warning system of adverse events often cannot catch signs of heart failure adverse events in time. Once the adverse events occur, the disease is serious or terminal, resulting in the difficult treatment and the limited effect of intervention. It is of great scientific significance and social value to actively develop the

[1] Chengdu Institute of Computer Applications, Chinese Academy of Sciences, Chengdu, 610041, China.

[2] Chongqing Institute of Green and Intelligent Technology, Chinese Academy of Sciences, Chongqing, 400714, China.

[3] School of Computer and Control Engineering, University of Chinese Academy of Sciences, Beijing, 100049, China.

[4] Department of Anesthesia, Southwest Hospital, Army Medical University, Chongqing, 400038, China.

[*] Corresponding Author: Yuwen Chen. Email: chenyuwen@cigit.ac.cn.

Received: 29 April 2020; Accepted: 21 May 2020.

risk prediction of heart failure adverse events, which is helpful in the early warning and intervention clues of adverse events.

Recently, artificial intelligence technology has been widely used in the medical field [Hannun, Rajpurkar, Haghpanahi et al. (2019); Gurovich, Hanani, Bar et al. (2019); Gottesman, Johansson, Komorowski et al. (2019); Esteva, Robicquet, Ramsundar et al. (2019); Attia, Kapa, Lopezjimenez et al. (2019); Chen, Sun and Zhong (2018); Chen, Wang, Lin et al. (2019); Chen, Zhong, Wang et al. (2018)]. Currently, the research on heart failure is mainly based on the data from patients' medical records, physical characteristics, auxiliary examination, the treatment plan, and the algorithm is used to build the model for studying, analyzing and classifying of diagnosis and prediction. In addition, most studies mainly analyzed the characteristics of electrocardiogram data and built the diagnostic model of heart failure [Yu and Lee (2012); Masetic and Subasi (2016); Melillo, Izzo, Orrico et al. (2015); Acharya, Fujita, Sudarshan et al. (2017)]. Zheng et al. [Zheng, Zhang, Yoon et al. (2015)] presented a method used support vector machine algorithm to analyze the data of patients with heart failure, including age, type of medical insurance, sensitivity assessment (audio-visual and thinking), complications, emergency treatment, the drug-induced risks, the period of last hospitalization, and built a prediction model for the readmission of patients with heart failure, with a prediction accuracy of 78.4%. Choi et al. [Choi, Schuetz, Stewart et al. (2017)] used the recurrent neural network algorithm to analyze the diagnostic data of patients with heart failure, including time series of doctor's orders, spatial density and other characteristics, to build a diagnostic model of heart failure, and verified by experiment that the area under the curve (AUC) of the diagnosis of this model was 0.883. Chen et al. [Chen, Zheng, Li et al. (2016)] analyzed 24 hours dynamic electrocardiogram of heart failure patients and healthy controls by using support vector machine (SVM) algorithm based on non-equilibrium decision tree. Shameer et al. [Shameer, Johnson, Yahi et al. (2017)] also utilized Naive Bayes algorithm to analyze about data variables of patients with heart failure, including diagnosis data, treatment data, examination data, records of doctor's orders, and vital signs data, and built a model for predicting readmission of patients with heart failure, with a predicted AUC of 0.78. However, these studies only focused on the numerical data of preoperative test or intraoperative monitoring. Its prediction of poor timeliness, low accuracy, and non-fusion of patient heterogeneity data. Specifically, the heterogeneity of patients are embodied in many aspects, such as the age, occupation, gender, various physiological indices [Koerkamp, Stijnen, Weinstein et al. (2011)] as well as various types of numerical laboratory data, textual diagnostic information. These indices are used to predict heart failure, the results are more practical and targeted. Therefore, in order to make the results of research and analysis more accurate and more convenient to apply to practice, in this paper, we use Deep Pyramid Convolutional Neural Networks (DPCNN) and Extreme Gradient Boosting (XGBOOST) method to model the risk prediction of heart failure after operation based on the preoperative and Intraoperative medical data of patients, so as to construct a scalable, low-cost and effective prediction solution for heart failure.

Our main contributions are in two areas: 1) A novel heart failure prediction framework is proposed, which integrates the advantages of deep neural network and XGBOOST model, and uses the structure and unstructured data of patients before and during operation to predict heart failure. 2) Experiments based on real heart failure data show that fusion of multi-source heterogeneous data can improve the accuracy of heart failure prediction.

The rest of this paper is organized as follows: The preliminary and related technology, and methodology of this paper is discussed in Section 2. Section 3 reports the experimental results and discusses the implications of the study. Finally, Section 4 discusses the conclusion of this paper.

2 Methods

2.1 Deep Pyramid Convolutional Neural Networks (DPCNN)

Deep Pyramid Convolutional Neural Networks (DPCNN) [Johnson and Zhang (2017)] is a low-complexity word-level deep convolutional neural network architecture for text categorization [Zhang and Wallace (2017), Liu, Qiu and Huang (2016)] that can efficiently represent long range associations in text. The specific structure is shown in the Fig. 1. The first layer performs text region embedding, which generalizes commonly used word embedding to the embedding of text regions covering one or more words. It is followed by stacking of convolution blocks (two convolution layers and a shortcut) interleaved with pooling layers with stride 2 for down sampling. The final pooling layer aggregates internal data for each document into one vector. It uses max pooling for all pooling layers.

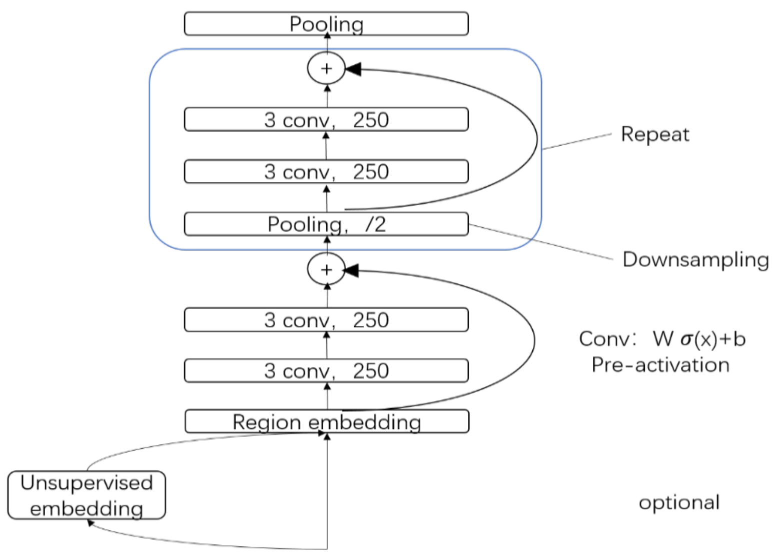


Figure 1: Network architecture of DPCNN model

The DPCNN can effectively extract the features of long-distance relationship of the text with low complexity and better effect than the previous CNN structure. Therefore, we use this model to extract the features of unstructured diagnosis text of patients to predict heart failure.

2.2 Extreme gradient boosting

XGBOOST [Chen and Guestrin (2016)] is used to extract the numerical features of patients, which are an integrated learning method proposed by Tianqi Chen based on GBDT [Ye, Chow, Chen et al. (2009)]. The improvement of XGBOOST algorithm to GBDT algorithm lies in that the second derivative is used to calculate the objective function in the process of model optimization, besides, the regularization term is added to the objective function to prevent the algorithm from over-fitting in the training process, moreover, XGBOOST algorithm uses the idea of random forest for reference in the training process, and does not use all samples in the iteration process, and does not use every iteration. The generalization ability of the model is effectively improved by sampling all the features of the samples and training some of the features of the samples. Different patients will have different preoperative tests and intraoperative monitoring according to different diseases. In order to facilitate the construction of data set, this paper integrates the test indicators and intraoperative monitoring indicators of all patients, so a large number of null values will be generated in the data set. XGBOOST can treat the missing values as a sparse matrix, so this paper uses XGBOOST model for prediction modeling.

2.3 DPCNN-XGBOOST models

The DPCNN- XGBOOST heart failure prediction model proposed in this paper is shown in Fig. 2, which combined with the advantages of DPCNN and XGBOOST. The DPCNN was used to automatically extract features of patients' diagnostic texts, and the text features were integrated with the preoperative examination and intraoperative monitoring values of patients, then the XGBOOST algorithm was used to construct the prediction model of heart failure. It includes three parts in Fig. 2.

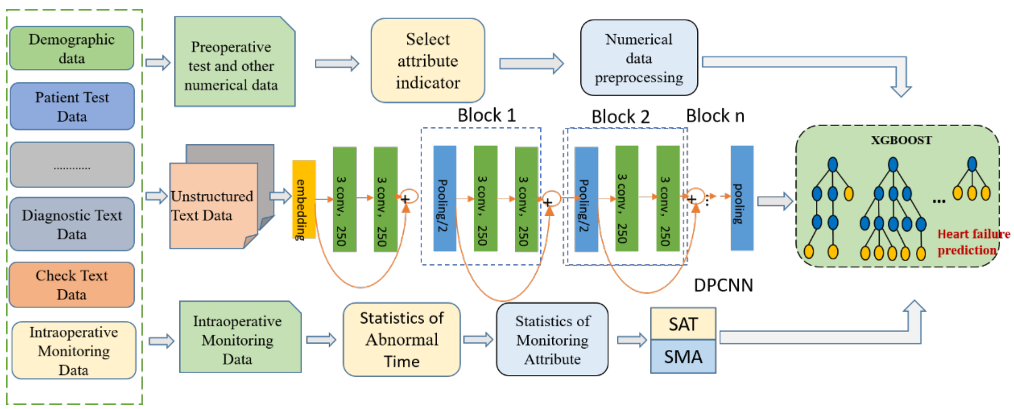


Figure 2: DPCNN-XGBOOST model architecture

2.3.1 Feature extraction of preoperative structured numerical data

Different patients will carry out different test items according to different diseases. In order to build standard data sets, this paper extracts indicators with common test attributes for positive patients and negative patients as feature items (see Tab. 1 for specific test attribute indicators). Due to the different dimension of test items, the test data of patients were standardized and normalized to facilitate data analysis, and the processed data can be classified and predicted as the input of XGBOOST. The top part of Fig. 2.

Table 1: Attributes of preoperative patient test

N	Attribute Name	N	Attribute Name	N	Attribute Name
1	neutrophil count	2	troponin	3	bilirubin
4	uric acid	5	d-dimer	6	creatine kinase
7	total protein	8	urea	9	troponin Ⅰ
10	international normalized ratio	11	monocyte percentage	12	glucose
13	fibrinogen degradation product	14	osmotic pressure	15	thrombin time
16	fibrinogen	17	aspartate amino transferase	18	prothrombin time
19	percentage of neutrophils	20	gamma glutamyl transferase	21	monocyte count
22	hydroxybutyrate dehydrogenase.	23	percentage of eosinophils	24	activated partial thromboplastin time
25	lymphocyte count	26	cholinesterase	27	sodium
28	creatinine	29	alanine aminotransferase	30	indirect bilirubin
31	white blood cell count	32	potassium	33	alkaline phosphatase
34	creatine kinase isoenzyme	35	percentage of basophils	36	red blood cell
37	total bile acid	38	total cholesterol	39	hemoglobin
40	eosinophils count	41	conjugated bilirubin	42	calcium
43	glycocholic acid	44	cystatin	45	albumin
46	albumin/globin	47	percentage of lymphocytes	48	basophilic granulocyte count
49	phosphatemia	50	red blood cell count	51	hematocrit
52	age	53	sex	54	anamnesis
55	the past operation	56	smoking history		

2.3.2 Feature extraction of preoperative unstructured text data

Table 2: Unstructured text attributes of preoperative patients

Number	Attribute Name
1	previous medical illness
2	chief complaint-positive symptom/sign
3	history of present illness
4	echocardiogram
5	electrocardiogram
6	chest X-ray
7	preoperative clinical diagnosis

Table 3: Normal value range of patient monitoring attribute during operation

Number	Attribute Name	**Value Interval**
1	HR	[50, 100]
2	SBP	[90, 140]
3	DBP	[60, 90]
4	CVP	[5, 12]
5	RR	[12, 20]
6	ETCO2	[35, 45]
7	Temperature	[36.2, 37.2]
8	Rectal T	[36.5, 37.7]
9	SpO2	[95, 100]
10	ADBP	[60, 90]
11	PULSE	[60, 100]
12	MAP	[60, 10000]
13	MBP	[60, 10000]
14	ASBP	[90, 140]
15	PAs	[15, 30]
16	PAd	[6, 10]
17	PAm	[12, 16]
18	RAPm	[5, 7]
19	BID	[40, 60]
20	TBlood	[36.2, 37.2]

The feature extraction of preoperative unstructured text data is the text that extracting the previous medical illness, chief complaint-positive symptom/sign, history of present illness, echocardiogram, electrocardiogram, chest X-Ray, preoperative clinical diagnosis of patients (See Tab. 2 for specific patient examination text attributes). The text was preprocessed and converted into a word vector by using Tencent open-sources word vector [Song, Shi, Li et al. (2018)], then the predictive classification features of heart failure were extracted by DPCNN, finally forming a text vector of 250 dimensions. The middle part of the Fig. 2.

2.3.3 Feature extraction of intraoperative monitoring data

Intraoperative monitoring data are time series data, in this paper, we convert intraoperative monitoring data into the fixed features by the feature engineering methods. The methods used in the model include statistical outliers and statistical characteristics of patient monitoring attributes. The outliers are the time (in minutes) that the monitoring attribute data of patients exceed the normal interval, as shown in Tab. 3. The statistics include the maximum, minimum, mean, variance, standard deviation, kurtosis and skewness of each attribute. The formula is as follows. According to different types of operation and patients' diseases, the attributes of intraoperative monitoring are different. In order to construct the standard data set, this paper counts the intraoperative monitoring attributes of all patients (negative and positive), as the monitoring attribute column of patients during operation. When patients do not monitor an attribute, the value is empty, and finally the intraoperative monitoring data of patients is converted into a fixed dimension vector.

$$\mu = \frac{1}{T}\sum_{i=1}^{T} x_i$$
$$\sigma^2 = \sum_{i=1}^{T} \frac{1}{T}(x_i - \mu)^2$$
$$\text{skewness(X)} = E\left[\left(\frac{X-\mu}{\sigma}\right)^3\right] = \frac{1}{T}\sum_{i=1}^{T}\frac{(x_i-\mu)^3}{\sigma^3}$$
$$\text{kurtosis(X)} = E\left[\left(\frac{X-\mu}{\sigma}\right)^4\right] = \frac{1}{T}\sum_{i=1}^{T}\frac{(x_i-\mu)^4}{\sigma^4} \quad (1)$$

3 Experimental and results

The experimental environment of this article was based on the server: Ubuntu 16.04 LTS was used as the operating system with Intel Xeon e5-2650 V4 processor and Nvidia GTX 1080 Ti GPU. The memory is 63 GB. Pytorch was used to build the Deep Pyramid convolutional neural network, and Python3.6 was used as the programming tool.

3.1 Database

The data used in this experiment were collected from surgery patient in hospital of China from 2014 to 2018. The number of patients in this data was 4700, including 536 positive patients, 4164 negative patients, 2718 female patients, and 1982 male patients. The age, gender and label distribution were shown in Figs. 3-5. The test set is randomly selected from the data set and accounts for 20% of the data set. In order to make the results statistically significant, the experimental method adopted 10-fold cross validation.

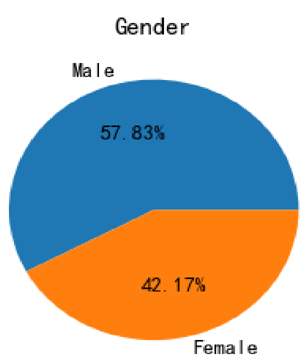


Figure 3: Gender distribution

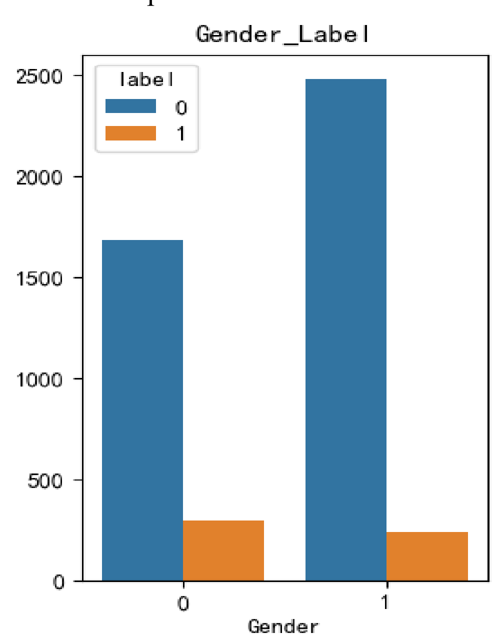


Figure 4: Histogram of patients' gender and label

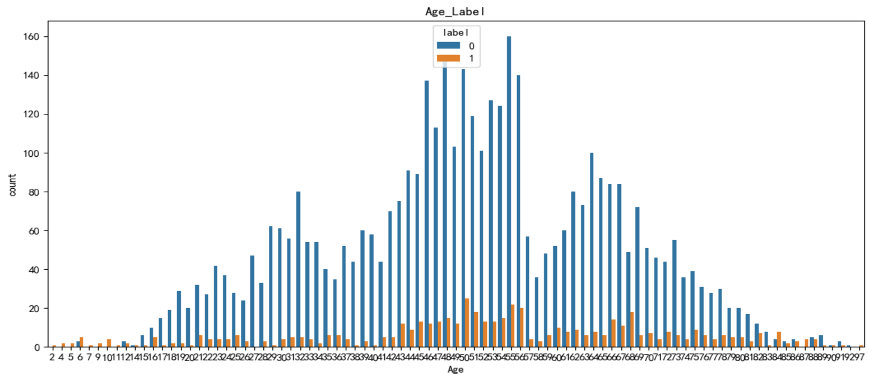


Figure 5: Statistical chart of patient data set age and label

3.2 Performance evaluation

In this paper, the prediction of heart failure is modeled and analyzed as a dichotomy problem, and the performance of the model can be evaluated through the confusion matrix. According to the ground truth and prediction, true positive (TP), true negative (TN), false positive (FP) and false negative (FN) can be classified. Precision, recall and F1 Score were used to test the accuracy of classification.

4 Experimental and discussion

We report the experiments with DPCNNs and XGBOOST in comparison with previous models and alternatives. The specific experimental description is as follows. The code is publicly available on the internet.

4.1 Preoperative test and demographic data for heart failure prediction

Only preoperative test and demographic data were used to predict heart failure. XGBOOST model was compared with Gaussian Bayesian model, logistic regression model, SVM model, random forest model and GBDT model. The experimental results using Python's sklearn library are shown in Tab. 4, and the ROC diagram of the model is shown in Figs. 6 and 7.

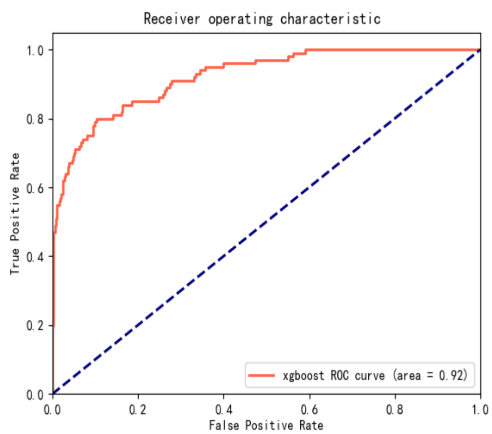


Figure 6: XGBOOST ROC curve

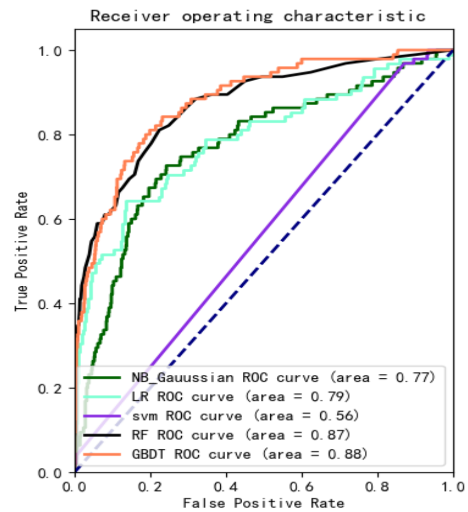


Figure 7: ROC curve of comparative model

Table 4: Performance comparison of models with different classifications

Model	Weighted avg precision(pos/neg)		Weighted avg recall (pos/neg)		Weighted avg f1-score (pos/neg)		Roc
XGBOOST	**0.94**	0.96	**0.94**	0.43	**0.93**	0.59	0.92
		0.94		0.99		0.97	
NB_Gauussian	0.87	0.22	0.66	0.77	0.72	0.34	0.77
		0.94		0.64		0.77	
LR	0.86	0.55	0.89	0.17	0.86	0.26	0.79
		0.90		0.98		0.94	
svm	0.94	1.00	0.52	0.03	0.50	0.06	0.56
		0.89		1.00		0.94	
RF	0.92	0.92	0.92	0.37	0.91	0.52	0.87
		0.92		1.00		0.96	
GBDT	0.91	0.72	0.92	0.45	0.91	0.55	0.88

Due to the imbalance of data set (negative patient: 4164, positive patient: 536), the positive recall rate (sensitivity) of all models is low, but the weighted average accuracy (94%), weighted average recall rate (94%), F1 score (93%) and ROC value (92%) based on XGBOOST model are higher than other models. In addition, because there are a lot of missing values in the data set, the missing values are filled with 0 when using LR and SVM models, and the effect is worse than the tree model. Therefore, based on the preoperative patient test data, the tree model is more suitable for patients with missing value.

4.2 Prediction of heart failure with intraoperative monitoring data

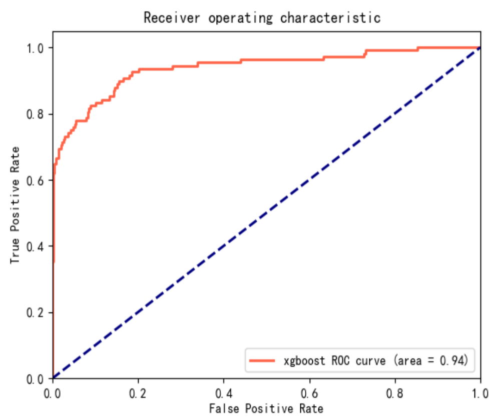


Figure 8: XGBOOST ROC curve

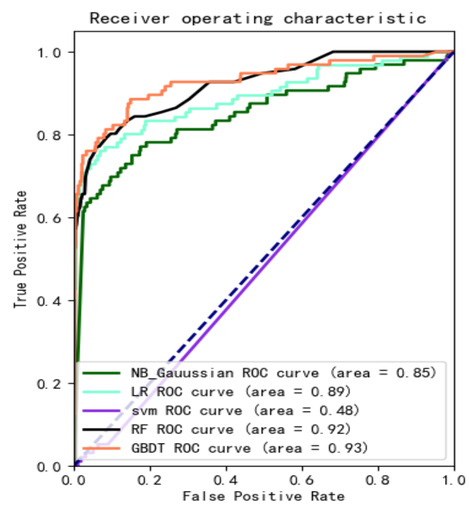


Figure 9: ROC curve of comparative model

In this paper, the outliers time and statistical characteristics of monitoring data were extracted, and the model in Experiment 1 was used to predict heart failure. The experimental results are shown in Tab. 5. The ROC curve of the model is shown in Figs. 8 and 9.

The experimental results show that the performance of XGBOOST model is the best (including weighted average accuracy (95%), weighted average recall rate (95%), F1 score (95%) and ROC value (94%)), but the positive recall rate is lower (XGBOOST (0.62), NB_Gauussian (0.58), LR (0.60), SVM (0.3), RF (0.54), GBDT (0.61). In order to improve the sensitivity of heart failure prediction model, the following experiments are carried out.

Table 5: Performance comparison of models with different classifications

Model	Weighted avg precision(pos/neg)		Weighted avg recall (pos/neg)		Weighted avg f1-score (pos/neg)		Roc
XGBOOST	**0.95**	0.95 0.94	**0.95**	0.62 0.99	**0.95**	0.75 0.97	**0.94**
NB_Gauussian	0.91	0.71 0.94	0.91	0.58 0.96	0.91	0.64 0.95	0.85
LR	0.95	0.94 0.95	0.95	0.60 0.93	0.94	0.75 0.95	0.89
svm	0.75	0.20 0.87	0.87	0.30 1.00	0.81	0.10 0.93	0.48
RF	0.94	0.94 0.93	0.94	0.54 1.00	0.93	0.69 0.96	0.92
GBDT	0.95	0.96 0.94	0.95	0.61 1.00	0.94	0.75 0.97	0.93

4.3 Prediction of heart failure based on unstructured text data of Preoperative patients

Firstly, DPCNN is used to extract the text features of patients. Two convolution layers of blocks are used. Each convolution core is 3 × 3, the size of pooling layer is 2 × 2, and each layer has 250 convolution cores. Finally, a full connection layer is connected for heart failure classification and prediction. Tab. 6 of experimental results, ROC curve as shown in Fig. 10.

It can be seen from the experimental results that the unstructured text data based on the preoperative examination of patients can better predict the postoperative heart failure of patients (sensitivity 90%, specificity 97%, ROC 98%). Analysis reason: the preoperative text includes the diagnosis information of patients and the conclusion of ECG examination. The positive patient's ECG examination and preoperative diagnosis information contain the disease information of patients' heart, while the negative patient's preoperative diagnosis and ECG examination conclusion are normal, so the text-based method is very effective for the prediction of heart failure. In order to further integrate the patient's diagnosis information, this paper integrates all the patient's diagnosis information, and carries out the following experiments.

Table 6: Prediction results of heart failure

	precision	recall	F1-score
negative	0.9881	0.9733	0.9806
positive	0.8144	0.9086	0.8589
macro avg	0.9012	0.9410	0.9198
weighted avg	0.9682	0.9660	0.9668

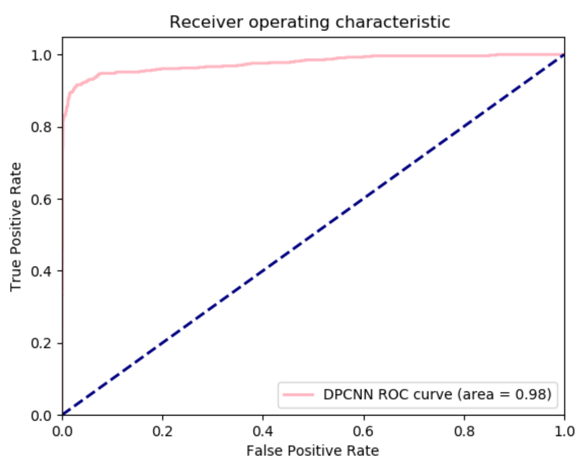
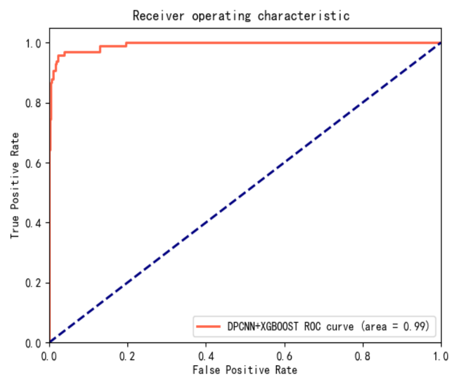


Figure 10: ROC curve of DPCNN model

4.4 Prediction of heart failure by fusion of preoperative and intraoperative data



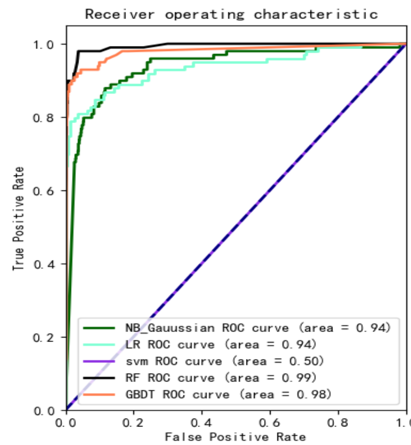


Figure 11: DPCNN+XGBOOST ROC curve **Figure 12:** ROC curve of comparative model

Based on DPCNN and XGBOOST, the patients' heart failure was predicted by fusing the numerical and textual features of preoperative test and the numerical features of intraoperative monitoring. In order to further verify the effectiveness of DPCNN-XGBOOST model, the experimental results are compared with Bayes, logistic regression, RF, GBDT and other models. The results are shown in Tab. 7 and ROC is shown in Figs. 11 and 12.

Table 7: Performance comparison of models with different classifications

Model	Weighted avg precision(pos/neg)		Weighted avg recall (pos/neg)		Weighted avg f1-score (pos/neg)		Roc
DPCNN+XGBOOST	**0.98**	**0.88**	**0.98**	**0.93**	**0.98**	**0.90**	**0.99**
		0.99		0.98		0.99	
DPCNN+NB	0.93	0.74	0.94	0.64	0.93	0.69	0.94
		0.96		0.97		0.96	
DPCNN+LR	0.96	0.91	0.96	0.70	0.96	0.79	0.94
		0.96		0.99		0.98	
DPCNN+svm	0.80	0.20	0.89	0.30	0.84	0.10	0.5
		0.89		1.00		0.94	
DPCNN+RF	0.97	0.88	0.97	0.77	0.97	0.84	0.97
		0.98		0.99		0.98	
DPCNN+GBDT	0.97	0.87	0.97	0.88	0.97	0.87	0.97
		0.99		0.98		0.98	

Fig. 13 shows that DPCNN-XGBOOST is the best predictor of heart failure (including weighted average accuracy (98%), weighted average recall rate (98%), F1 score (98%) and ROC value (99%)). Compared with preoperative test data, intraoperative monitoring data and preoperative text attributes, the sensitivity of the model increased by 50%, 31% and 3% respectively. It can be seen from the experimental results that the combination of DPCNN and XGBOOST classifier can improve the prediction of heart failure, and also verify the effectiveness of the proposed model.

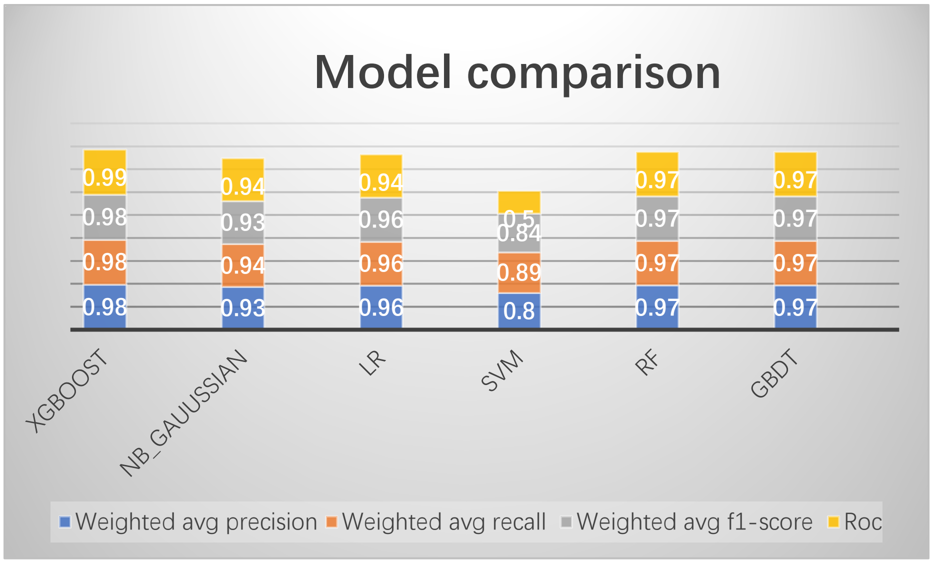


Figure 13: Model performance comparison

5 Discussion

Heart failure in the perioperative period is one of the most significant causes of postoperative death of patients. At present, most studies only focused on the numerical data of preoperative test or intraoperative monitoring, whose prediction has poor timeliness, low accuracy, and non-fusion of patient heterogeneity data. Thus, in order to make the results of research and analysis more accurate and more convenient to apply to practice, in this paper, we use Deep Pyramid Convolutional Neural Networks (DPCNN) and Extreme Gradient Boosting (XGBOOST) method to model the risk prediction of heart failure after operation based on the preoperative and Intraoperative medical data of patients. The proposed model is verified in four experiments. The results of Experiment 1 show that XGBOOST model has a better classification effect on patients' numerical structural data. The results of Experiment 2 showed that the monitoring data of patients during operation had certain precursory information for the prediction of heart failure. Experiment 3 shows that the text-based data of preoperative examination can better describe the patient's condition. In Experiment 4, the combination of numerical and textual data can further improve the accuracy of heart failure prediction. There are still several ways to improve this work, which is also the direction of our future work. First of all, the word vector used for text feature extraction is based on Tencent's trained word vector. In order to improve the description of patients' condition, we can train medical field word vector based on electronic medical record data for feature extractions. Secondly, the prediction results of the model are still black boxes for doctors. In order to improve the interpretability of the model, it is necessary to further integrate the deep neural network and the tree model, and use the tree model to interpret the results. Finally, the model is only tested in the event of heart failure at present, and we will try to apply the model to other critical diseases (liver failure, renal failure, etc.). In the future, we hope that such methods will be used to provide medical staff with the support to improve decision making for surgical surgeon.

6 Conclusion

This paper discusses an interesting medical problem. The prediction of heart failure in perioperative patients is a complex process, and whether or not adverse events occur after surgery depends entirely on the doctor experience and judgment, and the prediction accuracy of long-term and experienced doctors is higher, but the prediction accuracy of doctors with short working hours or inexperience is slightly lower. Additionally, the judgment has lag and the evaluation result cannot be applied directly and effectively, which is a serious problem. Therefore, based on the machine learning method, this paper establishes a prediction model of heart failure adverse events in patients to predict the risk of critical disease with the preoperative index of any patient. Firstly, the data of preoperative patients are preprocessed, and the patient data is divided into the structural data of numerical data and the unstructured data of textual data. The numerical data of the patient is constructed by the gradient boosting tree model, and the textual features of the patients are extracted by Deep Pyramid Convolutional Neural Networks of the text-based data. Fusion of textual features and numerical features, and finally through a XGBOOST model to predict the patient heart failure illness.

Availability of Data and Materials: Not applicable. The data and materials are related to the privacy of patients and cannot be disclosed at this time.

Funding Statement: This study was approved by the Ethics Committee of the First Affiliated Hospital of Army Medical University, PLA, and the Approved No. of ethic committee is KY201936. This work is supported by the National Key Research & Development Plan of China (2018YFC0116704) in data collection. In addition, it is supported by Chongqing Technology Innovation and application research and development project (cstc2019jscx-msxmx0237) in the design of the study.

Conflicts of Interest: The authors declare that they have no competing interests.

References

Acharya, U. R.; Fujita, H.; Sudarshan, V. K.; Oh, S. L.; Muhammad, A. et al. (2017): Application of empirical mode decomposition (EMD) for automated identification of congestive heart failure using heart rate signals. *Neural Computing and Applications*, vol. 28, no. 10, pp. 3073-3094.

Attia, Z. I.; Kapa, S.; Lopezjimenez, F.; Mckie, P. M.; Ladewig, D. J. et al. (2019): Screening for cardiac contractile dysfunction using an artificial intelligence-enabled electrocardiogram. *Nature Medicine*, vol. 25, no. 1, pp. 70-74.

Chen, T.; Guestrin, C. (2016): Xgboost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785-794.

Chen, W.; Zheng, L.; Li, K.; Wang, Q.; Liu, G. et al. (2016): A novel and effective method for congestive heart failure detection and quantification using dynamic heart rate variability measurement. *PLoS One*, vol. 11, no. 11.

Chen, Y. W.; Sun, Q. L.; Zhong, K. H. (2018): Semi-supervised spatio-temporal cnn for recognition of surgical workflow. *Eurasip Journal on Image and Video Processing*, pp. 9.

Chen, Y. W.; Wang, P.; Lin, X. G.; Zhong, K. H. (2019): Prediction of in-hospital mortality risk in intensive care unit based on deep neural network. *Basic and Clinical Pharmacology and Toxicology*, vol. 125, pp. 4.

Chen, Y.; Zhong, K.; Wang, F.; Wang, H.; Zhao, X. (2018): Surgical workflow image generation based on generative adversarial networks. *International Conference on Artificial Intelligence*, pp. 82-86.

Choi, E.; Schuetz, A.; Stewart, W. F.; Sun, J. (2017): Using recurrent neural network models for early detection of heart failure onset. *Journal of the American Medical Informatics Association*, vol. 24, no. 2, pp. 361-370.

Esteva, A.; Robicquet, A.; Ramsundar, B.; Kuleshov, V.; Depristo, M. A. et al. (2019): A guide to deep learning in healthcare. *Nature Medicine*, vol. 25, no. 1, pp. 24-29.

Gottesman, O.; Johansson, F. D.; Komorowski, M.; Faisal, A. A.; Sontag, D. et al. (2019): Guidelines for reinforcement learning in healthcare. *Nature Medicine*, vol. 25, no. 1, pp. 16-18.

Gurovich, Y.; Hanani, Y.; Bar, O.; Nadav, G.; Fleischer, N. et al. (2019): Identifying facial phenotypes of genetic disorders using deep learning. *Nature Medicine*, vol. 25, no. 1, pp. 60-64.

Hannun, A.; Rajpurkar, P.; Haghpanahi, M.; Tison, G. H.; Bourn, C. et al. (2019): Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nature Medicine*, vol. 25, no. 1, pp. 65-69.

Johnson, R.; Zhang, T. (2017): Deep pyramid convolutional neural networks for text categorization. *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, Vancouver, Canada, pp. 562-570.

Koerkamp, B. G.; Stijnen, T.; Weinstein, M. C.; Hunink, M. G. M. (2011): The combined analysis of uncertainty and patient heterogeneity in medical decision models. *Medical Decision Making*, vol. 31, no. 4, pp. 650-661.

Liu, P.; Qiu, X.; Huang, X. (2016): Recurrent neural network for text classification with multi-task learning. *Computation and Language*.

Masetic, Z.; Subasi, A. (2016): Congestive heart failure detection using random forest classifier. *Computer Methods and Programs in Biomedicine*, vol. 130, pp. 54-64.

Melillo, P.; Izzo, R.; Orrico, A.; Scala, P.; Attanasio, M. et al. (2015): Automatic prediction of cardiovascular and cerebrovascular events using heart rate variability analysis. *PLoS One*, vol. 10, no. 3.

Shameer, K.; Johnson, K. W.; Yahi, A.; Miotto, R.; Li, L. et al. (2017): Predictive modeling of hospital readmission rates using electronic medical record-wide machine learning: a case-study using mount sinai heart failure cohort. *Pacific Symposium on Biocomputing*, vol. 22, pp. 276-287.

Song, Y.; Shi, S.; Li, J.; Zhang, H. (2018): Directional skip-gram: explicitly distinguishing left and right context for word embeddings. *North American Chapter of the Association for Computational Linguistics*, vol. 2, pp. 175-180.

Ye, J.; Chow, J.; Chen, J.; Zheng, Z. (2009): Stochastic gradient boosted distributed decision trees. *Proceedings of the 18th ACM Conference on Information and Knowledge Management*, pp. 2061-2064.

Yu, S. N.; Lee, M. Y. (2012): Conditional mutual information-based feature selection for congestive heart failure recognition using heart rate variability. *Computer Methods and Programs in Biomedicine*, vol. 108, no. 1, pp. 299-309.

Zhang, Y.; Wallace, B. C. (2017): A sensitivity analysis of (and practitioners' guide to) convolutional neural networks for sentence classification. *International Joint Conference on Natural Language Processing*, vol. 1, pp. 253-263.

Zheng, B.; Zhang, J.; Yoon, S. W.; Lam, S. S.; Khasawneh, M. et al. (2015): Predictive modeling of hospital readmissions using metaheuristics and data mining. *Expert Systems with Applications*, vol. 42, no. 20, pp. 7110-7120.