MULTI-VIEW LEARNING BASED ON NON-REDUNDANT FUSION FOR ICU PATIENT MORTALITY PREDICTION

Yifan Wang, Ying Lan

School of Electronic and Computer Engineering, Peking University

ABSTRACT

In medical data research, mortality prediction in intensive care units (ICUs) has always been a research hotspot. The Apache-II death prediction system relies on scoring rules. Despite its extensive application, it also has apparent shortcomings, with its accuracy rate decreasing over time. In recent years, researchers have proposed machine learning and deep learning algorithms to establish predictive models for ICUs. Those predicting from a single perspective cannot fully apply multiple sources of information, while the fusion of multiple perspectives may produce much redundant information. Therefore, this paper proposes a multi-view fusion method based on non-redundant information learning, applying it to ICU patient mortality prediction. Collaboratively, it applies consistency and complementarity among different views to discover internal data patterns accurately and improve the effectiveness of data analysis. Experimental results indicate that the accuracy of this method in predicting the mortality of critically ill patients in ICUs reaches 90.43%, around 5.25% higher than the existing model.

***Index Terms*—** Multi-view learning, data fusion, heterogeneous information, non-redundant information learning

1. INTRODUCTION

Intensive care units (ICUs) in China have a history spanning only 30 years[1]; consequently, they are still unable to meet the enormous demand in major public medical incidents. Therefore, it is significant to improve the efficiency of using ICU wards. By estimating the mortality of patients and the severity of patients' conditions, those with more urgent conditions can realize priority ICU resources to save more lives[2, 3]. Therefore, studying mortality prediction in ICU wards is of much significance.

Multi-view learning is a vital technology that utilizes multi-source information[4]. Since specific single-view data cannot fully describe the information from all examples, it is necessary to obtain information from different perspective[5]. However, the multi-source and heterogeneous characteristics of data make the correlation among different views complex and challenging to predict, and even domain experts are difficult to apply when facing complex multi-source information effectively. For instance, ICU patient data has several sources, including basic patient information, patient vital signs measurement, and imaging reports[6]. Moreover, different information systems follow varying standards and protocols, data types, and consistency and complementarity among information. Thus, multi-view learning has an application value in this field.

Canonical correlation analysis (CCA) and its kernel function extension method are the representative techniques for multi-view learning. The deep CCA method [7] applies a general strategy to learn a high-level coupled multi-view feature representation after learning view-related features at a superficial level. It is not possible to learn good associations between multi-view data. Meanwhile, the multi-view SVM [8, 9] method combines data and label information using a classifier or regressor; it normalizes the classifier or regressor to make the results obtained by multiple views as consistent as possible. Multi-view Maximum Entropy Discrimination (MVMED) [10] proposes an interval consistency strategy, utilizing the margin between the sample and the hyperplane to depict the relationship between the model and the data. It enables the potential consistency of the classification results. Memory Fusion Network [11] employs the memory network to simulate the modal's internal information and obtains the relationship of the multi-modal information sequence over time through a multi-view gated memory mechanism.

Although the above feature fusion-based methods have shown superiority in disease prediction and medical image detection tasks, some information from varying views is essentially consistent in terms of ICU patient death prediction, including heart disease patients' admission symptoms and heart rate data [12]. Furthermore, mapping features to high-dimensional space can lead to data redundancy. During multi-view fusion, existing fusion methods cannot solve information redundancy.

Our contributions can be specified as follows: we designed a multi-view non-redundant information fusion network to predict ICU patient mortality.

(1) We extracted three views of ICU patient data from MIMIC III database, designed intra-view and inter-view layer to realize view-specific and cross-view interactions.

(2) We proposed the non-redundant information learning module to remove the essentially consistent from varying

Table 1. Formula Variable Description

N	Set of three views
M	Total number of words in a sentence
x_{n_m}	Representation of m-th word in view n
d_n	Dimension of mode n
x_n	Representation of view n
d_c	Dimension of LSTM hidden layer

views in terms of ICU patient death prediction.

(3) We designed a fusion module based on self-attention mechanism to fuse each view and yield the final prediction output. The experiment results demonstrated that the proposed method outperforms other state-of-the-art methods.

2. METHOD

2.1. Feature Presentation Layer

The extracted patient ICU data includes physiological indicators, treatment records, and hospitalization records after entering the ICU. Treatment records and hospitalization records are text data, which are represented by GloVe word vectors [13]. In the training phase, word vectors in the utterance are fine-tuned through backpropagation, and unregistered words are randomly initialized to obtain a sequence representation. Physiological indicators are numerical data, which are converted to sequence representation after normalization. The general expression formulas for three views are as follows.

$$N = \{p, t, h\}$$
$$H_n^u = \left[h_{n_m}^u : m \leq M, h_{n_m}^u \in R^{2\times d_c}, n = N\right]$$

N is a collection of three views of physiological indicators, treatment records, and hospitalization records. $n = N$ means that n covers the information of all views. Table 1 explains the variables in the formula.

2.2. Intra-view Layer

Use the private two-way LSTM layer to model the context information of a unit view.

$$h_{n_m}^u = \overrightarrow{LSTM}\left(x_{n_m}\right) \oplus \overleftarrow{LSTM}\left(x_{n_m}\right)$$

where u is the mark of a unit view, $\oplus$ splicing the forward information and backward information of the sentence, $h_{n_m}^u$ is the m-th word of view n. x_{n_m} is output through the hidden layer of the bidirectional LSTM. H_n^u is the internal information sequence representation of view n. Therefore, the single-view internal information of physiological indicators, treatment records, and hospitalization records are represented as H_p^u, H_t^u, H_h^u.

2.3. Inter-view Layer

The inter-view layer models the dynamic interaction between dual-view and three-view.

Dual-view Interaction The patient's physiological data, treatment records, and hospitalization information can form three different interaction modes, for which three shared two-way LSTM layers are introduced. Take physiological data-treatment records as an example to illustrate the dynamic interaction process of dual views:

$$h_{n_m}^{b_1} = \overrightarrow{LSTM}\left(x_{n_m}\right) \oplus \overleftarrow{LSTM}\left(x_{n_n}\right)$$
$$H_n^{b_1} = \left[h_{n_m}^{b_1} : m \leq M, h_{n_m}^{b_1} \in R^{2\times d_c}, n = N - \{a\}\right]$$

Where b_1 is the mark of the dynamic interaction between physiological data and treatment record, $h_{n_m}^{b_1}$ is the hidden layer output of the m-th word for view n after BiLSTM[14], and $H_n^{b_1}$ is the sequence representation of view n in the discourse. After dynamic interaction, the physiological data and treatment records are expressed as $H_p^{b_1}$, $H_t^{b_1}$.

Similarly, after the interaction of physiological data and hospitalization records, $H_p^{b_2}$ and $H_h^{b_2}$ can be obtained, and $H_t^{b_3}$ and $H_h^{b_3}$ can be obtained through the interaction of treatment records and hospitalization records.

Tri-view interaction Use the shared network layer to model physiological data-treatment records-hospitalization records, the formula are as follows.

$$h_{n_m}^t = \overrightarrow{LSTM}\left(x_{n_m}\right) \oplus \overleftarrow{LSTM}\left(x_{n_m}\right)$$
$$H_n^t = \left[h_{n_m}^t : m \leq M, h_{n_m}^t \in R^{2\times d_c}, n = N\right]$$

t is the mark of triple-view interaction, $h_{n_m}^t$ is the hidden layer output of m-th word x_{n_m} for view n after BiLSTM[14], and H_n^t is the sequence representation of view n. After the interaction of the three views of physiological data, treatment record and hospitalization record, the three views can be expressed as H_p^t, H_t^t, H_h^t.

2.4. Non-redundant Information Learning Layer

Inspired by recent research work [15, 16], we use orthogonal loss constraints to punish redundant information in the purpose of removing redundancy. Orthogonal loss constraint can extract more effective information from the three-view fusion information. Taking physiological data and treatment records as examples, the formula are as follows.

$$\text{Loss}^p = \frac{1}{M}\left\|\left(H_p^t\right)^T \cdot H_t^t\right\|^2$$
$$\text{Loss}^t = \frac{1}{M}\left\|\left(H_t^t\right)^T \cdot H_p^t\right\|^2$$

M is the total amount of training data, $Loss^p$ and $Loss^t$ are the orthogonal loss constraints from the views of physiological data and treatment records.

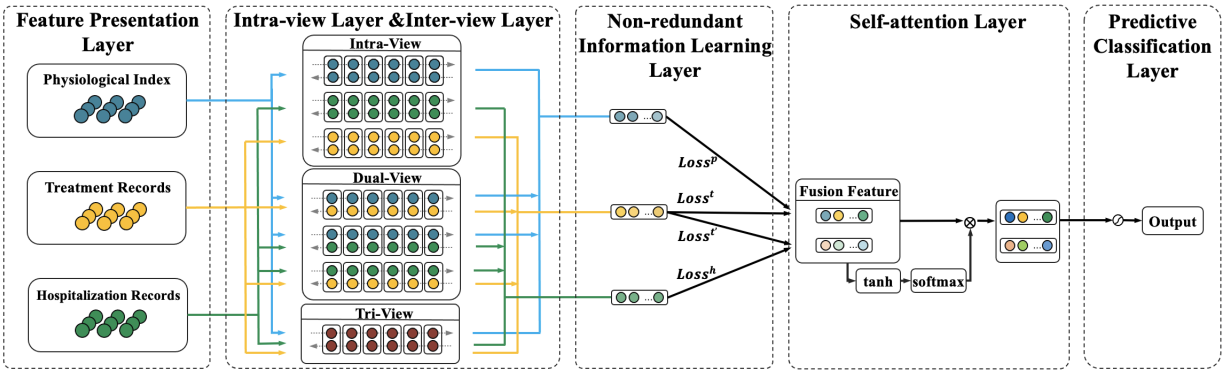

Fig. 1. The architecture of Multi-view Non-redundant Information Fusion Network

2.5. Self-attention Layer

The attention distribution of the task is calculated by the self-attention mechanism [17], the formula are as follows.

$$A^f = \text{softmax}\left(W_p \cdot \left(\tanh\left(W_q \cdot C^f\right)\right)\right)$$

$$F^f = A^f \cdot C^f$$

$$R^f = mp\left(\tanh\left(W_r \cdot F^f + b_r\right)\right)$$

C^f represents the fusion representation of the three views after being processed by the non-redundant information learning layer. A^f Represents the attention distribution of the multi-view fusion representation in the task. The weighted representation is obtained by weighting the original information. The dimensionality reduction is performed through pooling, W_p, W_q, and W_r is the weight of the network layer, b_r is the bias of the network layer. The final weighted fusion of multi-view information is expressed as R^f.

2.6. Predictive Classification Layer

Use the activation functions $tanh$ and $sigmoid$ to classify and predict the fusion representation of multiple views.

$$\hat{y}^f = \sigma\left(W_q \cdot \left(\tanh\left(W_p \cdot R^f + b_p\right)\right) + b_q\right)$$

R^f is the classification result obtained after multi-view information fusion, W_p and W_q, b_p and b_q are the weight and bias of the network layer.

2.7. Optimization Strategy

In the process of model training, use the minimized cross entropy error to optimize the multi-view fusion information prediction.

$$Loss(\hat{y}, y) = -\sum_{s=1}^{S}\sum_{c=1}^{C} y_s^c \cdot \log \hat{y}_s^c$$

Table 2. Extracted ICU Patient Statistics Table

	Population	Survival	Dead
Number	32310	28079	4321
Average Age	63.2	62.6	67.1
Gender(M)	17965	15612	2353

y_s^c is the true label of the sample s, $\hat{y}_s^c$ is the probability that the model predicts s as the category c, S is the total number of training samples, C is the number of target categories, and Adadelta is used in the experiment to optimize parameters.

3. EXPERIMENT

3.1. Data and Criterion

MIMIC-III (Medical Information Mart for Intensive Care III [18]) is a large free database containing medical data of patients in the intensive care unit of Beth Israel Dikang Medical Center (BIDMC) from 2001 to 2012. The MIMIC-III database covers demographics, vital signs measurement, laboratory test structure, nursing records and other information.

This article selects data from three perspectives: physiological indicators, treatment records, and hospitalization records[19]. Physiological indicators include patient mean arterial pressure, heart rate, respiratory rate, and other 20 check items. The treatment record includes diagnosis code, surgical operation, and other 17 check items. The hospitalization record includespatient number, race, reason for admission and other 11 check items. Table 2 shows the statistics of the number of ICU patients.

According Table 2, there is not much difference in the proportion of male patients and female patients, but the number of surviving patients is much larger than the number of dead patients, and the positive and negative samples are unbalanced. Therefore, this article uses the Synthetic Minority Oversamping Technique [20] to balance the data set. Synthe-

Table 3. Results statistics of different models(%)

Methods	Class	Acc.	Prec.	Recall	F1
CCA	D	76.53	65.61	65.01	65.31
	S		82.07	82.45	82.26
SVM	D	85.18	80.89	83.74	82.29
	S		87.18	85.83	86.50
MVMED	D	82.30	74.91	77.77	76.31
	S		87.22	85.76	86.48
MFN	D	83.30	74.35	76.55	75.43
	S		87.69	86.21	86.94
MNRIFN	D	90.43	83.33	86.13	84.71
	S		93.74	92.35	93.04

size new samples based on minority samples and add them to the data set to increase the number of minority samples. For each sample x in the minority sample set, calculate the Euclidean distance from x to all other minority samples to obtain its k nearest neighbors. Set the sampling magnification N, randomly select several samples from k nearest neighbors, assuming that the selected nearest neighbor is x_n. For each x_n, the new sample is $x_{new} = x + \text{rand}(0,1) \cdot (\tilde{x} - x)$

The model in this paper is binary classification model, and the commonly used evaluation criteria are accuracy, precision, recall, and F1 value.

3.2. Performance Versus Other Methods

This article applies the MNRIFN model to predict and compare it with the four strong baselines: CCA, SVM, MVMED, and MFN. The SVMCCA algorithm completes data fusion through CCA. SVM classifies the samples by finding a hyperplane. MVMED uses the margin between the sample and the hyperplane to describe the model-data relationship and applies the potential consistency of the classification results. Finally, MFN uses the memory network to simulate the modal's internal information and obtains the relationship of the multimodal information sequence over time through a multi-view gated memory mechanism.

Table 3 shows that, among the four comparison models, SVM exhibited the best model performance. SVM mapped the data of surviving and dead patients into a high-dimensional space, and their data could present good separability. In comparison, the effect of CCA was poor. Interpreted from a data level, the physiological data of patients were very close; that is, the data ranges of surviving and dead patients overlapped, leading to a decrease in the model's performance. The accuracy of the MNRIFN model employed in this paper reached 90.43%, and all evaluation indicators performed well. It indicated that in multi-view information fusion, MNRIFN still retained a good predictive capability while removing non-redundant information.

3.3. Ablation Experiment

To illustrate the effectiveness of the non-redundant information learning layer and the self-attention layer, the MNRIFN was compared. The method in this paper was further implemented without orthogonal loss constraints (MNRIFN w/o OC) and a self-attention mechanism (MNRIFN w/o SA). The experimental results indicated that the accuracy of death prediction had decreased without using orthogonal loss constraints to enable non-redundant data. It is because there was redundant information between specific medical records and physiological indicators. After data fusion, the role of specific indicators magnified, and learning other features became disturbed. Without the self-attention mechanism, the practical effect was slightly worse than the complete MNRIFN model, primarily because the attention mechanism focused on the details according to the predicted targets.

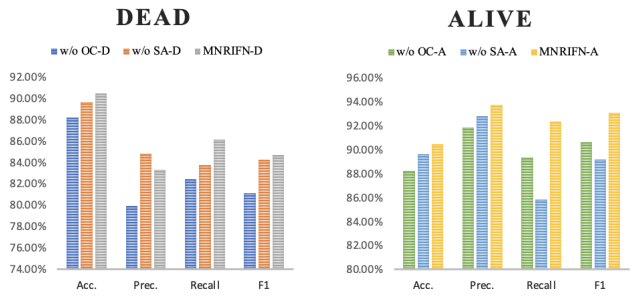

Fig.2. Results statistics of ablation experiment (Dead&Alive)

The results indicated that without using the orthogonal loss constraints to de-redundate the data, the accuracy of death prediction had decreased, since there was redundant information between certain medical records and the physiological indicators. After the fusion of the data, the role of certain indicators was magnified and the learning of other features was also disturbed. Without using the self-attention mechanism, the experimental effect was found to be slightly worse than the complete model of the MNRIFN, since the attention mechanism paid attention to the details according to the predicted target.

4. CONCLUSION

This paper proposes a multi-view fusion method based on non-redundant information learning, applying orthogonal loss constraints to learn non-redundant dynamic interaction information within and between views, and combining with the self-attention mechanism to enhance the concerned part capture. The said algorithm solves redundant information generated by multi-view data fusion in ICU patient mortality prediction. The encouraging empirical results on MIMIC-III have demonstrated that the proposed method outperforms other state-of-the-art methods.

5. REFERENCES

[1] Ruoran Li, Caitlin Rivers, Qi Tan, Megan B Murray, Eric Toner, and Marc Lipsitch, "The demand for inpatient and icu beds for covid-19 in the us: lessons from chinese cities," *MedRxiv*, 2020.

[2] Brett Ley, Harold R Collard, and Talmadge E King Jr, "Clinical course and prediction of survival in idiopathic pulmonary fibrosis," *American journal of respiratory and critical care medicine*, vol. 183, no. 4, pp. 431–440, 2011.

[3] Oscar Luaces, José Ramón Quevedo, Francisco Taboada, Guillermo M Albaiceta, Antonio Bahamonde, and Asturias-Spain Asturias-Spain, "Prediction of probability of survival in critically ill patients optimizing the area under the roc curve.," in *IJCAI*, 2007, pp. 956–961.

[4] Chang Xu, Dacheng Tao, and Chao Xu, "A survey on multi-view learning," *arXiv preprint arXiv:1304.5634*, 2013.

[5] Jing Zhao, Xijiong Xie, Xin Xu, and Shiliang Sun, "Multi-view learning overview: Recent progress and new challenges," *Information Fusion*, vol. 38, pp. 43–54, 2017.

[6] Jaspreet Kaur Mann, Farhad Kaffashi, Benjamin Vandendriessche, Frank J Jacono, and Kenneth Loparo, "Data collection and analysis in the icu," *Neurocritical Care Informatics*, pp. 111–134, 2020.

[7] Galen Andrew, Raman Arora, Jeff Bilmes, and Karen Livescu, "Deep canonical correlation analysis," in *International conference on machine learning*. PMLR, 2013, pp. 1247–1255.

[8] Xijiong Xie and Shiliang Sun, "Multi-view twin support vector machines," *Intelligent Data Analysis*, vol. 19, no. 4, pp. 701–712, 2015.

[9] Jason Farquhar, David Hardoon, Hongying Meng, John S Shawe-Taylor, and Sandor Szedmak, "Two view learning: Svm-2k, theory and practice," in *Advances in neural information processing systems*, 2006, pp. 355–362.

[10] Shiliang Sun and Guoqing Chao, "Multi-view maximum entropy discrimination," in *Twenty-third international joint conference on artificial intelligence*, 2013.

[11] Amir Zadeh, Paul Pu Liang, Navonil Mazumder, Soujanya Poria, Erik Cambria, and Louis-Philippe Morency, "Memory fusion network for multi-view sequential learning," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2018, vol. 32.

[12] Christina Twyman-Saint Victor, Andrew J Rech, Amit Maity, Ramesh Rengan, Kristen E Pauken, Erietta Stelekati, Joseph L Benci, Bihui Xu, Hannah Dada, Pamela M Odorizzi, et al., "Radiation and dual checkpoint blockade activate non-redundant immune mechanisms in cancer," *Nature*, vol. 520, no. 7547, pp. 373–377, 2015.

[13] Jeffrey Pennington, Richard Socher, and Christopher D Manning, "Glove: Global vectors for word representation," in *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, 2014, pp. 1532–1543.

[14] Tao Chen, Ruifeng Xu, Yulan He, and Xuan Wang, "Improving sentiment analysis via sentence type classification using bilstm-crf and cnn," *Expert Systems with Applications*, vol. 72, pp. 221–230, 2017.

[15] Behnaz Nojavanasghari, Deepak Gopinath, Jayanth Koushik, Tadas Baltrušaitis, and Louis-Philippe Morency, "Deep multimodal fusion for persuasiveness prediction," in *Proceedings of the 18th ACM International Conference on Multimodal Interaction*, 2016, pp. 284–288.

[16] Mohit Iyyer, Varun Manjunatha, Jordan Boyd-Graber, and Hal Daumé III, "Deep unordered composition rivals syntactic methods for text classification," in *Proceedings of the 53rd annual meeting of the association for computational linguistics and the 7th international joint conference on natural language processing (volume 1: Long papers)*, 2015, pp. 1681–1691.

[17] Han Zhang, Ian Goodfellow, Dimitris Metaxas, and Augustus Odena, "Self-attention generative adversarial networks," in *International conference on machine learning*. PMLR, 2019, pp. 7354–7363.

[18] Alistair EW Johnson, Tom J Pollard, Lu Shen, H Lehman Li-Wei, Mengling Feng, Mohammad Ghassemi, Benjamin Moody, Peter Szolovits, Leo Anthony Celi, and Roger G Mark, "Mimic-iii, a freely accessible critical care database," *Scientific data*, vol. 3, no. 1, pp. 1–9, 2016.

[19] Mengling Feng, Jakob I McSparron, Dang Trung Kien, David J Stone, David H Roberts, Richard M Schwartzstein, Antoine Vieillard-Baron, and Leo Anthony Celi, "Transthoracic echocardiography and mortality in sepsis: analysis of the mimic-iii database," *Intensive care medicine*, vol. 44, no. 6, pp. 884–892, 2018.

[20] Nitesh V Chawla, Kevin W Bowyer, Lawrence O Hall, and W Philip Kegelmeyer, "Smote: synthetic minority over-sampling technique," *Journal of artificial intelligence research*, vol. 16, pp. 321–357, 2002.