MEDINFO23
8 – 12 JULY 2023 | SYDNEY, AUSTRALIA

Multimodality Fusion Method based on Multiview Subspace Clustering for Pulmonary Embolism Diagnosis

zhuyan_me@zhejianglab.com

Yan Zhu

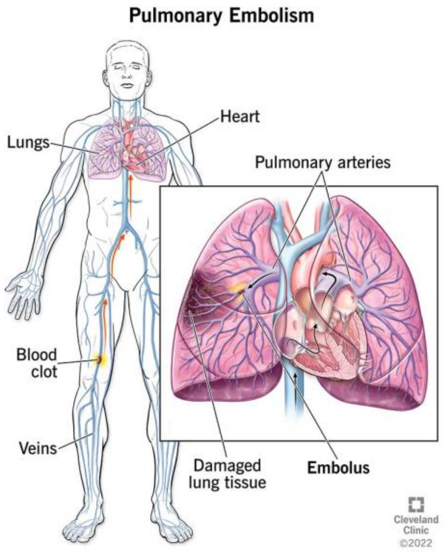
Postdoctoral,
Zhejiang Lab, China

之江实验室
ZHEJIANG LAB

AIDH IMIA

Introduction

- Pulmonary embolism (PE) is a blood clot in the blood vessels of the lung, which may lead to lung damage or be harmful to other organs
- Early diagnosis and timely treatment is important!
- CTPA is the gold standard for diagnosis, while physician fatigue, diagnosis error and poor image quality lead to miss diagnosis
- Deep learning methods have shown application potential in PE detection and diagnosis



Related work

Author	method	modality	performance
Huang *et al*	A 77-layer 3D CNN PENet	CTPA	0.85 (AUC)
Yang *et al.*	two-stage CNN that consists of detection net and classify net	CTPA	75.4% (sensitivity)
Huang *et al.*	late fusion -PENet for CT -ElasticNet for EMR	CTPA & EMR	0.947 (AUC)
Zhi *et al.*	joint fuson -MLP for EMR -2D CNN for CT	CTPA & EMR	97.3%（accuracy） 0.964（F1 score）

Disadvantages

- Few work focus on multimodality analysis
- Using concatenating or MLP to fuse original high-dimensional data

This work

- Multi-view subspace clustering-based feature selection method to fuse multimodality data
- A multimodality fusion architecture with early, joint and late fusion methods
- Best model achieves AUROC of 0.947

Methods

Materials

- **RadFusion**: 1837 studies from Stanford University Medical Center consists of CT imaging and EMR
- The EMR includes ICD9 codes, vitals, lab tests, demographics and inpatients and outpatient medications.
- **Label**: 0 for negative PE and 1 for positive PE
- **Set spilt**: 1454 studies for train, 193 for validation and 190 for test

Preprocessing

- Follow [4], the EMR data was screened to have 1505 features, and CT images are processed by PENet to get 2048 image features
- Feature selection: Mann-Whitney U test and Spearman correlation filter
- 632 image features and 65 EMR features remained

Methods

Multimodality fusion based on MVSC

- **Motivation**: take different modality data as different views of the object, use multiview subspace clustering guided feature selection (**MSCUFS**) method to fuse multimodality data
- **MSCUFS Model**:
  - View-specific self representation term: $\sum_{v=1}^{V} \|X_v - X_v Z_v\|_F^2$
  - Feature selection term: $\|X^T W - F\|_{2,1}$
  - Graph embedding term to preserve local geometry structure: $\sum_{v=1}^{V} tr(F^T L_v F)$
  - Sparse constraint on feature selection matrix: $\|W\|_{2,1}$

$$\arg \min_{W,F,Z_v,v=1,\dots,V} E(W,F,Z_v) = \sum_{v=1}^{V} \|X_v - X_v Z_v\|_F^2 + \lambda_1 \|X^T W - F\|_{2,1}$$
$$+\lambda_2 \sum_{v=1}^{V} tr(F^T L_v F) + \lambda_3 \|W\|_{2,1},$$
$$\text{s.t. } Z_v \mathbf{1} = \mathbf{1}, Z_v(i,i) = 0, v = 1, \dots, \mathrm{V}, \mathrm{F} \geq 0, F^T F = I_c.$$

- **Optimization**: select 200 anchor points by k-means, and use an iterative optimization approach
- **Results**: the importance of each feature can be ranked by $\|w_i\|_2$

Methods

Multimodality fusion architecture

- Multimodality fusion architecture is established with early, late and joint fusion strategies
- With grid search on validation, the number of features and model parameters are set
- 75 features consist of 52 image features and 23 EMR features are selected by MSCUFS
- ML models, i.e. SVM, Logistic regression, ElasticNet and neural network are used to construct the prediction models

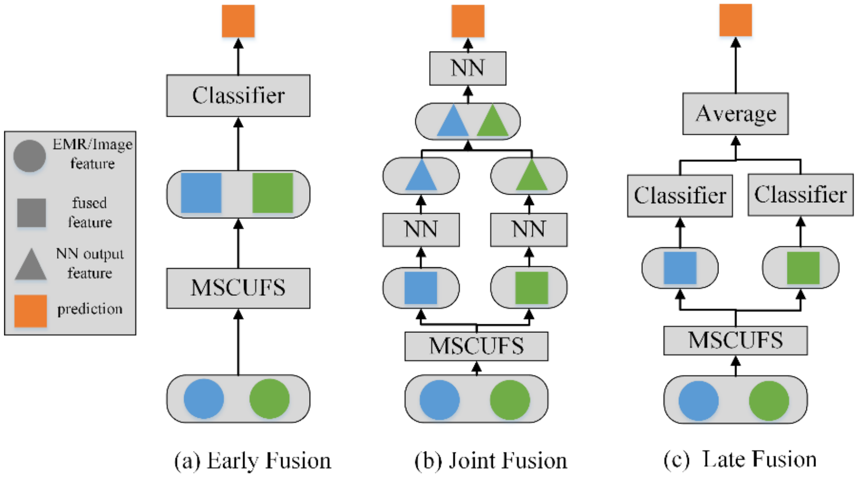


Figure 1. Fusion model architecture. (a) Early fusion, (b) Joint Fusion, (c) Late fusion. The input of each model is EMR and image features after data preprocessing

Results

Table 1. Fusion model results. Different fusion methods take the MSCUFS fused multimodality features as input and are constructed with different machine learning classifiers. Best performance metrics in bold text.

Evaluation Metrics	Early fusion			Late fusion average			Joint fusion
	Elastic	SVC	Logistic	Elastic	SVC	Logistic	NN
Accuracy	0.8842	0.8895	0.8947	0.8842	0.8842	0.8895	0.9000
AUROC	0.9331	0.9342	0.9313	0.9316	0.9183	0.9376	**0.9478**
Specificity	0.8818	0.8750	0.8625	0.9273	0.9455	0.9182	0.8500
Sensitivity	0.8875	0.9000	0.9182	0.8250	0.8000	0.8500	0.9364
PPV	0.9151	0.9083	0.9018	0.8793	0.8667	0.8938	0.8957
NPV	0.8452	0.8642	0.8846	0.8919	0.9143	0.8831	0.9067

PPV: positive predictive value, NPV: negative predictive value

Discussion

MSCUFS brings improvement in all three fusion methods.

Table 2. Comparison between best performing early fusion, late fusion and joint fusion models with or without MSCUFS-based feature selection.

Evaluation Metrics	Early Elastic fusion		Late Elastic fusion		Joint NN fusion	
	baseline	MSCUFS	baseline	MSCUFS	baseline	MSCUFS
Accuracy	0.8316	0.8842	0.8737	0.8842	0.8368	0.9000
AUROC	0.8801	0.9331	0.9277	0.9316	0.8800	0.9478
Specificity	0.8000	0.8818	0.9000	0.9273	0.7500	0.8500
Sensitivity	0.8750	0.8875	0.8375	0.8250	0.9000	0.9364
PPV	0.8980	0.9151	0.8839	0.8793	0.8319	0.8957
NPV	0.7609	0.8452	0.8590	0.8919	0.8451	0.9067

PPV: positive predictive value, NPV: negative predictive value

Discussion

- multimodality data has better performance than imaging or EMR alone in PE diagnosis
- EMR data achieves much higher accuracy than imaging model, and the combination of both brings better sensitivity and PPV

Table 3. Comparison between best performing multimodality and single modality models.

Evaluation Metrics	Imaging SVC model	EMR Elastic model	Joint NN model
Accuracy	0.7421	0.9053	0.9000
AUROC	0.7840	0.9239	0.9478
Specificity	0.7091	0.9636	0.8500
Sensitivity	0.7875	0.8250	0.9364
PPV	0.8211	0.8833	0.8957
NPV	0.6632	0.9429	0.9067

PPV: positive predictive value, NPV: negative predictive value

Conclusion

- we propose a novel multimodality fusion method that adopt multi-view subspace clustering guided feature selection (MSCUFS) to fuse imaging and EMR data
- Experiments show the effectiveness of MSCUFS in improving PE classifier performance and the superiority of multimodality model than imaging-only or EMR-only model.

Thank you for your attention