MEDINFO 2023 — The Future Is Accessible
J. Bichel-Findlay et al. (Eds.)

doi:10.3233/SHTI231098

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

Peijun HU[a], Qianqian QI[a], Yanxia ZHAO[a], Miaomiao FU[a], Jingsong LI[a,b,1]
[a] *Research Center for Healthcare Data Science, Zhejiang Lab, Hangzhou, China*
[b] *Engineering Research Center of EMR and Intelligent Expert System, Ministry of Education, College of Biomedical Engineering and Instrument Science, Zhejiang University, Hangzhou, China*

Abstract. Pulmonary embolism (PE) is an important clinical disorder that will result in lung tissue damage or low blood oxygen levels, which need early diagnosis and timely treatment. While computed tomographic pulmonary angiography (CTPA) is the gold standard to diagnose PE, previous studies have verified the effectiveness of combing CTPA and EMR data in computer-aided PE detection or diagnosis. In this paper, we proposed a multimodality fusion method based on multi-view subspace clustering guided feature selection (MSCUFS). The extracted high-dimensional image and EMR features are firstly selected and fused by the MSCUFS, and then are feed into different machine learning models with different fusion strategy to construct the PE classifier. The experiment results showed that the joint fusion strategy with MSCUFS achieved best AUROC of 0.947, surpassing other early fusion and late fusion models. The comparison between single modality and multimodality also illustrated the effectiveness of the proposed method.

Keywords. Pulmonary embolism, multimodality fusion, multi-view subspace clustering, feature selection

1. Introduction

Pulmonary embolism (PE) is an important clinical disorder that needs early diagnosis for timely and proper treatment. A pulmonary embolism will result in a lack of blood flow, which may lead to lung damage or cause low blood oxygen levels that will be harmful to other organs or body [1]. The gold standard for PE diagnosis is computed tomographic pulmonary angiography (CTPA), which can provide accurate diagnosis by radiologists. However, due to the limited healthcare resources of experts, physician fatigue, diagnosis error and poor image quality, miss diagnosis of PE remains frequently happened.

With the development of deep learning in recent decade, it has shown application potential in medical imaging, including the detection and diagnosis of PE. Huang *et al.* [2] developed and evaluated an end-to-end deep learning model for detecting PE with

[1] Corresponding Author: Jingsong Li, email: ljs@zju.edu.cn.
The project is supported in part by the National Natural Science Foundation of China (No. 12101571) and Key Research Project of Zhejiang Lab (No. 2022ND0AC01).

CTPA. The model achieved AUROC of 0.85 on external dataset. Vijayachitra *et al.* [1] designed a deep learning neural network classifier to detect left-side/right-side PE or normal from CT images. Yang *et al.* [3] proposed a two-stage CNN for PE detection on CTPA images. Despite the promising results of image classification models, researches have pointed out that using multimodality data such as image and EMR data can bring improvement. Huang *et al.* [4] developed different multimodality fusion architectures and applied them in PE detection, in which the late fusion model with imaging and EMR data outperformed image-only or EMR-only models. Based on this work, an MLP-2D CNN multimodality model [5] was proposed for PE diagnosis, in which a MDS algorithm was applied to EMR data to reduce feature dimension and 2D CNN was designed to replace original 3D CNN model.

Although above-mentioned multimodality fusion brings improvement over single-modality data, these work generally used concatenating or MLP to perform modality fusion on original high-dimensional data, which ignores relationships between these data and may face curse of dimensionality. In this paper, we propose to use multi-view subspace clustering-based feature selection for multimodality fusion. In addition, we compare different fusion architectures of early fusion, joint fusion and late fusion based on the feature selection results. The experimental results showed the joint fusion NN model achieves AUROC of 0.947 on the test set of RadFusion [6], surpassing early and late fusion models.

2. Methods

In this section, we first introduced the materials used in this study. Then, the proposed multimodality fusion method and classification model are illustrated. Figure 1 illustrates the multimodality fusion method with early, join and late fusion architectures. The input image and EMR data are extracted features from CT and EMR, which are then fused by the MSCUFS model. The output used features are feed into different machine learning classifiers or joint NN models to construct the PE diagnosis classifier.

2.1. Materials

The evaluation dataset comes from RadFusion [6], in which 1837 studies from Stanford University Medical Center (SUMC) were enrolled. CT imaging and patient EMR were pulled from the PASCS and SUMC Epic database. The EMR includes ICD9 codes, vitals, lab tests, demographics and inpatients and outpatient medications. The label is given as 0 or 1, which refers to negative PE and positive PE, respectively. The studies are split into a training set (1454 studies), a validation set (193 studies) and a test set (190 studies).

2.2. Data preprocessing

Follow [4], original EMR and CT scans are processed to obtaine structural EMR and imaging features. After screening, the EMR data contains 1505 features. The CT images are processed by the image-only PENet [2] and obtained 2048 image features. Before fusion, we applied a feature selection pipeline on each EMR and image features to reduce dimension. First, Mann-Whitney U test is applied to each feature, and features are abandoned if $p > 0.05$. Then, Spearman correlation between feature is

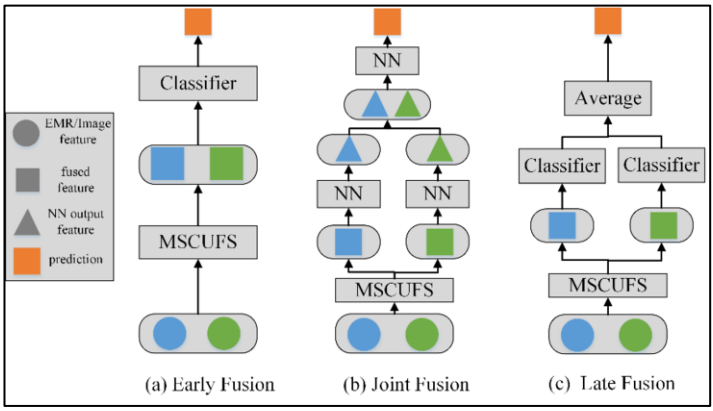

Figure 1. Fusion model architecture. (a) Early fusion, (b) Joint Fusion, (c) Late fusion. The input of each model is EMR and image features after data preprocessing.

calculated. If the correlation coefficient is greater than 0.9, then the latter feature is deleted. After feature selection on training dataset, 632 image features and 65 EMR features are obtained for subsequent multimodality fusion.

2.3. Multimodality fusion based on multi-view subspace clustering

The EMR and imaging data reflect different aspects and can be taken as multi-view data of subjects. In this section, we introduce the multi-view subspace clustering guided feature selection (MSCUFS) model, which is extended from single-view feature selection model SCUFS [7].

Denote features in the v-th view as $X^v = [x_1^v, x_2^v, \dots, x_n^v] \in \mathcal{R}^{d_v \times n}$, $x_1^v \in \mathcal{R}^{d_v \times 1}$, d_v is the dimension of features in the v-th view, n is the number of samples. The feature matrixes in V views $\{X^v\}_{v=1}^V$ can be concatenated to represent the overall feature matrix as $X = [(X^1)^T, (X^2)^T, \dots, (X^V)^T]^T \in \mathcal{R}^{d \times n}, d = \sum_{v=1}^V d_v$. In the case of fusing EMR and CT features, V=2. The MSCUFS model is constructed as:

$$\arg \min_{W,F,Z_v,v=1,\dots,V} E(W,F,Z_v) = \sum_{v=1}^V \|X_v - X_v Z_v\|_F^2 + \lambda_1 \|X^T W - F\|_{2,1}$$

$$+\lambda_2 \sum_{v=1}^V tr(F^T L_v F) + \lambda_3 \|W\|_{2,1},$$

$$\text{s.t.} Z_v \mathbf{1} = \mathbf{1}, Z_v(i,i) = 0, v = 1, \dots, \text{V}, \text{F} \geq 0, F^T F = I_c.$$

The first term is view-specific self-representation term that ensures data structure, the second term is feature selection term, the third term is graph embedding term that maintains local geometry structure, and the last term poses sparse constraint on feature selection matrix. In the objective function, $F \in \mathcal{R}^{n \times c}$ is relaxed cluster indictor matrix, c is the number of clusters, $W \in \mathcal{R}^{d \times c}$ is the feature selection matrix, $Z_v \in \mathcal{R}^{n \times n}$ is the view-specific self-representation matrix, and L_v is the Laplacian matrix of v-th view. Notations λ_1, λ_2 and λ_3 are the tradeoff parameters.

The objective loss function is optimized by an iterative approach. Since the time complexity of optimization algorithm is $\mathcal{O}(d^3 + n^3\text{d})$, we select 200 anchor data points by k-means to performed MSCFUS-based feature fusion. With optimized $\widehat{\text{W}}$, the importance of each feature can be ranked by $\|w_i\|_2$,where w_i is the i-th row of $\widehat{\text{W}}$.

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	Early fusion			Late fusion average			Joint fusion
Metrics	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

2.4. Classification model construction

To find the best fusion architecture with MSCUFS, early, late and joint fusion models are constructed based on the fused EMR and image features by MSCUFS. The early fusion takes fused multimodality features as input, and construct PE classify model with machine learning models, i.e. SVM, Logistic and ElasticNet. By grid search on validation set, optimal parameters of λ_1, λ_2 and λ_3 in MSCUFS and the number of features are set w.r.t. AUROC. Finally, 75 features consist of 52 image features and 23 EMR features are selected. The SVM model achieved best AUROC of 0.9342. As illustrated in Fig.1(b), a two–path Neural Network (NN) model with selected EMR features and image features as input is constructed. The single modality features are first feed into each NN, then the outputs are concatenated to form input for joint NN model to get predicted probability of PE. In the late fusion model as shown in Fig.1(c), each selected feature set is used to construct individual classify model. Then the predicted results are fused by averaging.

3. Results

A set of metrics, i.e., accuracy, AUROC, specificity, sensitivity, PPV (positive predictive value) and NPV (negative predictive value) are used as evaluation metrics. Table 1 shows different fusion methods with MSCUFS fused multimodality features. The listed models include early fusion and late fusion methods with Elastic, SVM, and Logistic classifiers, and joint fusion with NN model. Compared results shows the joint NN fusion model achieves best AUROC of 0.9478.

4. Discussion

To show the effectiveness of MSUCFS-based feature fusion, early fusion, late fusion and joint fusion models with and without MSCUFS are compared in Table 2. As can be seen, MSCUFS brings improvement in all three fusion methods. Table 3 compares single modality model with multimodality model. As can be seen, the imaging model and EMR model achieved AUROCs of 0.7840 and 0.9239, inferior than multimodality model that achieved AUROC of 0.9478. This result verifies the assumption that multimodality data have better performance than imaging or EMR alone in PE diagnosis. Indeed, we see that EMR data achieves much higher accuracy than imaging model, and the combination of both brings better sensitivity and PPV, which is meaningful to clinic application.

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

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

5. Conclusions

In this paper, we propose a novel multimodality fusion method that adopt multi-view subspace clustering guided feature selection (MSCUFS) to fuse imaging and EMR data. Three fusion architectures, i.e. early, late and joint fusion models are constructed with the MSCUFS fused features. Experiments show the effectiveness of MSCUFS in improving PE classifier performance and the superiority of multimodality model than imaging-only or EMR-only model.

References

[1] Vijayachitra S, Prabhu K, Abarana M, Deepa A, Loga Priya L. Deep Learning Technique-Based Pulmonary Embolism (PE) Diagnosis. In Advances in Electrical and Computer Technologies: Select Proceedings of ICAECT 2021 2022 Jun 26 (pp. 695-702). Singapore: Springer Nature Singapore, doi: 10.1007/978-981-19-1111-8_52.

[2] Huang SC, Kothari T, Banerjee I, Chute C, Ball RL, Borus N, Huang A, Patel BN, Rajpurkar P, Irvin J, Dunnmon J. PENet—a scalable deep-learning model for automated diagnosis of pulmonary embolism using volumetric CT imaging. NPJ Digit Med. 2020 Apr;3(1):61, doi: 10.1038/s41746-020-0266-y.

[3] Yang X, Lin Y, Su J, Wang X, Li X, Lin J, Cheng KT. A two-stage convolutional neural network for pulmonary embolism detection from CTPA images. IEEE Access. 2019 Jun;7:84849-57, doi: 10.1109/ACCESS.2019.2925210.

[4] Huang SC, Pareek A, Zamanian R, Banerjee I, Lungren MP. Multimodal fusion with deep neural networks for leveraging CT imaging and electronic health record: a case-study in pulmonary embolism detection. Sci Rep. 2020 Dec;10(1):22147, doi: 10.1038/s41598-020-78888-w.

[5] Zhi Z, Elbadawi M, Daneshmend A, Orlu M, Basit A, Demosthenous A, Rodrigues M. Multimodal Diagnosis for Pulmonary Embolism from EHR Data and CT Images. Annu Int Conf IEEE Eng Med Biol Soc. 2022 Jul;2022:2053-2057, doi: 10.1109/EMBC48229.2022.9871041.

[6] Zhou Y, Huang SC, Fries JA, Youssef A, Amrhein TJ, Chang M, Banerjee I, Rubin D, Xing L, Shah N, Lungren MP. Radfusion: Benchmarking performance and fairness for multimodal pulmonary embolism detection from ct and ehr. arXiv preprint arXiv:2111.11665. 2021 Nov 23, doi: 10.48550/arXiv.2111.11665.

[7] Zhu P, Zhu W, Hu Q, Zhang C, Zuo W. Subspace clustering guided unsupervised feature selection. Pattern Recognit. 2017 Jun;66:364-74, doi: 10.1016/j.patcog.2017.01.016.