



# Joint Clinical Data and CT Image Based Prognosis: A Case Study on Postoperative Pulmonary Venous Obstruction Prediction

Xinrong Hu<sup>1</sup>(✉), Zeyang Yao<sup>2</sup>, Furong Liu<sup>2</sup>, Wen Xie<sup>2</sup>, Hailong Qiu<sup>2</sup>, Haoyu Dong<sup>2</sup>, Qianjun Jia<sup>2</sup>, Meiping Huang<sup>2</sup>, Jian Zhuang<sup>2</sup>, Xiaowei Xu<sup>1</sup>, and Yiyu Shi<sup>1</sup>

<sup>1</sup> University of Notre Dame, Notre Dame, IN 46556, USA

{xhu7,xxu8,yshi4}@nd.edu

<sup>2</sup> Guangdong Provincial People's Hospital, Guangdong 510000, China

**Abstract.** Very often doctors diagnose diseases and prescribe treatments through cross-referencing patients' clinical data as well as radiology reports. On the other hand, while a few existing machine learning frameworks for diagnosis, treatment planning, and prognosis have used both clinical data and medical images, they all have prior knowledge about what information should be extracted from medical images. However, this is not the case for many diseases. For example, cardiac anatomical structure and tissue shapes are essential for pulmonary venous obstruction (PVO) prediction after correction of total anomalous pulmonary venous connection (TAPVC), but the exact graphical features in the computed tomography (CT) images that should be measured remain unclear. In this paper, we propose to use convolutional neural network to automatically obtain features from CT images and combine them with clinical data in an end-to-end trainable manner. We further collect a dataset consisting of 132 TAPVC patients for evaluation, and find that jointly using clinical data and CT images to predict postoperative PVO outperforms the method based on either clinical data or CT images alone. Our dataset is released to the community to promote further research.

**Keywords:** Cardiac CT images · Pulmonary venous obstruction · Total anomalous pulmonary venous connection · Convolutional neural network

## 1 Introduction

Over the past decades, machine learning has demonstrated great power in the medical domain and been applied for different tasks, including diagnosis, treatment planning, and disease prognosis. In an analogy to the common practice of human doctors where both the clinical data (e.g. demographic information

and lab results) and radiology reports (e.g., X-ray, CT, and MRI) are cross-referenced, there is a growing trend involving both clinical data and medical images in machine learning based frameworks to handle various diseases (e.g., [2, 8, 9, 12]). In [8], all metrics in coronary computed tomographic angiography (CCTA) are measured visually by experienced cardiologists or radiologists, which combined with hospital records are used to predict 5-year all-cause mortality. [9] benefits from considering both the auxiliary information with the imaging data and non-imaging information in a graphical neural network for brain analysis in populations. Recently, [2] combines clinical information and myocardial perfusion imaging (MPI) data to predict major adverse cardiac events. They take advantage of professional software to preprocess MPI and to automatically quantify variables in images.

However, all these works require predefined features of medical images, either measured by software or manually, to be combined with clinical data. Prior knowledge about what information should be extracted from medical images is needed. Unfortunately, such knowledge is not always available for many diseases. In this paper, we use postoperative pulmonary venous obstruction (PVO) prediction problem as a vehicle to explore the practicability of combining clinical data and CT images under such a scenario. The relationship between graphical features on CT images and PVO recurrence remains unknown as of today.

Total anomalous pulmonary venous connection (TAPVC) contributes to about 3% of all congenital heart diseases [5] and yet has notoriously high mortality achieving nearly 80% if without intervention [3]. It is characterized by failure of the pulmonary venous confluence (PVC) to be absorbed into the dorsal portion of the left atrium (LA) in combination with a persistent splanchnic connection to the systemic venous systems. Even with surgical repair, the death rate is still reported as 5% to 7%. PVO is one of the most frequent causes of death after operations. Accurate prediction model can identify patients with high PVO recurrence risk, and then their chance of survival will be improved by early preventive treatment.

Previous works only focus on finding features in the clinical data that have a close association with recurrent PVO, like [6]. An automatic model for PVO prediction is still missing. In addition, it remains unknown if CT images would be helpful in such prediction, and if so, the graphical features that may be of relevance. To answer these questions, in this paper we build a prediction model that explicitly indicates whether a patient will suffer from PVO after surgery. Specifically, we put forward three machine learning based methods: i) applying logistic regression with features selected from the clinical data; ii) using a convolutional neural network (CNN) to extract features from patients' 3D cardiac CT images for prediction; and iii) building an architecture where clinical data features are combined with graphical features through an end-to-end trainable CNN.

The limited number of TAPVC patients and biased distribution of postoperative PVO and non-PVO cases introduce additional challenges. To address these issues, we adopt a series of practical techniques regarding data preprocessing and

learning strategies: i) image augmentation is utilized to increase the size of CT dataset; ii) for training epochs, positive cases are over-sampled in an attempt to balance the biased distribution; iii) we modify the loss function, adding L2 normalization of weights in the last fully connected layer of the CNN for generalization performance, and adjusting the weights of the PVO and non-PVO samples.

The main contributions of this paper are concluded as follows:

- For the first time in the literature, we present an automatic prognosis model for PVO after TAPVC surgery and demonstrate its promising accuracy on test dataset.
- We propose clinical data based, CT image based, and joint data and image based methods for the prognosis model. By comparison, we find that jointly considering both clinical data and medical images in an end-to-end trainable manner can indeed provide better prediction.
- Several empirical techniques of data preprocessing and training are discussed in order to deal with small and biased dataset pertinent to TAPVC.
- The method can possibly be extended to various diagnosis, treatment planning and prognosis problems where both clinical data and medical images are important yet the related graphical feature is unknown.
- Our dataset composed of both clinical data and CT images of 132 patients is released [1] to promote further research on this topic.

## 2 Methods

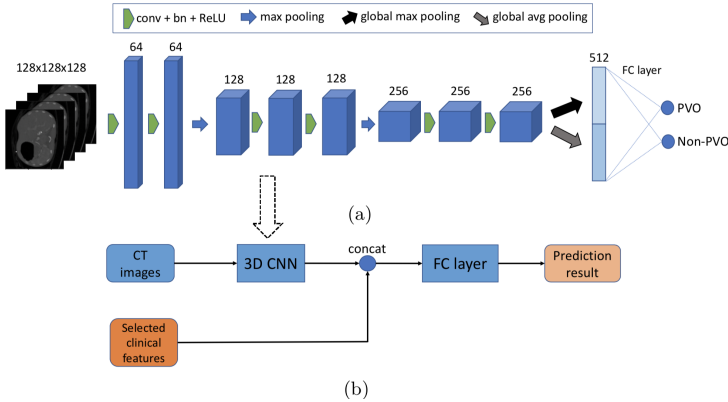
### 2.1 Clinical Data Based Method

Prognosis about the onset of PVO after TPAVC correction can be treated as a binary classification problem. With only clinical data taken into consideration, each patient  $A_k$  is defined by several features  $\{x_{k1}, x_{k2}, x_{k3}, \dots\}$ , and a target label  $y_k$  indicating whether PVO occurs after surgery (either 1 or 0). The problem is then to generate predictions  $\tilde{y}_1, \tilde{y}_2, \tilde{y}_3, \dots$  and to optimize a defined loss function  $\sum_k L(y_k, \tilde{y}_k)$ . The most prevalent methods for such a binary classification problem include logistic regression, support vector machine, and random forest, etc. Among these machine learning models, we find that logistic regression is the most effective method for PVO prognosis in practice.

To decide which features to feed into the prediction model, we initially select a set of candidate features which cardiologists believe to have direct or indirect effects on the recurrence of PVO. For example, sutureless operation is believed to have a potential impact on the geometric distortion of pulmonary venous suture line and thus on the postoperative PVO, according to the previous study [16]. Then, we recursively prune features from the candidate set in the ascending order of the features' importance score until the optimal prediction accuracy is achieved. Such an iterative pruning can help to remove redundant information and reduce potential overfitting.

## 2.2 CT Image Based Method

Cardiac CT images have never been used in traditional prognosis of TAPVC repair. Yet CT images provide spatial information like anatomical structure and tissue development, which clinical data fails to provide. However, those graphical features are not specified by any measurable parameters. Hence, we apply a convolutional neural network to extract information from patients' CT and then to make a prediction on postoperative PVO. The input of the network is 3D images, and thus it resembles a 3D classification network. Because of the memory bottleneck induced by 3D medical images, very deep CNNs like VGG [13] and ResNet [4] is not suitable. Figure 1(a) illustrates our 3D CNN's architecture. The original 3D cardiac CT images are first cropped to the region of interest and resized to a uniform shape, after which a batch of images are passed to the model. The model consists of 3 resolution stages, and for each resolution stage, there are two convolutional layers followed by max pooling with stride two to half the resolution as well as double the number of channels. The size of feature maps maintains the same for every convolutional operation. We replace the max pooling with global pooling to flatten the feature maps in the last stage, which will be followed by a fully connected layer for predicting postoperative PVO. For the global pooling layer, inspired by [15], we use a combination of global max pooling and global average pooling. These two types of pooling act like two filters with different frequency responses, so that our model takes advantage of both high-frequency and low-frequency information. All the convolutional and pooling operations mentioned above are 3D counterparts. Besides, the number of resolution stages and initial filters are both adjustable. After comparison of different combinations, we decide on this architecture that strikes a balance between computational efficiency and generalization performance.



**Fig. 1.** (a) Architecture of CT image based 3D CNN. The blue cuboid represents 3D feature maps, and the number of channels is marked above it. (b) Framework of joint clinical data and image based postoperative PVO prediction method. (Color figure online)

### 2.3 Joint Data and Image Based Method

Actually, CT images and clinical data give the prediction based on factors from two distinct domains. Thus, it is a natural thought to combine the knowledge from those two “experts” skilled in different fields, in other words, using CT images and clinical data jointly for the prediction task. We concatenate the selected features from clinical data (based on the method described in Sect. 2.1) with the flattened activations right before the fully connected layer in our 3D CNN, as shown in Fig. 1(b). In this way, we build an end-to-end deep learning model making use of both CT images and clinical data, which we find demonstrates better performance than the methods using either one of the data sources solely.

## 3 Dataset

Our study enrolls 306 patients who were diagnosed with TAPVC and received surgical treatment from January 2009 to September 2019 with complete clinical data and CT images. There are four types of TAPVC, including supracardiac, infracardiac, cardiac, and mixed. In terms of image-based methods, these four types have heterogeneous anatomical structure, so it is hard for a single neural network to learn all distinct spatial information. Thus, we only consider the most common type, the supracardiac, that accounts for 30%–50% of total TAPVC. Among the 306 patients, there are 132 (43.1%) cases of supracardiac TAPVC which are used to create the dataset.

For clinical data, suggested by expert cardiologists, we focus on 14 features, which are patient weight at the time of operation (operation weight), length of hospital stay (hospital stay), Alanine Transaminase value (ALT), Aspartate Aminotransferase value (AST), Total Bilirubin value (TBIL), Direct Bilirubin value (DBIL), INR (International Normalized Ratio), Prothrombin activity time (PT), aortic cross-clamp time (cross-clamp), CPB time (Cardiopulmonary Bypass time), Deep hypothermic circulatory arrest time (DHCA), and binary features including gender, the use of sutureless operation (sutureless), and ligation. All lab results are collected before the operation. These features along with their ranges are summarized in Table 1. As for medical images, all 3D CT were captured by a Siemens Biography 64 CT scanner, and the typical voxel size is  $0.25\text{ mm} \times 0.25\text{ mm} \times 0.5\text{ mm}$ .

Based on the patients’ follow-up record after surgery (which is not part of this dataset and is not used in model training), when at least one of the following three conditions is met, there is postoperative PVO recurrence: i) blood flow in the vertical vein or common trunk of pulmonary vein greater than  $1.8\text{ m/s}$ ; ii) atrial septal defect smaller than  $3\text{ mm}$ ; iii) color ultrasonic diagnosis suggesting obstruction. Finally, 10% of the patients in the dataset are labeled with postoperative PVO. This ratio agrees with epidemiology studies, suggesting that our dataset is relatively unbiased.

**Table 1.** The 14 candidate features in the clinical data that may be relevant to post-operative PVO recurrence, their respective ranges, and the importance score (IS) from logistic regression (clinical data based method).

Feature	Range	IS	Feature	Range	IS
Operation weight	2.63–53.0 kg	<b>21.8</b>	Hospital stay	0–77 d	<b>0.819</b>
ALT	5–948 IU/L	0.771	AST	20–2420 IU/L	0.237
TBIL	6.8–211.6 $\mu$ mol/L	0.324	DBIL	1.9–104.9 $\mu$ mol/L	0.498
INR	0.83–2.55	<b>44.8</b>	PT	31–140 s	0.040
Cross-clamp	0–153 min	0.409	CPB	40–290 min	<b>1.38</b>
DHCA	0–40 min	<b>4.35</b>	Gender	{0, 1}	<b>19.0</b>
Sutureless	{0, 1}	<b>55.4</b>	Ligation	{0, 1}	<b>59.1</b>

## 4 Preprocessing and Learning Techniques

### 4.1 Image Augmentation

To make full use of all cardiac CT images in the limited dataset, we deploy classic augmentation methods to enlarge the dataset. Traditional image augmentation methods include translation, rotation, scaling, and flipping [10, 11]. In order to preserve the spatial structure and relative position of atrial and vessels, we only adopt rotation to increase both the training set and test set. Firstly, we crop the CT images so that the region of interest is centered. Then, we rotate the images clockwise by ( $10^\circ$ ,  $-10^\circ$ ,  $90^\circ$ ,  $180^\circ$ ,  $270^\circ$ ). As a result, we can expand the dataset by six times. Note that we first divide all the data into four folds with PVO patients evenly distributed. Only then do we augment the images in each fold and do four-fold cross-validation. This trick only works for the two methods described in Sects. 2.2 and 2.3, as it cannot expand clinical data.

### 4.2 Resampling

The distribution of our dataset is highly biased, which genuinely reflects the recurrence rate of PVO. To deal with the imbalance problem, resampling the dataset is a simple and effective method. There are two main methods called over-sampling and under-sampling. Over-sampling is to add copies of samples from the under-presented class, and under-sampling is to remove instances belonging to the over-presented class. In practice, we find over-sampling fits our dataset

better than under-sampling because the dataset is already small, and reducing training samples would compromise models’ generalization performance.

### 4.3 Loss Function Modification

The loss function  $L$  we choose for the 3D CNN is cross entropy. However, during training, we find the training loss is smaller than validation loss, indicating potential overfitting of our model on the training dataset. L2 normalization is a common way to alleviate overfitting by restraining weights from growing too large. The modified loss function for every batch is defined as:

$$L_{batch} = - \sum_{k \in B} weight[class] * \ln s_k[class] + \beta |w|^2, \quad (1)$$

in which  $B$  is the batch set,  $s_k[class]$  is the softmax value of a class (PVO or non-PVO) at the output layer,  $w$  is the weight vector of the last fully connected layer, and  $\beta$  is an adjustable parameter which we set to be 2. Additionally, we assign different  $weight[class]$  to the two classes, since positive samples appear at a lower frequency in a batch. A larger weight for the PVO class forces the model to be more sensitive to wrongly predicting PVO patients as non-PVO, i.e., false negative, which is more critical than the other way around. On the other hand, if it is too large, it would lead to high false positive rate. Our experiments suggest that the optimal of weights for PVO and non-PVO classes is (30, 1).

**Table 2.** Impact of the techniques described in Sect. 4 on the three prediction methods. “8” or “14” stands for the number of features used (pruned features v.s. all candidate features). For the CT image based method and the joint method, over-sampling is always applied.

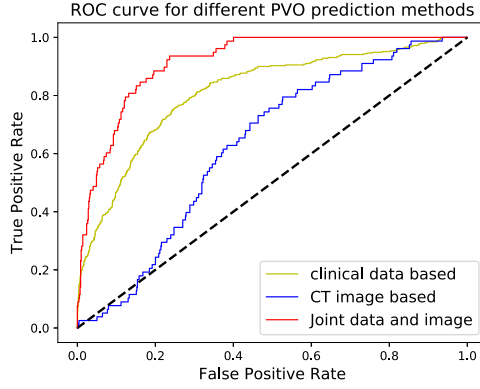
	Clinical data based (8, w/o over-sample)	Clinical data based (8, w/over-sample)	CT image based (-, w/L2 norm)
AUC	$0.824 \pm 0.060$	$0.856 \pm 0.061$	$0.618 \pm 0.061$
	Joint method (14, w/L2 norm)	Joint method (8, w/o L2 norm)	Joint method (8, w/L2 norm)
AUC	$0.818 \pm 0.081$	$0.933 \pm 0.041$	<b><math>0.941 \pm 0.027</math></b>

## 5 Experiments

The metric we use to evaluate and compare different methods is AUC, which represents the area under the ROC (Receiver Operating Characteristics) curve,

and is an important measurement for a classification model’s performance as used in [7, 14]. In our setting, it shows the ability of a method to predict the postoperative PVO recurrence of a patient. For the four-fold cross-validation, the AUC scores of four test sets are recorded and the resulting average is reported.

For clinical data based method, logistic regression is applied to make the prediction as well as to calculate the importance score of each feature. The importance score is obtained from the absolute value of features’ coefficients in the logistic regression model, as listed in Table 1. We gradually remove some features from the 14 candidates following the ascending order of importance score. We observe that, when the number of predictors is eight, logistic regression model achieves the best performance. The eight features that are finally selected as postoperative PVO predictors by the clinical data based method are operation weight, hospital stay, INR, CPB, DHCA, gender, sutureless, and ligation.



**Fig. 2.** ROC curves of the three different prediction methods. All the techniques in Sect. 4 are applied as appropriate on the three methods. The joint data and image based method is consistently better than the other two.

Table 2 displays the impact of the techniques discussed in Sect. 4 on AUC when applied on the three prediction methods. For the sake of clarity, we choose not to enumerate all the possible combinations in this ablation study; rather we illustrate six cases which are sufficient to observe the effectiveness of the techniques. From the table, several interesting points can be observed. First, comparing the clinical data based method with and without data over-sampling, the technique can increase the AUC by 0.032. Second, comparing the joint method with and without L2 norm, when both methods use eight features, the technique can increase the AUC by 0.008. Third, comparing the joint method with eight features and 14 features, when both methods use L2 norm, the technique can increase the AUC by 0.123. This suggests that feature pruning is not only effective for clinical data based method, but also the joint one. Finally, comparing all the methods, the joint method achieves the highest AUC score of 0.941. The corresponding ROC curves of the three methods are shown in Fig. 2, where the

joint method is consistently better than the other two. This result supports the speculation that the 3D CNN is capable of extracting useful information from CT images for PVO prediction, complementing that from the clinical data.

## 6 Conclusion

In this paper, we explore how to combine clinical data and CT images for prognosis when the relationship between graphical features on CT images and the outcome is unknown. We use postoperative PVO prediction as a case study. A novel neural network architecture that jointly learns from clinical data and CT images in an end-to-end trainable manner is built. We also introduce a group of implementation tips involving data preprocessing and learning to manipulate the limited and biased dataset pertinent to the disease. Experimental results clearly demonstrate the advantage of the proposed method of joint learning.

## References

1. link omitted for blind review
2. Betancur, J., et al.: Prognostic value of combined clinical and myocardial perfusion imaging data using machine learning. *JACC: Cardiovascular Imaging* **11**(7), 1000–1009 (2018)
3. Burroughs, J.T., Edwards, J.E.: Total anomalous pulmonary venous connection. *Am. Heart J.* **59**(6), 913–931 (1960)
4. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778 (2016)
5. Herlong, J.R., Jaggars, J.J., Ungerleider, R.M.: Congenital heart surgery nomenclature and database project: pulmonary venous anomalies. *Annal. Thorac. Surg.* **69**(3), 56–69 (2000)
6. Husain, S.A., et al.: Total anomalous pulmonary venous connection: factors associated with mortality and recurrent pulmonary venous obstruction. *Annal. Thorac. Surg.* **94**(3), 825–832 (2012)
7. Kourou, K., Exarchos, T.P., Exarchos, K.P., Karamouzis, M.V., Fotiadis, D.I.: Machine learning applications in cancer prognosis and prediction. *Comput. Struct. Biotechnol. J.* **13**, 8–17 (2015)
8. Motwani, M., et al.: Machine learning for prediction of all-cause mortality in patients with suspected coronary artery disease: a 5-year multicentre prospective registry analysis. *Eur. Heart J.* **38**(7), 500–507 (2017)
9. Parisot, S., et al.: Spectral graph convolutions for population-based disease prediction. In: Descoteaux, M., Maier-Hein, L., Franz, A., Jannin, P., Collins, D.L., Duchesne, S. (eds.) *MICCAI 2017. LNCS*, vol. 10435, pp. 177–185. Springer, Cham (2017). [https://doi.org/10.1007/978-3-319-66179-7\\_21](https://doi.org/10.1007/978-3-319-66179-7_21)
10. Roth, H.R., et al.: Improving computer-aided detection using convolutional neural networks and random view aggregation. *IEEE Trans. Med. Imaging* **35**(5), 1170–1181 (2015)
11. Setio, A.A.A., et al.: Pulmonary nodule detection in ct images: false positive reduction using multi-view convolutional networks. *IEEE Trans. Med. Imaging* **35**(5), 1160–1169 (2016)

12. Shah, S.J., et al.: Phenomapping for novel classification of heart failure with preserved ejection fraction. *Circulation* **131**(3), 269–279 (2015)
13. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint [arXiv:1409.1556](https://arxiv.org/abs/1409.1556) (2014)
14. Weng, S.F., Reps, J., Kai, J., Garibaldi, J.M., Qureshi, N.: Can machine-learning improve cardiovascular risk prediction using routine clinical data? *PloS one* **12**(4) (2017)
15. Yuan, B., Xing, W.: Diagnosing cardiac abnormalities from 12-lead electrocardiograms using enhanced deep convolutional neural networks. In: Liao, H., Balocco, S., Wang, G., Zhang, F., Liu, Y., Ding, Z., Duong, L., Phellan, R., Zahnd, G., Breininger, K., Albarqouni, S., Moriconi, S., Lee, S.-L., Demirci, S. (eds.) *MLMECH/CVII-STENT -2019. LNCS*, vol. 11794, pp. 36–44. Springer, Cham (2019). [https://doi.org/10.1007/978-3-030-33327-0\\_5](https://doi.org/10.1007/978-3-030-33327-0_5)
16. Yun, T.J., et al.: Conventional and sutureless techniques for management of the pulmonary veins: evolution of indications from postrepair pulmonary vein stenosis to primary pulmonary vein anomalies. *J. Thorac. Cardiovasc. Surg.* **129**(1), 167–174 (2005)