International Journal of Computer Assisted Radiology and Surgery (2021) 16:609–617
https://doi.org/10.1007/s11548-021-02335-y

ORIGINAL ARTICLE

A machine learning-based pulmonary venous obstruction prediction model using clinical data and CT image

Zeyang Yao[1] · Xinrong Hu[2] · Xiaobing Liu[3] · Wen Xie[1] · Yuhao Dong[4] · Hailong Qiu[3] · Zewen Chen[3] · Yiyu Shi[2] · Xiaowei Xu[2] · Meiping Huang[4] · Jian Zhuang[3]

Received: 25 August 2020 / Accepted: 25 February 2021 / Published online: 31 March 2021

Abstract

Purpose In this study, we try to consider the most common type of total anomalous pulmonary venous connection and established a machine learning-based prediction model for postoperative pulmonary venous obstruction by using clinical data and CT images jointly.

Method Patients diagnosed with supracardiac TPAVC from January 1, 2009, to December 31, 2018, in Guangdong Province People's Hospital were enrolled. Logistic regression were applied for clinical data features selection, while a convolutional neural network was used to extract CT images features. The prediction model was established by integrating the above two kinds of features for PVO prediction. And the proposed methods were evaluated using fourfold cross-validation.

Result Finally, 131 patients were enrolled in our study. Results show that compared with traditional approaches, the machine learning-based joint method using clinical data and CT image achieved the highest average AUC score of 0.943. In addition, the joint method also achieved a higher sensitivity of 0.828 and a higher positive prediction value of 0.864.

Conclusion Using clinical data and CT images jointly can improve the performance significantly compared with other methods that using only clinical data or CT images. The proposed machine learning-based joint method demonstrates the practicability of fully using multi-modality clinical data.

Keywords Total anomalous pulmonary venous connection · Pulmonary venous obstruction · Prediction · Deep learning

Zeyang Yao, Xinrong Hu and Xiaobing Liu have contribute equally to this work

✉ Xiaowei Xu
xiao.wei.xu@foxmail.com

✉ Meiping Huang
huangmeiping@126.com

✉ Jian Zhuang
zhuangjian5413@163.com

Zeyang Yao
zeyang_yao@foxmail.com

Xiaobing Liu
liuxb21@aliyun.com

Wen Xie
xiewen0231@qq.com

Yuhao Dong
barbardong1@outlook.com

Hailong Qiu
hailong0518@qq.com

Zewen Chen
zerwing@msn.cn

1 School of Medicine, South China University of Technology Guangdong Cardiovascular Institute, Guangdong Provincial Key Laboratory of South China Structural Heart Disease, Guangdong Provincial People's Hospital, Guangdong Academy of Medical Sciences, Dongchuan Rd 96, Guangzhou 510080, China

2 Guangdong Cardiovascular Institute, Guangdong Provincial Key Laboratory of South China Structural Heart Disease, Guangdong Provincial People's Hospital, Guangdong Academy of Medical Sciences, Dongchuan Rd 96, Guangzhou 510080, China

3 Department of Cardiac Surgery, Guangdong Cardiovascular Institute, Guangdong Provincial Key Laboratory of South China Structural Heart Disease, Guangdong Provincial People's Hospital, Guangdong Academy of Medical Sciences, Dongchuan Rd 96, Guangzhou 510080, China

4 Department of Catheterization Lab, Guangdong Cardiovascular Institute, Guangdong Provincial Key Laboratory of South China Structural Heart Disease,

Introduction

Along with a variety of clinical features associated with the development of cardiovascular disease, various prediction models have been developed to identify high-risk individuals especially for cardiac surgery related disease [1–3]. Traditional methods for clinical end-points events prediction models usually use multivariate regression-based analyses. These models identify clinical features with odds or hazard ratios and then provide risks for target disease [4,5], which enables treatment strategies tailored to an individual.

However, these traditional pre-specified clinical features used prediction must meet the assumption of independence between features [2,6]. In an ideal medical research, because pre-selected clinical features are measured at pre-identified time points, the collected clinical feature information can be fully exploited by statistical methods. While, the logistically challenging of the establishment of traditional prediction methods making medical researches is not ideal. Given the growing volume of data and manual data processing, traditional methods is time-consuming and expensive, results in traditional prediction models may not fulfill their expectations [7]. In the current study [8], traditional clinical features derived from prior cohort studies such as BMI and TC are not significant predictors for cardiovascular disease in regression models, which corresponds to previous findings derived from analysis of various clinical data that shows many traditional clinical features were less significant factors for cardiovascular disease occurrence.

Deep learning (DL) is a class of machine learning algorithms and demonstrates excellent performance in classification nowadays [9]. The overall transformations have multiple layers in deep learning and this capacity could enhance predictive model performance in complex time-varying datasets. Under large training biomedical data and advanced computing power, DL has been applied to the development of risk prediction models using real word data [10–13]. Weng et al. [11] used routine clinical data from 378,256 patients from UK family practices, exploited establishing machine learning-based high-accuracy cardiovascular disease prediction model. Parisot et al. [12] benefited from considering both the auxiliary information with the imaging data and non-imaging information in a graphical neural network for brain analysis in populations. Particularly, Husain et al. [13] focused on finding features in the clinical data that have a close association with recurrent PVO, while having not exploited radiographs into the model exhibition.

Even though DL shows promising results in many medical fields, an automatic prediction model for PVO prediction which fully exploits clinical data and CT images is still missing. To solve these problems, we combine clinical data and CT image to build a prediction model based on machine learning that explicitly indicates whether a patient will suffer from PVO after surgery. Specifically, we put forward three machine learning-based methods:

- applying logistic regression with features selected from the clinical data;
- using a convolutional neural network (CNN) to extract features from patients' 3D cardiac CT images for prediction;
- building an architecture where clinical data features are combined with graphical features through an end-to-end trainable CNN.

In this paper, we have used the postoperative pulmonary venous obstruction prediction problem as a vehicle to explore the practicability of combining clinical data and CT images to evaluate the discriminative accuracy of a machine learning-based prediction model, and to integrate repeated-measures health examination data for prediction of PVO. Finally, we investigated the feasibility of our method for a subsequent prospective observational study.

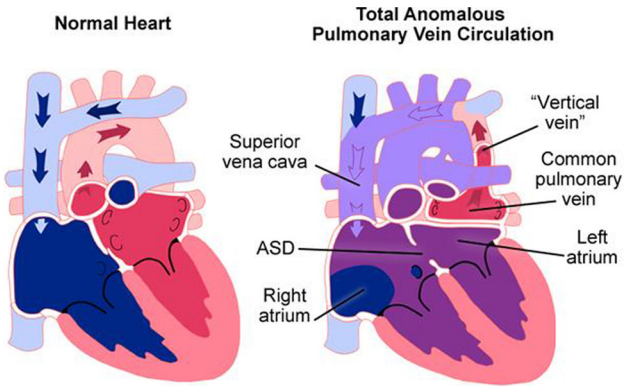


Fig. 1 A brief illustration of total anomalous pulmonary venous connection

Background

Congenital heart disease is the leading cause of mortality from birth defects, and total anomalous pulmonary venous connection (TAPVC) contributes to about 3% of all congenital heart diseases [14]. Specifically, about 1 in every 7,809 babies born in the USA each year are born with Total Anomalous Pulmonary Venous Connection [8]. 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

Guangdong Provincial People's Hospital, Guangdong Academy of Medical Sciences, Dongchuan Rd 96, Guangzhou 510080, China

the systemic venous systems. TAPVC has notoriously high mortality reaching nearly 80% without intervention [15]. Even with surgical repair, the death rate is still reported as 5–7%. Pulmonary venous obstruction is one of the most frequent causes of death after operations. An accurate prediction model can identify patients with high PVO recurrence risk, and then their chance of survival will be improved by early preventive treatment.

There are four types of TAPVC referred to as supracardiac, cardiac, intracardiac type, and mixed type. And the most common subtypes are supracardiac type which accounts for 30–50% of total TAPVC. In the supracardiac type, the common "ventricle vein" (showed in Fig. 1) connects to the superior vena cava system via an anomalous "vertical vein" , which is the main blood vessel that brings oxygen-poor blood from the upper part of the body to the heart.

Material and methodology

Dataset

As for the types of TAPVC, these four types have heterogeneous anatomical structures, from clinical perspectives it can be regarded as four different diseases. It is hard for a single prediction model to fit all distinct information, and a limited number in some types refrain our model construction. In our study, we choose the most common types, which account for 30–50% of the total number. Finally, 131 patients who were diagnosed with supracardiac TAPVC and received surgical treatment from January 1, 2009, to December 31, 2018, in Guangdong Provincial Hospital were enrolled. All patients with completely CT images and clinical data.

For CT images, all 131 patient's pre-operation CT images were used. For clinical data, suggested by expert cardiologists, we focused on 14 features, which were 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 were collected before the operation. Detail description of our dataset and features along with their ranges are summarized in Table 1. As for medical images, all 3D CT was captured by a Siemens Biography 64 CT scanner, and the typical voxel size were $0.25mm \times 0.25mm \times 0.5mm$.

Based on the patients' follow-up record after surgery (which was not part of this dataset and was not used in model training), when at least one of the following three conditions was met, there was postoperative PVO recurrence:

- Blood flow in the vertical vein or common trunk of pulmonary vein greater than 1.8 m/s;
- Atrial septal defect smaller than 3mm;
- Follow-up echocardiography diagnoses obstruction.

About 10% of the patients in the dataset were labeled with postoperative PVO.

The ratio of positive and negative samples, and the relatively limited number of our dataset comes from real-word clinical data introduce additional challenges to our prediction model.

To address these issues, we adopt a series of practical techniques regarding data preprocessing and learning strategies: i) image augmentation was utilized to increase the size of CT dataset; ii) for training epochs, positive cases were oversampled in an attempt to balance the biased distribution; iii) we modified the loss function, added L2 normalization of weights in the last fully connected layer of the CNN for generalization performance, and adjusted the weights of the PVO and non-PVO samples.

Clinical data based method

Prognosis about the onset of PVO after TPAVC correction could be treated as a binary classification problem. With only clinical features taken into consideration, each patient A_k was defined by several features $\{x_{k1}, x_{k2}, x_{k3}, ...\}$, and a target label y_k indicating whether PVO would occur after surgery (either 1 or 0). The problem was then to generate predictions $\widetilde{y_1}, \widetilde{y_2}, \widetilde{y_3}, ...$ and to optimize a defined loss function $\sum_k L(y_k, \widetilde{y_k})$. There are many methods for clinical features selection, among these methods, we found that logistic regression was the most effective method for PVO prognosis in practice, which is also the most prevalent method used in the medical problem research. Meanwhile, there are a number of techniques proposed to perform binary classification. Neural network is one of the artificial intelligence techniques that has many successful examples when applying to this problem [16–18].

To decide which features to feed into the prediction model, we initially selected a set of candidate features from clinical suggestion, and then we recursively prune features from the candidate set in the ascending order of the features' importance score until the optimal prediction accuracy was achieved.

CT image-based method

Cardiac CT images have never been used in the traditional prognosis of TAPVC repair. Yet CT images provide spatial

Table 1 The 14 candidate features in the clinical data that may be relevant to postoperative 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	$2.63 \sim 53.0\ kg$	**21.8**	hospital stay	$0 \sim 77\ d$	**0.819**
ALT	$5 \sim 948\ IU/L$	0.771	AST	$20 \sim 2420\ IU/L$	0.237
TBIL	$6.8 \sim 211.6\ \mu mol/L$	0.324	DBIL	$1.9 \sim 104.9\ \mu mol/L$	0.498
INR	$0.83 \sim 2.55$	**44.8**	PT	$31 \sim 140\ s$	0.040
cross-clamp	$0 \sim 153\ min$	0.409	CPB	$40 \sim 290\ min$	**1.38**
DHCA	$0 \sim 40\ min$	**4.35**	gender	$\{0, 1\}$	**19.0**
sutureless	$\{0, 1\}$	**55.4**	ligation	$\{0, 1\}$	**59.1**

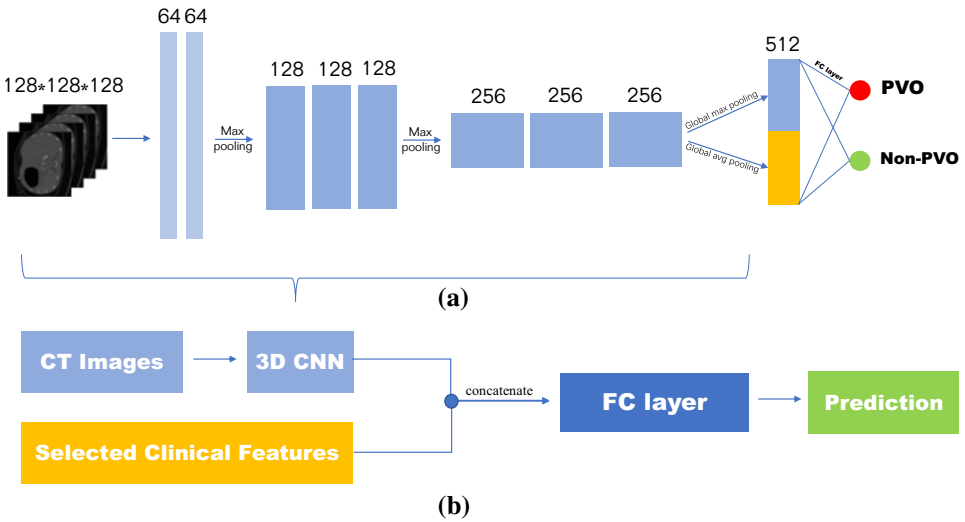


Fig. 2 **a** Architecture of CT Image-Based 3D CNN. The blue cuboid represents 3D feature maps, and the number of channels is marked above it. **b** Framework of joint clinical data and Image-Based postoperative PVO prediction method. *FC layer: fully connected layer*

information like anatomical structure and tissue development, which clinical data fails to provide. However, those graphical features are not specified by measurable parameters. Hence, we applied a convolutional neural network to extract information from patients' CT and then made a prediction on postoperative PVO. The input of the network was 3D images, and thus it resembled a 3D classification network. Many network structures such as VGG [19] and ResNet [20] are potentially well suited to handle such data. As our goal is to design a clinically usable predictive model for PVO rather than to compare the accuracy and efficiency of different networks, the 3D U-Net, as one of the most classical and easy to deploy neural network structures, is a very suitable choice. In this paper, we use it as our basic network structure. Because of the memory bottleneck induced by 3D medical images, very deep CNNs like VGG [19] and ResNet [20] was not suitable.

Figure 2a illustrates our 3D CNN's architecture. The original 3D cardiac CT images were first cropped to the region of interest and resized to a uniform shape, after which a batch of images was passed to the model. Three resolution stages consisted of our model, for each resolution stage, there were 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 replaced the max pooling with global pooling to flatten the feature maps in the last stage, which would be followed by a fully connected layer for predicting postoperative PVO. For the global pooling layer, inspired by [21], we used 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 were 3D counterparts. Besides, the number of resolution stages and initial filters were both adjustable. After comparison of different combinations, we decide on this architecture that strikes a balance between computational efficiency and generalization performance.

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 concatenated the selected features from clinical data (based on the method described in Sect. 3.2) with the flattened activations right before the fully connected layer in our 3D CNN,

as shown in Fig. 2(b). In this way, we built an end-to-end deep learning model making use of both CT images and clinical data, which we found demonstrates better performance than the methods using either one of the data sources solely.

Implementation details

Image augmentation

To make full use of all cardiac CT images in the limited dataset, we deployed classic augmentation methods to enlarge the dataset. Traditional image augmentation methods include translation, rotation, scaling, and flipping [22,23], To preserve the spatial structure and relative position of atrial and vessels, while we used a relatively conservative image augmentation methods, that we only adopted rotation to increase both the training set and test set. Firstly, we cropped the CT images so that the region of interest was centered. Then, we rotated the images clockwise by (10°, -10°, 90°, 180°, 270°). As a result, we could expand the dataset by six times. Noting that we first divided all the data into four folds with PVO patients evenly distributed. Only then did we augment the images in each fold and do fourfold cross-validation. This trick only works for the two methods described in Sect. 3.3 and Sect. 3.4, as it cannot expand clinical data.

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 were two main methods called over-sampling and under-sampling. Over-sampling was to add copies of samples from the under-presented class, and under-sampling was to remove instances belonging to the over-presented class. In practice, we found over-sampling fits our dataset better than under-sampling because the dataset was already small, and reducing training samples would compromise models' generalization performance.

Loss function Optimization

The loss function we choose for the 3D CNN was cross-entropy. However, during training, we found the training loss was smaller than the 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 was defined as:

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

in which B was the batch set, $s_k[class]$ was the softmax value of a class (PVO or non-PVO) at the output layer, w was the weight vector of the last fully connected layer, and β was an adjustable parameter which we set to be 2. Additionally, we assigned 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 a high false-positive rate. Our experiments suggest that the optimal weights for PVO and non-PVO classes are (30, 1).

Evaluation

The evaluation metrics were sensitivity,

specificity, positive predictive value (PPV) and negative predictive value (NPV), since they directly reflect the performance of our postoperative PVO prediction models. And the area under the ROC (Receiver Operating Characteristics) curve - AUC, was reported as an important measurement of a classification model's performance. In our setting, the AUC showed the ability of a method to predict the postoperative PVO recurrence of a patient. And for better explaination of our model to physicians, we used saliency maps for promoting understandability of the methods.

Result

For the 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 scores are obtained from the absolute value of features' coefficients in the logistic regression model, as listed in Table 1. We gradually remove some features with the minimum importance score from the 14 candidates. We have observed that, when the number of predictors is eight, the logistic regression model achieves the optimal performance. The finally selected 8 features as postoperative PVO predictors in the clinical data-based method include operation weight, hospital stay, INR, CPB, DHCA, gender, sutureless, and ligation.

For the impact of the techniques described in Sect. 3.5 on the prediction methods, the average AUC scores of fourfold cross-validation are shown in Table 2. We can notice that the joint method with 8 clinical features and L2 norm achieves the highest average AUC score of 0.943. Specifically, the L2 norm can improve the average AUC score of 0.025 in the joint method when using 8 clinical features only, and the joint method with 8 clinical features outperforms those with 14 by 0.132 when both methods use the L2 norm. This suggests that feature pruning is not only effective for clinical data-based method, but also the joint one. The highest average AUC

Table 2 The prediction performance of the methods with different configurations and techniques in fourfold cross validation. The item "8" or "14" indicates the number of adopted clinical features. For the CT image-based method and the joint method, over-sampling is always applied

Methods	Fold 1	Fold 2	Fold 3	Fold 4	Average
CT Image-Based (with L2 norm)	0.605	0.615	0.520	0.618	0.590
Clinical Data-Based (8, with over-sample)	0.823	0.801	0.797	0.813	0.801
Clinical Data-Based (8, w/o over-sample)	0.813	0.805	0.792	0.759	0.792
Joint Method (8, with L2 norm)	0.973	0.962	0.916	0.915	**0.943**
Joint Method (8, w/o L2 norm)	0.947	0.936	0.885	0.907	0.918
Joint Method (14, with L2 norm)	0.814	0.921	0.774	0.735	0.811

of the clinical data-based method is 0.801 with 8 features involved, and the over-sample method can slightly improve the performance by 0.009. The highest average AUC of the CT image-based method is 0.590, which is the lowest in the prediction methods. The ROC curves of the optimal configurations of the three methods are shown in Fig. 3. The joint method is consistently better than the other two, shows the benefits our prediction model gained from features extracted automatically from CT images.

The joint method also achieves the highest sensitivity of 0.828 and the highest PPV of 0.864, which is a significant difference between the other methods. The joint method has a slightly improvement in specificity and NPV, which reaches 0.910 and 0.972, respectively. The joint method has a significant improvement in the prediction of pvo compared with other methods. However, the performance of the different methods is basically stable non-pvo predictions. Detailed evaluation results can be found in Fig. 4.

The saliency maps in Fig. 5 are used to explain the prediction results of the joint method. Saliency maps are shown in the different slices, original CT images are pixeled in $128\times128\times128$. The joint method has more activation region which is concentrated in the pulmonary area roughly, while the CT image-based method is activated more decentralized. All the identified imaging features for the methods are hard to understand for the clinicians.

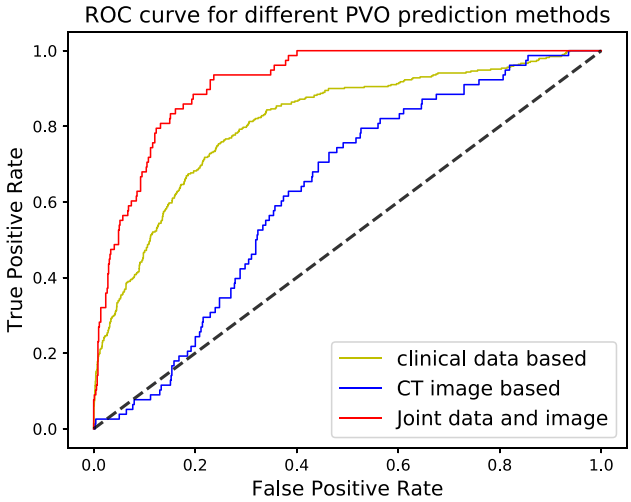
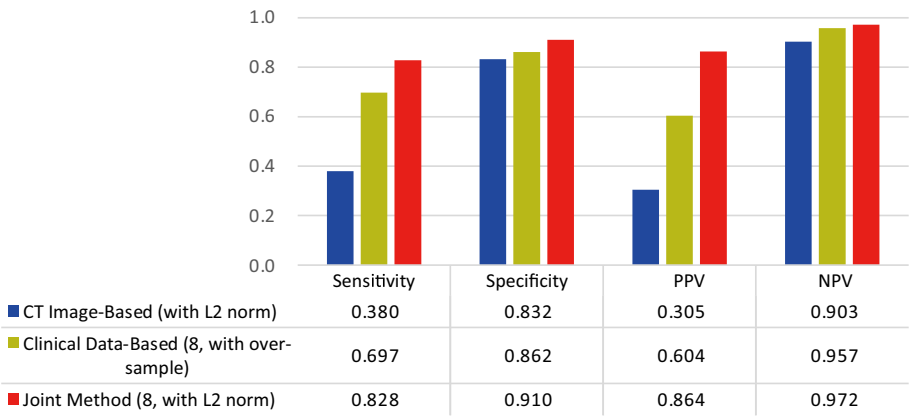


Fig. 3 ROC curves of the three different prediction methods. All the techniques in Sect. 3.5 are applied as appropriate on the three methods. The joint method is consistently better than the other two

Discussion

The goal of this paper is to showcase the first attempt and methodology toward a machine learning-based pulmonary venous obstruction prediction model among supracardiac TAPVC repair patients. Our prediction model offers the possibility to integrate typical multi-modality clinical data like ALT, INR, and individual CT images, etc. into a neural network without assuming all features are independent of each other. Among the selected clinical features, the surgery technique [24], DHCA [25], gender [26,27], and operation weight [27] are researched to be related to PVO according to previous studies. Meanwhile, the operation information (CPB, cross-clamp, ligation, and hospital stay) [24–26] were extremely useful, as the statistical variability between the PVO and non-PVO groups is extremely significant in many comparative studies. And many studies have found that the information of patient-specific laboratory test results (ALT, AST, INT, PT, TBIL, DBIL) [28,29] can be used as a non-invasive test to predict specific clinical medical events, which is not fully used in the previous PVO prediction model. Our method fully uses the multi-modality data, and provides a possible method for better prediction of PVO events. Compared with results from clinical data and CT images only based method, our model shows similar performance reported in other references [2,30]. Although for the sake of clarity, we choose not to enumerate all the possible combinations in clinical data only method, while the six combinations we illustrated are sufficient to observe the effectiveness of the techniques.

To decide which features to feed into the prediction model, we initially selected a set of candidate features which cardiologists believe to have direct or indirect effects on the recurrence of PVO. For example, the sutureless operation is believed to have a potential impact on the geometric distortion of the pulmonary venous suture line and thus on the postoperative PVO, according to the previous study [24]. Then, we recursively prune features from the candidate set



	Sensitivity	Specificity	PPV	NPV
CT Image-Based (with L2 norm)	0.380	0.832	0.305	0.903
Clinical Data-Based (8, with over-sample)	0.697	0.862	0.604	0.957
Joint Method (8, with L2 norm)	0.828	0.910	0.864	0.972

Fig. 4 Evaluation results of sensitivity, specificity, PPV and NPV among the three different prediction methods. All the techniques in Sect. 3.5 are applied as appropriate on the methods. The joint method is consistently better than the other two. *PPV: Positive predictive value; NPV: Negative predictive value*

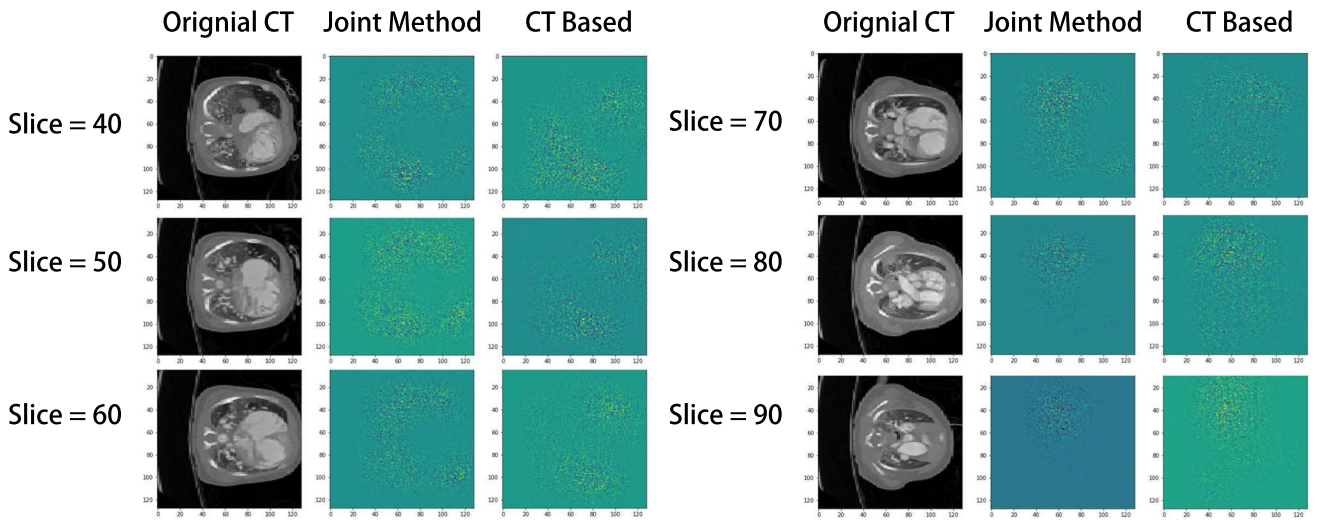


Fig. 5 Saliency map for the two image involved method. The joint method has more activation area which is concentrated in the pulmonary area roughly, while the CT image-based method is activated more decentralized

in the ascending order of the features' importance score until the optimal prediction accuracy was achieved. Such an iterative pruning can help to remove redundant information and reduce potential overfitting.

For image augmentation methods we used in the data preprocessing, it is worth noting that traditional image augmentation methods include limited translation, rotation, scaling, and flipping, etc, and all of those methods could preserve the spatial structure and relative position of vessels. However, in our work, we find that it is not clear to what extent the transformation is limited, and we are concerned that it might be difficult to interpret the transformation of the image data for physicians. Therefore, we temporarily adopt a relatively conservative image data augmentation approach. To some extent, this is a limitation of our study. We will further investigate this challenge in our future work as well, seeking a suitable balance between image data augmentation methods and physician interpretation.

By more deeply examining the various test results, we notice that in a certain degree we cannot guarantee that models based only on clinical data will not have the possibility of overfitting. In other words, higher AUC values may appear when we further select clinical parameters for predictive model construction, or further, adopt more network training skills. However as results overfitting in a single dataset almost an inherent shortcoming of a neural network, it is acceptable not to pursue the highest average AUC value for clinical use. To balance the implications of our results with the potential limitations of observational studies, during our collaboration with clinicians, we conclude that clinical predictive models should be based on clinical utility, prediction model results and model scalability should be balanced to some degree.

A good, but not the best, model it is acceptable, and is sufficient evidence to support the feasibility of a subsequent prospective observational study. Although our dataset is very scarce, which is the largest and the only study that attempts to combine clinical data and CT image in supracardiac TAPVC patients in the world as far as we know, but the disadvantages of a single-center retrospective research study still exist. Our

results only did the traditional fourfold cross-validation, and using another external validation dataset to test our method is needed. However, owing to the diseases we study is still a rare disease and the high cost of follow-up, the amount of data is very small in a single center. Moreover, it is difficult to convince other medical centers to share their data with our team, these limitations limit the clinical efficacy of our method. To verify the clinical efficacy of our model, one possible solution is to collect more data through prospective observational experiment in our hospital in the future, and use this kind of prospective data as an external dataset. We hope that this will advance the application of our method in the supracardiac TAPVC.

Moreover, it is important to highlight that whilst it is not our objective to detail the the mathematical basis of machine learning, we aim to underscore how such a mathematical approach could constructively impact medicine. Our use of machine learning has concluded that integrate clinical data and CT images to a postoperative complication prediction model is possible. In future research, we plan to extend our method to the dataset that comes from our prospective observational study. Further, the development will focus on the actual impact of clinical doctors and patients.

Statement of Clinicians: We are not completely convinced by the PVO predictions derived from this model, but we can use it as a reference. The main reason for not being able to trust it is our inability to understand how the model was processed. The output and presentation of high-dimensional data are not entirely satisfactory, and there is no satisfactory answer as to what constitutes these correlations. The clinical focus is more on knowledge and more on causality rather than on correlation. But if the associations are accurate, it may be acceptable to accept the predictions of the model. Even if the model is accurate, predictions derived from single-center data need to be prospectively validated by multiple centers, to achieve a high-level evidence under the hierarchy of evidence-based medicine.

Conclusion

In this paper, we explored the practicability of combining clinical data and CT images for PVO prediction. A novel neural network architecture that jointly learns from clinical data and CT images in an end-to-end trainable manner are built. We also introduced a group of implementation tips involving data preprocessing and learning to manipulate the limited and biased dataset pertinent to the disease. The significant improvement of experimental results demonstrate the advantage of the proposed method of our joint learning method. Finally, the feasibility of our method proffers enough evidence for a subsequent prospective observational study.

Funding This work was funded by the National Key Research and Development Program of China [2018YFC1002600], the Science and Technology Planning Project of Guangdong Province, China [No.2017A070701013, 2017B090904034, 2017B030314109, 2018B090944002, 2019B020230003], Guangdong Peak Project [DFJH201802].

Declarations

Conflicts of interest The authors declare that they have no conflict of interest.

Ethical approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Declaration of Helsinki and its later amendments or comparable ethical standards.

Informed consent For this type of study formal consent is not required.

References

1. Hannan EL, Wu C, Bennett EV, Carlson RE, Culliford AT, Gold JP, Higgins RS, Smith CR, Jones RH (2007) Risk index for predicting in-hospital mortality for cardiac valve surgery. The Annals of Thoracic Surgery 83(3):921–929
2. Siregar S, Nieboer D, Versteegh MI, Steyerberg EW, Takkenberg JJ (2019) Methods for updating a risk prediction model for cardiac surgery: a statistical primer. Interactive cardiovascular and thoracic surgery 28(3):333–338
3. Parker DM, Everett AD, Stabler ME, Vricella L, Jacobs ML, Jacobs JP, Parikh CR, Pasquali SK, Brown JR (2020) Novel biomarkers improve prediction of 365-day readmission after pediatric congenital heart surgery. The Annals of thoracic surgery 109(1):164–170
4. Hippisley-Cox J, Coupland C, Vinogradova Y, Robson J, May M, Brindle P (2007) Derivation and validation of qrisk, a new cardiovascular disease risk score for the united kingdom: prospective open cohort study. Bmj 335(7611):136
5. D'Agostino RB, Grundy S, Sullivan LM, Wilson P (2001) Validation of the framingham coronary heart disease prediction scores: results of a multiple ethnic groups investigation. Jama 286(2):180–187
6. Goldstein BA, Navar AM, Carter RE (2017) Moving beyond regression techniques in cardiovascular risk prediction: applying machine learning to address analytic challenges. European heart journal 38(23):1805–1814
7. Pasquali SK, Gaies M, Banerjee M, Zhang W, Donohue J, Russell M, Gaynor JW (2019) The quest for precision medicine: unmeasured patient factors and mortality after congenital heart surgery. The Annals of thoracic surgery 108(6):1889–1894
8. Kennedy EH, Wiitala WL, Hayward RA, Sussman JB (2013) Improved cardiovascular risk prediction using nonparametric regression and electronic health record data. Medical care 51(3):251
9. Deo RC (2015) Machine learning in medicine. Circulation 132(20):1920–1930

10. Miotto R, Wang F, Wang S, Jiang X, Dudley JT (2018) Deep learning for healthcare: review, opportunities and challenges. Briefings in bioinformatics 19(6):1236–1246
11. Weng SF, Reps J, Kai J, Garibaldi JM, Qureshi N (2017) Can machine-learning improve cardiovascular risk prediction using routine clinical data? PloS one 12(4):e0174944
12. Parisot S, Ktena SI, Ferrante E, Lee M, Moreno RG, Glocker B, Rueckert D (2017) Spectral graph convolutions for population-based disease prediction. In: International conference on medical image computing and computer-assisted intervention, Springer, pp 177–185
13. Husain SA, Maldonado E, Rasch D, Michalek J, Taylor R, Curzon C, Neish S, Calhoon JH (2012) Total anomalous pulmonary venous connection: factors associated with mortality and recurrent pulmonary venous obstruction. The Annals of thoracic surgery 94(3):825–832
14. Herlong JR, Jaggers JJ, Ungerleider RM (2000) Congenital heart surgery nomenclature and database project: pulmonary venous anomalies. The Annals of thoracic surgery 69(3):56–69
15. Burroughs JT, Edwards JE (1960) Total anomalous pulmonary venous connection. American heart journal 59(6):913–931
16. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, Thrun S (2017) Dermatologist-level classification of skin cancer with deep neural networks. Nature 542(7639):115–118
17. Krinitskiy M, Verezemskaya P, Grashchenkov K, Tilinina N, Gulev S, Lazzara M (2018) Deep convolutional neural networks capabilities for binary classification of polar mesocyclones in satellite mosaics. Atmosphere 9(11):426
18. Veredas F, Mesa H, Morente L (2009) Binary tissue classification on wound images with neural networks and bayesian classifiers. IEEE transactions on medical imaging 29(2):410–427
19. Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:14091556
20. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 770–778
21. Yuan B, Xing W (2019) Diagnosing cardiac abnormalities from 12-lead electrocardiograms using enhanced deep convolutional neural networks. In: Machine Learning and Medical Engineering for Cardiovascular Health and Intravascular Imaging and Computer Assisted Stenting, Springer, pp 36–44
22. Roth HR, Lu L, Liu J, Yao J, Seff A, Cherry K, Kim L, Summers RM (2015) Improving computer-aided detection using convolutional neural networks and random view aggregation. IEEE transactions on medical imaging 35(5):1170–1181
23. Setio AAA, Ciompi F, Litjens G, Gerke P, Jacobs C, Van Riel SJ, Wille MMW, Naqibullah M, Sánchez CI, van Ginneken B (2016) Pulmonary nodule detection in ct images: false positive reduction using multi-view convolutional networks. IEEE transactions on medical imaging 35(5):1160–1169
24. Wu Y, Xin L, Zhou Y, Kuang H, Jin X, Li Y, Wu C (2019) Is sutureless technique beneficial in the primary repair of total anomalous pulmonary venous connection? a systematic review and meta-analysis. Pediatric cardiology 40(5):881–891
25. Padalino MA, Cavalli G, De Franceschi M, Mancuso D, Maschietto N, Vida V, Milanesi O, Stellin G (2014) Surgical outcomes of total anomalous pulmonary venous connection repair: a 22-year experience. Journal of Cardiac Surgery: Including Mechanical and Biological Support for the Heart and Lungs 29(5):678–685
26. Sakamoto T, Nagashima M, Umezu K, Houki R, Ikarashi J, Katagiri J, Yamazaki K (2018) Long-term outcomes of total correction for isolated total anomalous pulmonary venous connection: lessons from 50-years' experience. Interactive CardioVascular and Thoracic Surgery 27(1):20–26
27. Harada T, Nakano T, Oda S, Kado H (2019) Surgical results of total anomalous pulmonary venous connection repair in 256 patients. Interactive cardiovascular and thoracic surgery 28(3):421–426
28. Furukawa T, Akimoto K, Ohtsuki M, Sato K, Suzuki M, Takahashi K, Kishiro M, Shimizu T, Kawasaki S (2011) Non-invasive assessment of liver fibrosis in patients after the fontan operation. Pediatrics International 53(6):980–984
29. Valle JA, Graham L, Derussy A, Itani K, Hawn MT, Maddox TM (2014) Association of pre-operative inr evaluation with risk of major adverse cardiac events and bleeding complications in cardiac patients undergoing non-cardiac surgery. Circulation: Cardiovascular Quality and Outcomes 7(suppl 1):A284–A284
30. Quesada JA, Lopez-Pineda A, Gil-Guillén VF, Durazo-Arvizu R, Orozco-Beltrán D, López-Domenech A, Carratalá-Munuera C (2019) Machine learning to predict cardiovascular risk. International journal of clinical practice 73(10):e13389