Improved Esophageal Varices Assessment from Non-Contrast CT Scans

Chunli Li[1,2,3*], Xiaoming Zhang[2,3(✉,*)], Yuan Gao[2,3*], Xiaoli Yin[1], Le Lu[2], Ling Zhang[2], Ke Yan[2,3✉], and Yu Shi[1✉]

[1] Department of Radiology, Shengjing Hospital of China Medical University, 110004, Shenyang, China
[2] DAMO Academy, Alibaba Group
[3] Hupan Lab, 310023, Hangzhou, China
zxiaoming360@gmail.com; yanke.yan@alibaba-inc.com; 18940259980@163.com

Abstract. Esophageal varices (EV), a serious health concern resulting from portal hypertension, are traditionally diagnosed through invasive endoscopic procedures. Despite non-contrast computed tomography (NC-CT) imaging being a less expensive and non-invasive imaging modality, it has yet to gain full acceptance as a primary clinical diagnostic tool for EV evaluation. To overcome existing diagnostic challenges, we present the Multi-Organ-cOhesion-Network (MOON), a novel framework enhancing the analysis of critical organ features in NC-CT scans for effective assessment of EV. Drawing inspiration from the thorough assessment practices of radiologists, MOON establishes a cohesive multi-organ analysis model that unifies the imaging features of the related organs of EV, namely esophagus, liver, and spleen. This integration significantly increases the diagnostic accuracy for EV. We have compiled an extensive NC-CT dataset of 1,255 patients diagnosed with EV, spanning three grades of severity. Each case is corroborated by endoscopic diagnostic results. The efficacy of MOON has been substantiated through a validation process involving multi-fold cross-validation on 1,010 cases and an independent test on 245 cases, exhibiting superior diagnostic performance compared to methods focusing solely on the esophagus (for classifying severe grade: AUC of 0.864 versus 0.803, and for moderate to severe grades: AUC of 0.832 versus 0.793). To our knowledge, MOON is the first work to incorporate a synchronized multi-organ NC-CT analysis for EV assessment, providing a more acceptable and minimally invasive alternative for patients compared to traditional endoscopy.

Keywords: Esophageal varices assessment · Liver · Spleen · CT

1 Introduction

Esophageal varices (EV) are a significant complication often stemming from chronic liver diseases, with portal hypertension being the pivotal cause driving

---

blood into smaller collaterals within the lower esophagus [7], potentially leading to life-threatening hemorrhages and shock. Detecting EV at an early stage is difficult, with confirmation usually requiring invasive endoscopic procedures. Given that the extent of venous dilation correlates with hemorrhage risk, a precise assessment of EV severity is vital.

Invasive endoscopy is the conventional method for EV evaluation and, despite its effectiveness, it is associated with discomfort, potential complications like infection and bleeding, and higher healthcare costs. As a less invasive approach, dynamic contrast-enhanced CT (DCE-CT) allows for detailed visualization of EV and related collaterals, improving patient compliance and reducing discomfort and bleeding risks linked to endoscopy. However, methods such as Yan et al.'s [26], which use radiomic features from DCE-CT, have limited generalizability, and iodinated contrast agents pose risks of adverse reactions.

Non-contrast CT (NC-CT) offers a quick and low-radiation alternative, chest NC-CT scanning can be valuable for grading EV. Liver characteristics such as fibrosis and alterations in volume can indicate the pressure within the portal venous [20], which is critical in appraising EV risk and severity. Portal hypertension-induced changes in the spleen, such as enlargement, indirectly reflect this pressure and suggest EV bleeding risk. Therefore, a combined analysis of the liver, spleen, and esophagus could improve the understanding and grading of EV risks, and enhance clinical decisions. However, it faces several hurdles in EV assessment: differentiation difficulties due to lower contrast resolution, limited visibility of smaller varices, and the inconsistent distribution of varices throughout the esophagus.

Previous attempts to evaluate EV have focused on DCE-CT imaging of the liver, spleen, esophagus, and portal venous system's vascular features [17,24,26]. However, no current grading methods fully consider the esophagus's close relationships with adjacent structures. To optimize the assessment of EV, we have drawn upon radiologists' comprehensive diagnostic techniques and physicians' clinical diagnostic knowledge, introducing the framework of Multi-OrgancOhesion-Network (MOON) based on chest NC-CT. Our approach enhances the accuracy of EV evaluation and surpasses previous radiomics methods reliant on DCE-CT imaging.

MOON adopts a holistic strategy, simultaneously analyzing the regions of interest (ROIs) from the NC-CT scans of these organs. It employs a multiorgan framework that incorporates the Organ Representation Interaction (ORI) module, allowing features from the liver and spleen to inform the analysis of esophageal characteristics at each step within the network. Additionally, the Hierarchical Feature Enhancement (HFE) module is designed to highlight the esophagus' distinct morphology by synthesizing features across various levels, focusing on areas that exhibit abnormal varices. To achieve uniformity in decision-making across the various organ domains, which are vital for an accurate final diagnosis, we incorporate ordinal regression loss [3] and canonical correlation analysis (CCA) [1] loss into our training process. This synergistic use of loss

functions fosters a cohesive feature integration, thereby reducing inconsistencies in the decision-making process and enhancing the overall diagnostic performance.

Our studied dataset is sufficiently extensive, comprising chest NC-CT scans from 1255 patients diagnosed with EV, each confirmed via endoscopic examination and categorized into three differentiated levels of severity. The goal of MOON is to predict the grade of EV. It exhibits strong diagnostic proficiency, attaining an AUC performance of **0.864** versus 0.736 using DCE-CT [26] for severe EV cases and, **0.832** versus 0.802 using DCE-CT [24] for moderate to severe cases, showcasing its effectiveness in evaluating EV from NC-CT imagery.

2 Method

Deep learning, particularly convolutional neural networks (CNNs) [8,12,13], have significantly advanced the detection of pathologies or diseases with subtle imaging differences in Hounsfield Units on NC-CT images, often imperceptible to radiologists [25]. There are four key designs in our method: (1) To assess EV, a global receptive field is essential. Uniformer [14] combines convolution with self-attention, which is utilized to adeptly capture esophageal features, providing the comprehensive coverage necessary for accurate EV evaluation. (2) An effective EV assessment necessitates combining data from multiple organs. Our ORI module facilitates this, integrating liver and spleen features with those of the esophagus to improve EV detection. (3) The diversity in varices sizes and esophageal anatomy requires a multi-scale strategy. We answer this need with the HFE module, which assimilates features at various scales to enhance classification adaptability. (4) For effective integration across multiple organ branches and precise identification of the severity grades of EV, we train our model using a hybrid loss function that combines ordinal regression and CCA. This approach enables the model to maintain the ordinal relationships inherent to the various EV severity levels while enhancing the correlation between multi-organ features, reducing bias from reliance on any single organ, and improving the overall diagnostic accuracy and generalizability of the model.

2.1 Multi-Organ-cOhesion-Network (MOON)

Pipeline of MOON. The UniFormer architecture [14], combining local and global feature extraction, addresses the medical imaging challenge of accurately assessing EV by providing a comprehensive receptive field suitable for the esophagus's unique structure, thereby enabling precise identification of esophageal features without the constraints of traditional 3D CNNs [9,23] or Vision Transformers [5]. MOON, depicted in Fig. 1(a), is a multi-stage, multi-organ framework incorporating UniFormer as its core for each branch. It starts with an off-the-shelf nnUNet [9] for organ segmentation to localize the esophagus, liver, and spleen. Post-segmentation, respective image data are fed dedicated branches to extract features across scales. Within this framework, the ORI module plays a pivotal role, adeptly handling the unique anatomical features of the organs,
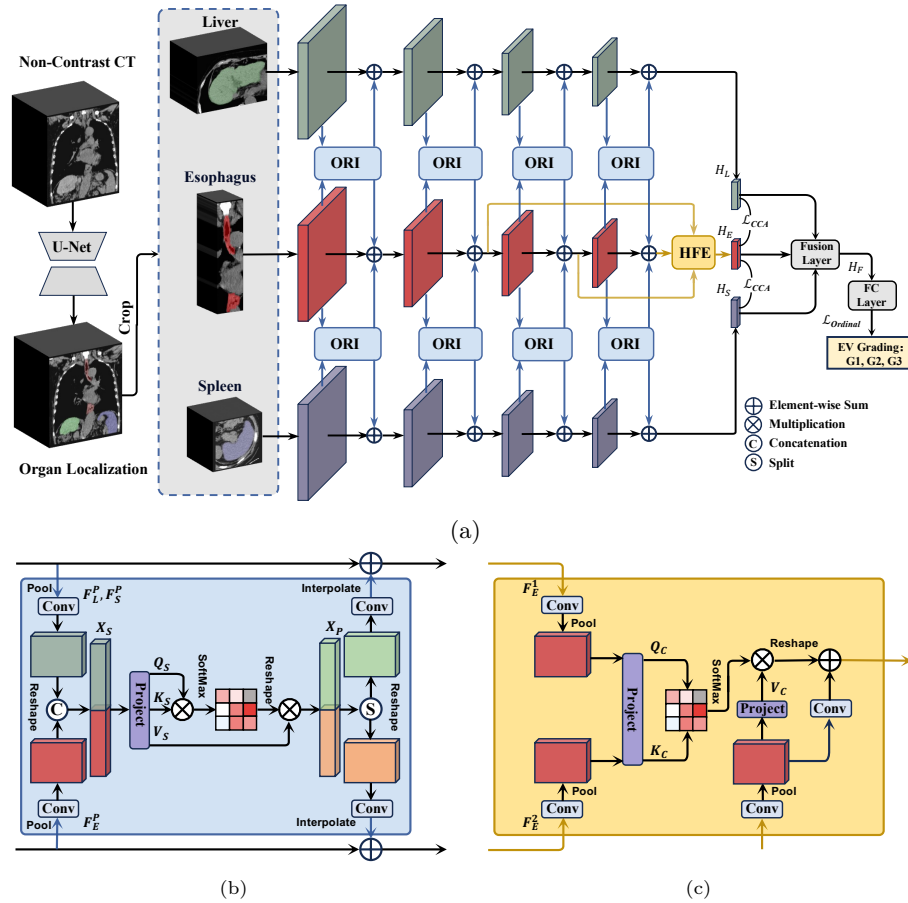
Fig. 1: An overview of Multi-Organ-cOhesion-Network (MOON). (a) The pipeline of multi-stage pipeline. (b) Organ representation interaction (ORI). (c) Hierarchical feature enhancement (HFE).

e.g., the tendency for spleen enlargement and liver lesions in the portal venous system. Its capacity to engage with each organ's distinct properties is vital for accurate EV assessment. The HFE module, in turn, is specifically designed to enhance the esophagus branch, which is essential given that experiments in Sec. 3 indicate esophageal features are most indicative of EV. Through its focus on the esophagus, the HFE module promotes effective feature integration and sharpens the identification of key esophageal details for precise EV detection.

Organ Representation Interaction. To address the nuanced diagnosis of EV stemming from portal hypertension, a comprehensive CT scan analysis of the liver and spleen is required due to the subtle morphological and texture changes they undergo. As shown in Fig. 1(b), the ORI module is designed to enhance the interpretation of these images by refining liver and spleen imaging

features into a coherent form, directing attention to critical areas. It streamlines the imaging features of these organs, amplifying essential characteristics and employing a self-attention mechanism that operates on tokenized 3D representation to link esophageal features to hepatic and splenic indicators of EV. Given that the features of the esophagus, liver, and spleen at each stage, defined as $\mathbf{F}_E \in \mathbb{R}^{H_e \times W_e \times D_e \times C}$, $\mathbf{F}_L \in \mathbb{R}^{H_l \times W_l \times D_l \times C}$, and $\mathbf{F}_S \in \mathbb{R}^{H_s \times W_s \times D_s \times C}$, respectively. $\mathbf{F}_E^P, \mathbf{F}_L^P, \mathbf{F}_S^P \in \mathbb{R}^{H_n \times W_n \times D_n \times C}$ are pooled features. An attention-ready form $X_S \in \mathbb{R}^{H_n W_n \times D_n \times 2C}$ is produced by concatenation and reshaping. The *query* (Q_S), *key* (K_S), and *value* (V_S) matrices are generated from X_S through learned transformations. The scaled dot-product attention is applied as $\text{Attention}(Q_S, K_S, V_S) = \text{SoftMax}\left(\frac{Q_S K_S^T}{\sqrt{d_k}}\right) V_S$, with d_k representing the dimensionality of the keys, allowing the model to capture feature interactions across the global scope of the input space. The outputs from the attention mechanism are aggregated into a tensor T_S and then projected to an enhanced feature space, resulting in $X_P = T_S W_P$, where the projection matrix W_P is of size $\mathbb{R}^{2C \times 2C}$. This process consolidates the learned information of adjacent organs and infuses them into the esophageal representation, thereby strengthening the subtle interconnections specific to EV. By finally splitting and then reintegrating X_P, followed by refinement through convolution and subsequent interpolation, ORI module achieves an interwoven feature representation.

Hierarchical Feature Enhancement. The HFE module refines the central branch that analyzes esophageal features, sharpening the detection of localized characteristics indicative of EV. It advances beyond former multi-scale fusion methods [15,22] by implementing a cross-attention mechanism. This feature promotes interaction among different depth levels and conveys detail-rich information from the shallow layers, imperative for EV detection, to the in-depth features that represent localized details within an expansive anatomical context. Illustrated in Fig. 1(c), through a process of convolution, pooling, and cross-attention, the HFE module adeptly blends shallow and deep features. This strategic combination solidifies the main branch's feature representation, bolstering the network's overall diagnostic accuracy. The HFE module uses convolution and pooling to adjust the dimensions of features from varying depths, starting with the intermediate $\mathbf{F}_E^1$ and moving to the deeper $\mathbf{F}_E^2$, ultimately aligning them with the dimensions of the deepest esophagus feature: $\mathbf{F}_E^3 \in \mathbb{R}^{H_n \times W_n \times D_n \times C}$. After pooling, these features are reorganized into an attention-ready form $X_C \in \mathbb{R}^{H_n W_n \times D_n \times C}$. Using a form of cross-attention, the attention mechanism then transforms X_C to generate the *query* (Q_C), *key* (K_C), and *value* (V_C) matrices, where Q_C originates from the intermediate features $\mathbf{F}_E^1$, K from the deeper features $\mathbf{F}_E^2$, and V from the deepest features $\mathbf{F}_E^3$. The scaled dot-product attention is applied as $\text{Attention}(Q_C, K_C, V_C) = \text{SoftMax}\left(\frac{Q_C K_C^T}{\sqrt{d_k}}\right) V_C$. The HFE module orchestrates the selective enhancement of local features and their integration with deeper, more globally aware features. By fusing shallow and deep features through the HFE module, with attention-focused features refined by point-wise convolutions for a more robust representation.

2.2   Training Paradigm of MOON

MOON is proposed and designed to tackle the intricacies of EV diagnosis by integrating multi-organ information. Within this scope, we deploy an ordinal regression loss, symbolized as $\mathcal{L}_{\text{Ordinal}}$, which is applied to the unified logits $\mathbf{H}_F$. These logits consolidate data from the esophagus, liver, and spleen branches, in line with [3]. The ordinal regression loss is crucial for measuring the model's capability to accurately classify the different stages of EV, as it verifies the alignment between $\mathbf{H}_F$ and the ground truth labels $\mathbf{Y}$. Additionally, the framework integrates CCA loss $\mathcal{L}_{\text{CCA}}$, from multi-view learning, which is key to maximizing the correlation between two feature sets for improving cross-modal retrieval [1]. Unlike approaches that seek complementary features, MOON employs $\mathcal{L}_{\text{CCA}}$ to capitalize on the interconnectedness of organ features by amplifying their correlation. Specifically, it targets the alignment of logits from the esophageal branch ($\mathbf{H}_E$) with those from the liver ($\mathbf{H}_L$) and spleen ($\mathbf{H}_S$) branches. By serving as a regularization mechanism, $\mathcal{L}_{\text{CCA}}$ facilitates the convergence of characteristics derived from multiple organs, thus enhancing the model's capability to interpret interrelated organ information.

The total loss $\mathcal{L}_{\text{Overall}}$ for a given batch is a composite of the $\mathcal{L}_{\text{Ordinal}}$ and $\mathcal{L}_{\text{CCA}}$, weighted appropriately, that not only accounts for accurate staging of EV but also reinforces the interconnectedness of the multi-organ representations:

$$\mathcal{L}_{\text{Overall}} = \lambda \cdot \mathcal{L}_{\text{Ordinal}}(\mathbf{H}_F, \mathbf{Y}) + (1-\lambda) \cdot (\mathcal{L}_{\text{CCA}}(\mathbf{H}_E, \mathbf{H}_L) + \mathcal{L}_{\text{CCA}}(\mathbf{H}_E, \mathbf{H}_S)). \quad (1)$$

3   Experiments

Datasets. We collected a dataset in a retrospective manner, with NC-CT images stratified into three grades of EV severity as outlined by [4,10]: $G1$ for *Mild* ($F1$, RC-), $G2$ for *Moderate* ($F1$, $RC+$ or $F2$, RC-), $G3$ for *Severe* ($F2$, $RC+$, $F3$, $RC+$ or $F3$, RC-). This dataset comprises images from 1,010 patients diagnosed with varying stages of EV ($G1$: 331, $G2$: 252, $G3$: 427), acquired with Philips Ingenuity 4 and Siemens Sensation 64 CT scanners. All patients were confirmed to have EV through endoscopy, and all CT scans were performed within one month prior to the endoscopic examination. The collected images were processed using a standardized abdominal window setting. To rigorously evaluate the proposed method, a 5-fold cross-validation approach was employed. Additionally, an independent test cohort comprising 245 subjects was curated from a single center to substantiate the method's validity. Eligibility for data inclusion required being over 18 years old and not having received treatment before the CT scans.

Implementation Details. Throughout the training process, several data augmentation techniques were applied, such as random rescaling, flipping, and cutout operations. For localizing the esophagus, liver, and spleen, we utilized the pretrained nnUNet [9] configured for low-resolution image processing as our localization network. Once localized, the ROIs for the esophagus, liver, and spleen were resized to dimensions of $40 \times 40 \times 100$, $256 \times 196 \times 36$, and $152 \times 196 \times 24$, respectively. To construct each branch of the MOON, we employed the Uniformer-S

model [14], which was pre-trained on the Kinetics-600 [2]. The MOON framework was trained employing the Adam [11] optimizer with an initial learning rate set to 10^{-5}. The training spanned across 100 epochs, with the learning rate being reduced by a factor of two every 20 epochs, following a piece-wise constant decay schedule. We set $\lambda = 0.8$ for Eqn. (1). The framework was implemented using Pytorch and run on Nvidia V100 GPUs.

Main Experimental Results. Table 1 consolidates experimental results demonstrating that for evaluating EV, single features derived from the esophagus alone surpass those from the liver or spleen in terms of accuracy. This underscores the critical importance of the esophagus in EV detection. As a result of portal hypertension, there is a marked enlargement of the spleen, potentially leading to changes in the internal textural characteristics of the organ. These alterations can be detected by neural networks. Hence, in terms of accuracy during 5-fold cross-validation, spleen features rank second, marginally surpassing those from the liver. Compared to single-organ input, integrating data from multiple organs achieved a better outcome with an increase of 0.003 to 0.039 in AUC for the evaluation of G3 on an independent test set. However, the performance improvement for assessing ≥G2 was not significant. This may be due to the fact that solely relying on post-fusion strategies does not adequately and effectively integrate features originating from multiple organs. The incorporation of the ORI and HFE modules into MOON has led to a performance that outperforms simple post-fusion strategies. In an independent test set, the AUC for G3 evaluation saw an increase of 0.006-0.040, and for the single-organ esophageal model, there was a boost of 0.038-0.061. Moreover, in the assessment of ≥G2, there was also a significant enhancement in AUC (0.016-0.034) relative to simple post-fusion strategies.In the fusion layer, we employ several strategies, with concatenation [6] emerging as the most effective, particularly in the challenging G3 on the independent test set. G3 grades show higher detection accuracy compared to ≥G2, likely due to G3's clearer clinical features. Additionally, FiLM [18] stands out in 5-fold validation, achieving the highest accuracy. These findings emphasize the significance of advanced post-fusion techniques in enhancing the accuracy of multi-organ EV, further augmented by ORI and HFE.
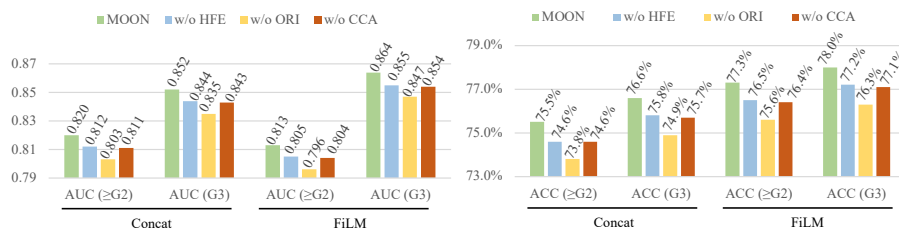

Fig. 2: Ablation experiments on different strategies on 5-fold validation.

Table 1: Comparative analysis between single-organ and multi-organ methodologies, examining various strategies for feature fusion. ‡ denotes that without the interaction from ORI and HFE modules. ACC(%): Accuracy Values; AUC: Area under Curve Values. Standard variation of the values are in %.

| Methods	5-fold ($n = 1010$)				Independent test ($n = 245$)			
	$\geq$G2		G3		$\geq$G2		G3	
	ACC	AUC	ACC	AUC	ACC	AUC	ACC	AUC
Single-organ (Eso.)	$72.5_{\pm 0.8}$	$0.754_{\pm 0.7}$	$75.8_{\pm 1.5}$	$0.821_{\pm 1.9}$	74.3	0.793	74.3	0.803
Single-organ (Liver)	$68.9_{\pm 0.5}$	$0.677_{\pm 1.8}$	$63.5_{\pm 0.4}$	$0.672_{\pm 0.7}$	71.8	0.679	64.5	0.680
Single-organ (Spleen)	$71.0_{\pm 1.0}$	$0.681_{\pm 1.5}$	$61.0_{\pm 2.0}$	$0.660_{\pm 1.7}$	68.6	0.668	64.1	0.684
MOON‡ (Concat [6])	$73.2_{\pm 2.4}$	$0.793_{\pm 2.7}$	$74.9_{\pm 2.1}$	$0.836_{\pm 3.5}$	75.1	0.799	73.9	0.827
MOON‡ (PredSum [21])	$75.4_{\pm 3.2}$	$0.776_{\pm 3.3}$	$76.2_{\pm 3.6}$	$0.830_{\pm 3.6}$	75.5	0.805	71.4	0.823
MOON‡ (LowRank [16])	$73.5_{\pm 3.5}$	$0.771_{\pm 1.57}$	$76.1_{\pm 3.8}$	$0.824_{\pm 4.4}$	73.9	0.779	71.4	0.806
MOON‡ (FiLM [18])	$74.9_{\pm 1.4}$	$0.780_{\pm 2.6}$	$77.0_{\pm 2.1}$	$0.843_{\pm 3.1}$	75.5	0.799	73.9	0.842
MOON (Concat [6])	$75.5_{\pm 2.2}$	$\underline{0.820}_{\pm 2.1}$	$76.6_{\pm 4.8}$	$0.852_{\pm 3.1}$	**78.0**	**0.832**	**77.6**	**0.864**
MOON (PredSum [21])	$75.6_{\pm 2.8}$	$\mathbf{0.824}_{\pm 2.6}$	$\underline{78.3}_{\pm 2.6}$	$0.851_{\pm 3.1}$	76.0	0.810	74.3	0.841
MOON (LowRank [16])	$\underline{76.0}_{\pm 2.8}$	$0.813_{\pm 3.4}$	$\mathbf{79.0}_{\pm 3.7}$	$\underline{0.858}_{\pm 4.6}$	76.8	0.821	76.8	0.846
MOON (FiLM [18])	$\mathbf{77.3}_{\pm 2.1}$	$0.810_{\pm 2.6}$	$78.0_{\pm 4.8}$	$\mathbf{0.864}_{\pm 4.4}$	**78.1**	_0.823_	_76.7_	_0.848_
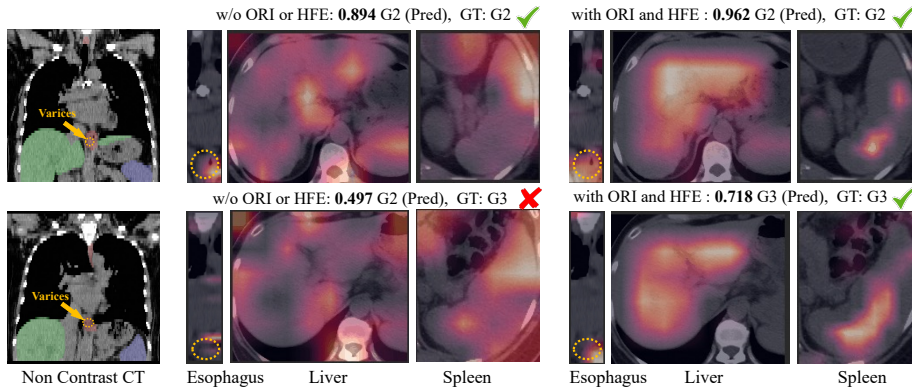
Fig. 3: Comparison of Grad-CAM visualizations across different organs.

Ablations Study. Information regarding the comparison of EV performance evaluation during 5-fold validation with the inclusion of the ORI module, HFE module, and CCA loss strategy is provided in Fig. 2. We discovered that the ORI module is more critical than the HFE module and CCA loss in the evaluation of EV, indicating that ORI facilitates comprehensive and effective interactive learning between the liver, spleen, and esophagus, capturing more valuable information pertinent to the assessment of EV. Additionally, other methods also hold significant value for the assessment of EV. Results of the performance on the independent test set are presented in the supplementary. We observed results that were similar to those in the 5-fold validation set. Fig. 3 highlights the role of the ORI and HFE modules in improving cross-organ interaction within

the model. Using Grad-CAM [19], we can pinpoint the specific areas within the model's field of focus. Correlating with endoscopic findings, we have annotated the variceal regions within the esophagus, which typically pose a detection challenge in NC-CT imaging. Without the incorporation of ORI and HFE modules for organ-specific information exchange, the model tends to overlook critical areas associated with varices, leading to potential misclassification of EV severity. In contrast, incorporating the ORI and HFE modules directs the MOON to accurately zone in on medically significant regions, and precisely targets the distended vessels in the esophagus, the hepatic portal vein, and the spleen, which are indicative of portal hypertension and EV.

4   Conclusion

We introduce the Multi-Organ-Cohesion-Network, designed for the non-invasive evaluation of EV using NC-CT scans. By simulating the diagnostic processes of radiologists who assess multiple organs implicated in EV, we fill the diagnostic gap in EV evaluation with NC-CT. Our extensive analysis confirms that NC-CT imaging effectively evaluates EV, outperforming previous DCE-CT-based methods. In our future work, we plan to incorporate adjunctive clinical data, e.g., blood test markers, to refine the classification efficacy of EV.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Andrew, G., Arora, R., Bilmes, J., Livescu, K.: Deep canonical correlation analysis. In: International Conference on Machine Learning. pp. 1247–1255 (2013)
2. Carreira, J., Noland, E., Banki-Horvath, A., Hillier, C., Zisserman, A.: A short note about kinetics-600. arXiv preprint arXiv:1808.01340 (2018)
3. Cheng, J., Wang, Z., Pollastri, G.: A neural network approach to ordinal regression. In: IEEE International Joint Conference on Neural Networks. pp. 1279–1284 (2008)
4. Dahong, D.: Trial scheme of endoscopic diagnosis and treatment of esophageal and gastric varices. Chin J Digest Endoscopy **17**, 198–199 (2000)
5. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al.: An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929 (2020)
6. Feichtenhofer, C., Pinz, A., Zisserman, A.: Convolutional two-stream network fusion for video action recognition. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 1933–1941 (2016)

7. Garcia-Tsao, G., Abraldes, J.G., Berzigotti, A., Bosch, J.: Portal hypertensive bleeding in cirrhosis: Risk stratification, diagnosis, and management: 2016 practice guidance by the american association for the study of liver diseases. Hepatology **65**(1), 310–335 (2017)

8. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 770–778 (2016)

9. Isensee, F., Jaeger, P.F., Kohl, S.A., Petersen, J., Maier-Hein, K.H.: nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. Nature Methods **18**(2), 203–211 (2021)

10. Kawano, Y., Sasaki, A., Kai, S., Endo, Y., Iwaki, K., Uchida, H., Shibata, K., Ohta, M., Kitano, S.: Short-and long-term outcomes after hepatic resection for hepatocellular carcinoma with concomitant esophageal varices in patients with cirrhosis. Annals of surgical oncology **15**, 1670–1676 (2008)

11. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)

12. LeCun, Y., Boser, B., Denker, J., Henderson, D., Howard, R., Hubbard, W., Jackel, L.: Handwritten digit recognition with a back-propagation network. Advances in Neural Information Processing Systems **2** (1989)

13. LeCun, Y., Boser, B., Denker, J.S., Henderson, D., Howard, R.E., Hubbard, W., Jackel, L.D.: Backpropagation applied to handwritten zip code recognition. Neural Computation **1**(4), 541–551 (1989)

14. Li, K., Wang, Y., Zhang, J., Gao, P., Song, G., Liu, Y., Li, H., Qiao, Y.: Uniformer: Unifying convolution and self-attention for visual recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence (2023)

15. Lin, T.Y., Dollár, P., Girshick, R., He, K., Hariharan, B., Belongie, S.: Feature pyramid networks for object detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2117–2125 (2017)

16. Liu, Z., Shen, Y., Lakshminarasimhan, V.B., Liang, P.P., Zadeh, A., Morency, L.P.: Efficient low-rank multimodal fusion with modality-specific factors. arXiv preprint arXiv:1806.00064 (2018)

17. Luo, R., Gao, J., Gan, W., Xie, W.B.: Clinical-radiomics nomogram for predicting esophagogastric variceal bleeding risk noninvasively in patients with cirrhosis. World Journal of Gastroenterology **29**(6), 1076 (2023)

18. Perez, E., Strub, F., De Vries, H., Dumoulin, V., Courville, A.: Film: Visual reasoning with a general conditioning layer. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 32 (2018)

19. Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D.: Gradcam: Visual explanations from deep networks via gradient-based localization. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 618–626 (2017)

20. Shi, Y., Qi, Y.f., Lan, G.y., Wu, Q., Ma, B., Zhang, X.y., Ji, R.y., Ma, Y.j., Hong, Y.: Three-dimensional mr elastography depicts liver inflammation, fibrosis, and portal hypertension in chronic hepatitis b or c. Radiology **301**(1), 154–162 (2021)

21. Simonyan, K., Zisserman, A.: Two-stream convolutional networks for action recognition in videos. Advances in Neural Information Processing Systems **27** (2014)

22. Tan, M., Pang, R., Le, Q.V.: Efficientdet: Scalable and efficient object detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 10781–10790 (2020)

23. Tran, D., Wang, H., Torresani, L., Ray, J., LeCun, Y., Paluri, M.: A closer look at spatiotemporal convolutions for action recognition. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 6450–6459 (2018)
24. Wan, S., He, Y., Zhang, X., Wei, Y., Song, B.: Quantitative measurements of esophageal varices using computed tomography for prediction of severe varices and the risk of bleeding: a preliminary study. Insights into Imaging **13**(1), 47 (2022)
25. Wasserthal, J., Meyer, M., Breit, H., Cyriac, J., Yang, S., Segeroth, M.: Totalsegmentator: Robust segmentation of 104 anatomical structures in ct images 2022. arXiv (2022)
26. Yan, Y., Li, Y., Fan, C., Zhang, Y., Zhang, S., Wang, Z., Huang, T., Ding, Z., Hu, K., Li, L., et al.: A novel machine learning-based radiomic model for diagnosing high bleeding risk esophageal varices in cirrhotic patients. Hepatology International **16**(2), 423–432 (2022)