

HDMF: Hypergraph-Driven Multimodal Fusion Network for COVID-19 Prognosis

Zhuoru Wu, Wennuo Yang, Meidan Ding, Shaonan Liu, Filippo Ruffini, Xuechen Li, Domenico Albano, Eliodoro Faiella, Deborah Fazzini, Domiziana Santucci, Paolo Soda, Valerio Guerrasi, Xiaoling Luo, Wenting Chen, and Linlin Shen, *Senior Member, IEEE*

Abstract—COVID-19 prognosis requires the accurate prediction of disease progression to guide clinical decisions. While multimodal fusion combining chest X-ray (CXR) and clinical tabular data shows promise, existing methods face three critical limitations: (1) lack of zone-specific severity analysis that radiologists employ, (2) implicit modeling of clinical attributes without explicitly capturing the inter-attribute relationship or adaptive weighting, and (3) shallow fusion strategies that fail to explore the deep cross-modal correlation. To address these challenges, we propose HDMF (Hypergraph-Driven Multimodal Fusion), a novel network leveraging hypergraph structures to model the high-order relationship within and across modalities. HDMF comprises three modules: a Zone-Severity Hypergraph Fusion Module (ZS-HFM) that correlates lung zones enhanced by severity scores, an Attribute-Confidence Hypergraph Fusion Module (AC-HFM) that explicitly models the clinical interrelationship with confidence-based weighting, and a Multimodal Prognosis Module (MPM) that captures the cross-modal correlation.

This work was supported in part by the National Natural Science Foundation of China under Grant 82261138629 and 12326610, Natural Science Foundation of Guangdong Province under Grant 2025A1515010184, Guangdong Provincial Key Laboratory under Grant 2023B1212060076, and the project of Shenzhen Science and Technology Innovation Committee under Grant JCYJ20240813141424032. This work was also supported in part by the Italian Ministry of Foreign Affairs and International Cooperation under Grant PGR01156, PNRR MUR project PE0000013-FAIR, and PNRR – DM 118/2023. (*Corresponding authors: Wenting Chen and Linlin Shen*.)

Wenting Chen is with the Department of Radiation Oncology, Stanford University, CA, USA (e-mail: wentchen@stanford.edu).

Zhuoru Wu, Wennuo Yang, Meidan Ding, Shaonan Liu, Xiaoling Luo, and Linlin Shen are with the College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, China. Linlin Shen is also with the AI Research Center for Medical Image Analysis and Diagnosis, Shenzhen University, Shenzhen, China, and the National Engineering Laboratory for Big Data System Computing Technology, Shenzhen University, Shenzhen, China (e-mail: llshen@szu.edu.cn).

Filippo Ruffini, Paolo Soda, and Valerio Guerrasi are with the Unit of Artificial Intelligence and Computer Systems, Department of Engineering, Università Campus Bio-Medico di Roma, Roma, Italy.

Xuechen Li is with the School of Electronics and Information Engineering, Wuyi University, Jiangmen, Guangdong, China.

Domenico Albano is with the Department of Radiology, ASST Grande Ospedale Metropolitano Niguarda, Milan, Italy, and with the Dipartimento di Scienze Biomediche, Chirurgiche ed Odontoiatriche, Università degli Studi di Milano, Milan, Italy.

Eliodoro Faiella and Domiziana Santucci are with the Unit of Radiology and Interventional Radiology, Fondazione Policlinico Universitario Campus Bio-Medico, Rome, Italy, and with the Research Unit of Radiology, Department of Medicine and Surgery, Università Campus Bio-Medico di Roma, Rome, Italy.

Deborah Fazzini is with the Department of Diagnostic Imaging and Stereotactic Radiosurgery, Centro Diagnostico Italiano S.p.A., Italy.

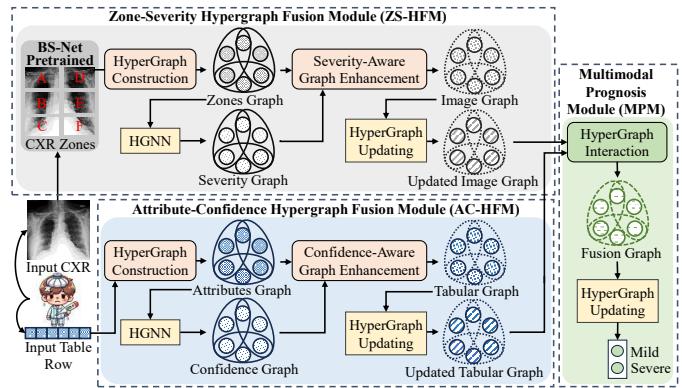


Fig. 1. An illustration of our HDMF network for COVID-19 prognosis. We capture intra-modal prognostic information by enriching the hypergraph representation modeling and further assist in final classification by mining the overall inter-modal correlation through hypergraph interaction.

through hypergraph matching. We implement hypergraph updating using Hypergraph-adapted Mamba and Hypergraph Neural Networks to capture complementary local and global perspectives. Extensive experiments on two public COVID-19 datasets demonstrate HDMF's superior performance over existing methods. Comparison with radiologists' assessments confirms HDMF closely aligns with clinical practice while maintaining high interpretability, validating its potential for real-world prognostic applications. Source code is to be released.

Index Terms—COVID-19 Prognosis, Multimodal Fusion, Hypergraph

I. INTRODUCTION

The ongoing threat of COVID-19, including emerging variants and immune evasion, underscores the need for accurate prognostic tools to predict disease progression and guide clinical decisions. COVID-19 prognosis is a binary classification problem distinguishing mild from severe patients. Traditional methods relying on unimodal data, such as Chest X-ray (CXR) [6], [7] or clinical indicators [8], [9], show promise but may not fully capture the complexity of patient conditions. Multimodal fusion, combining complementary information from different modalities [10], has emerged as a promising approach to improve predictive accuracy. As shown in TABLE I, many studies have applied multimodal data (CXR and clinical table) for COVID-19 classification, using CNNs for

TABLE I
EXISTING MULTIMODAL CLASSIFICATION METHODS IN COVID-19.

Existing Method	Mission Target	Image Encoder	Lung-Zone Modeling	Table Encoder	Attribute-Relationship Modeling	Fusion Mode	Inter-Modal Interaction
Soda <i>et al.</i> (2021) [1]	COVID-19 Prognosis	ResNet-50, VGG-11, GoogleNet	✗	MLP	✗	Intermediate Fusion	✗
Wu <i>et al.</i> (2023) [2]	Risk Prediction	ResNet-50, DenseNet-121, EfficientNet-B2	✗	MLP	✗	Intermediate Fusion	✗
Guarrasi <i>et al.</i> (2023) [3]	COVID-19 Prognosis	ResNeXt-50, VGG13-BN, GoogleNet	✗	MLP	✗	Later Fusion	✗
Wu <i>et al.</i> (2024) [4]	Mortality Prediction	ResNet-50+LSTM	✗	MLP	✗	Intermediate Fusion	✗
Guarrasi <i>et al.</i> (2024) [5]	COVID-19 Prognosis	ResNet-101	✗	MLP	✗	Intermediate Fusion	✗
HDMF	COVID-19 Prognosis	ZS-HFM	✓	AC-HFM	✓	MPM	✓

image features and MLPs for tabular features, with integration via intermediate or later fusion. While effective, these methods still face challenges in modeling different modalities.

The first challenge is that current methods mainly learn global features of CXR, lacking severity analysis for local lung zones. In clinical practice, radiologists often focus on abnormal zones based on their clinical expertise. They summarize the severity of these zones to conduct the prognosis, as analyzing pathological conditions across zones provides crucial insights into overall prognosis [11]. Moreover, Signoroni *et al.* [12] propose BS-Net that uses structured severity scores called brixia scores to assess COVID-19 severity across six lung zones, emphasizing zone-specific modeling in CXR. These severity scores support zone-specific analysis and show significant statistical correlations with treatment outcomes, validating their prognostic value [13]. However, present approaches typically extract CXR features by CNNs like pretrained ImageNet [14] models in a global or uniform manner, while ignoring the local analysis of lung zones. This lack of zone-specific modeling may prevent the model from fully capturing localized information crucial for prognosis. Therefore, it is necessary to leverage the lung-zone severity scores as prior pathological knowledge to guide prognosis.

Although existing methods can leverage clinical tabular data for prognosis, they ignore the explicit interrelationship among attributes. In a clinical table, attributes are not isolated data points but collectively reflect the patient's prognosis, indicating meaningful interrelationship among these clinical indicators [15]. For instance, D-dimer has a strong interrelationship with C-Reactive Protein (CRP), White Blood Cells (WBC), and Lactate Dehydrogenase (LDH), whose collectively elevated levels signify robust inflammation and compromised coagulation, representing the grave prognosis [16]. However, current methods [17] primarily leverage MLP for implicitly learning feature interactions, lacking explicit modeling of clinical interrelationship and adaptive attributes weighting. Without architectural inductive biases for attribute interactions, MLP cannot readily incorporate domain knowledge about which interrelationships are prognostically significant. Furthermore, MLP assigns equal confidence to all attributes, potentially allowing noisy features to dilute critical clinical signals. Consequently, explicitly modeling attributes' interrelationship and adaptively weighting their contributions is essential for improved prognostic prediction.

The third challenge is the limited deep interaction between modalities during multimodal fusion. In prognostic diagnosis, radiologists typically estimate prognosis integrating both imaging findings (e.g., abnormal lung zones) and clinical indicators (e.g., age, medical history), which jointly provide a comprehensive reflection of the patient's prognosis. A simultaneous decrease in SaO_2 and PaO_2 values in tabluar data along with the presence of diffuse infiltrates in CXR, is highly likely to occur in a patient with severe prognosis [18]. However, current multimodal fusion methods directly concatenate the image and tabular features for fusion or minimize the distance between them [10], [19], [20]. As a result, they fail to fully explore the deep cross-modal correlation between CXR and tabular data, hindering their ability to accurately simulate radiologists' decision-making for prognosis. Thus, it is essential to explore critical cross-modal correlation between CXR and tabular data to imitate clinical decision-making.

To address these challenges, we propose a Hypergraph-Driven Multimodal Fusion (HDMF) network for COVID-19 prognosis. In Fig. 1, HDMF comprises three modules: a *Zone-Severity Hypergraph Fusion Module* (ZS-HFM) for CXR zones analysis, an *Attribute-Confidence Hypergraph Fusion Module* (AC-HFM) for tabular data modeling, and a *Multimodal Prognosis Module* (MPM) for effective CXR-table fusion. To capture the critical severity variations across CXR zones, ZS-HFM uses hypergraph structures to establish high-order correlation among zones. We enhance the representation of zones with a severity-aware graph enhancement that integrates estimated severity scores directly into the zones graph, highlighting areas of particular concern. To exploit the important *inter-attribute relationship*, AC-HFM employs hypergraph modeling to construct clinical relevance among attributes. We prioritize the most relevant attributes and their interrelationship by a confidence-aware graph enhancement that assigns confidence weights to each attribute, strengthening the attributes graph where clinical significance ranks highly. To consider the latent deep *cross-modal correlation*, MPM features a HyperGraph Interaction (HGI) that performs hypergraph matching between image and tabular graphs, to emulate the clinical decision-making process with both visual and tabular information. This captures high-order correlation between zones and attributes, generating a fusion graph that mirrors clinical reasoning patterns and improves prognosis accuracy. Besides, the HyperGraph Updating in HDMF is implemented by specially designed Hypergraph-adapted Mamba (HG-Mamba) together with Hypergraph Neural Network (HGNN) [21]. It aims to analyze hypergraphs from complementary local and global perspectives, respectively, enhancing the modeling capability. Extensive evaluations on public datasets demonstrate the effectiveness and superiority of HDMF over existing methods. Notably, compared with radiologists' survey results, HDMF is closely aligned with clinical practice while offering high interpretability. Our contributions can be summarized as follows:

- We introduce a novel hypergraph-driven multimodal fusion network for COVID-19 prognosis, consisting of a zone-severity hypergraph fusion module (ZS-HFM), an attribute-confidence hypergraph fusion module (AC-HFM), and a multimodal prognosis module (MPM).
- We design a ZS-HFM to leverage the lung-zone severity as guidance, and develop an AC-HFM to highlight the important tabular attributes and their interrelationship related to prognosis.
- To exploit the latent high-order cross-modal correlation, we propose a HyperGraph Interaction within MPM, promoting hypergraph matching across modalities to imitate radiologists' decision-making behaviors.
- **HDMF** achieves the best performance on two public multimodal COVID-19 prognosis datasets in contrast to existing methods and aligns well with clinical practice compared to radiologists' assessment, showing its strong interpretability.

II. RELATED WORK

A. Multimodal COVID-19 Prognosis

Multimodal learning has garnered increasing attention in COVID-19 prognosis. Many scholars incorporate both CXR and clinical tabular data to improve the performance of predictive classification tasks [1]–[5]. For instance, Guerrasi *et al.* [5] apply joint learning of CXR and tabular data for modality reconstruction and sample classification in COVID-19 prognosis, and propose an interpretable method that reveals the most contributing features of each modality to the decision-making process. Althenayan *et al.* [22] introduce a hierarchical classification framework, which processes CXR by ResNet and then integrates CXR and tabular data to not only identify COVID-19 but also to distinguish it from other types of pneumonia. Sangeetha *et al.* [23] extract salient features from each modality (medical imaging, genomic, and clinical tabular data) and concatenate them to create a unified high-dimensional feature vector, thereby enhancing the diagnostic accuracy of lung cancer classification. However, such the modality reconstruction or extraction as a general approach may be unable to explore the task-relevant characteristics of each modality. Given the connection between CXR zones or tabular attributes and COVID-19 prognosis [11], [15], it is necessary to introduce the targeted intra-modal modeling that divides lung zones for CXR modeling and utilizes attributes' interrelationship for table modeling.

B. Multimodal Fusion Approaches

Effective multimodal fusion has emerged as an active topic of interest in current multimodal learning research. Various modes are developed to effectively handle the data heterogeneity issue in multimodal learning [24]. Simple multimodal fusion modes include early fusion (EF), later fusion (LF), and intermediate fusion (IF), which are typically categorized based on the level of data abstraction [10]. EF combines raw multimodal inputs into a unified representation before feature extraction [25]. Its requirement for consistent data formats makes it unsuitable for heterogeneous modalities like

image and tabular data. LF integrates final decision outputs from independent unimodal models [26], and trusted multi-view classification (TMVC) is based on the Dempster-Shafer theory that employs the variational Dirichlet distribution to represent the class probabilities [27]. IF merges high-level features extracted by modality-specific encoders [28], which is commonly used in multimodal COVID-19 classification, illustrated in TABLE I. However, these modes only perform superficial fusion at the high-dimensional feature level, ignoring the correlation across modalities. Beyond these, other fusion modes, while acknowledging the need to leverage cross-modal information, merely focus on aligning features across different modalities to form a unified feature representation, lacking interpretability. For example, deep canonical correlation analysis (DCCA) maximizes modality-specific features correlation to learn a joint embedding space, enhancing the complementary information [19]. Furthermore, deep canonically correlated autoencoder (DCCAE) is introduced not only to maximize inter-modal correlation to learn the shared representation but also to reconstruct each modality by the autoencoder, preserving important intra-modal information while aligning multimodal features [20]. Therefore, we establish a novel multimodal fusion mode that separately constructs the CXR and the table into hypergraphs, and enables the hypergraph matching between them to perform the explainable inter-modal interaction.

III. METHOD

A. Overview

In Fig. 2, **HDMF** comprises three distinct modules: the zone-severity hypergraph fusion module (ZS-HFM), the attribute-confidence hypergraph fusion module (AC-HFM), and the multimodal prognosis module (MPM). CXR and table row are input into ZS-HFM and AC-HFM, respectively. In ZS-HFM, CXR zones are encoded by the pretrained lung-zone severity estimator and formed into a zones hypergraph G_Z , which HGNN processes to generate a severity hypergraph G_S . Both G_Z and G_S are used in Severity-Aware Graph Enhancement to create the image hypergraph G_I , which is updated through HGNN and HG-Mamba to produce G_I^U . The severity scores from G_S are used to calculate the CE loss \mathcal{L}_{SEV} with the brixia scores from the pretrained lung-zone severity estimator, ensuring targeted modeling of lung zones. In AC-HFM, we construct an attributes hypergraph G_A and derive a confidence hypergraph G_C . Both are used in Confidence-Aware Graph Enhancement to form the tabular hypergraph G_T , which is updated through HGNN and HG-Mamba to yield G_T^U . The confidence scores from G_C are used to calculate the MSE loss \mathcal{L}_{CGE} , enabling precise modeling of attributes' interrelationship. Finally, MPM takes G_I^U and G_T^U as inputs, matches them through HyperGraph Interaction to form the fusion hypergraph G_F , and reduces redundant correlation across modalities via the sparsification entropy loss \mathcal{L}_{SPA} . The final prognosis outcome is obtained by updating G_F using HGNN and HG-Mamba, followed by calculating the CE loss \mathcal{L}_{CLS} with ground truth.

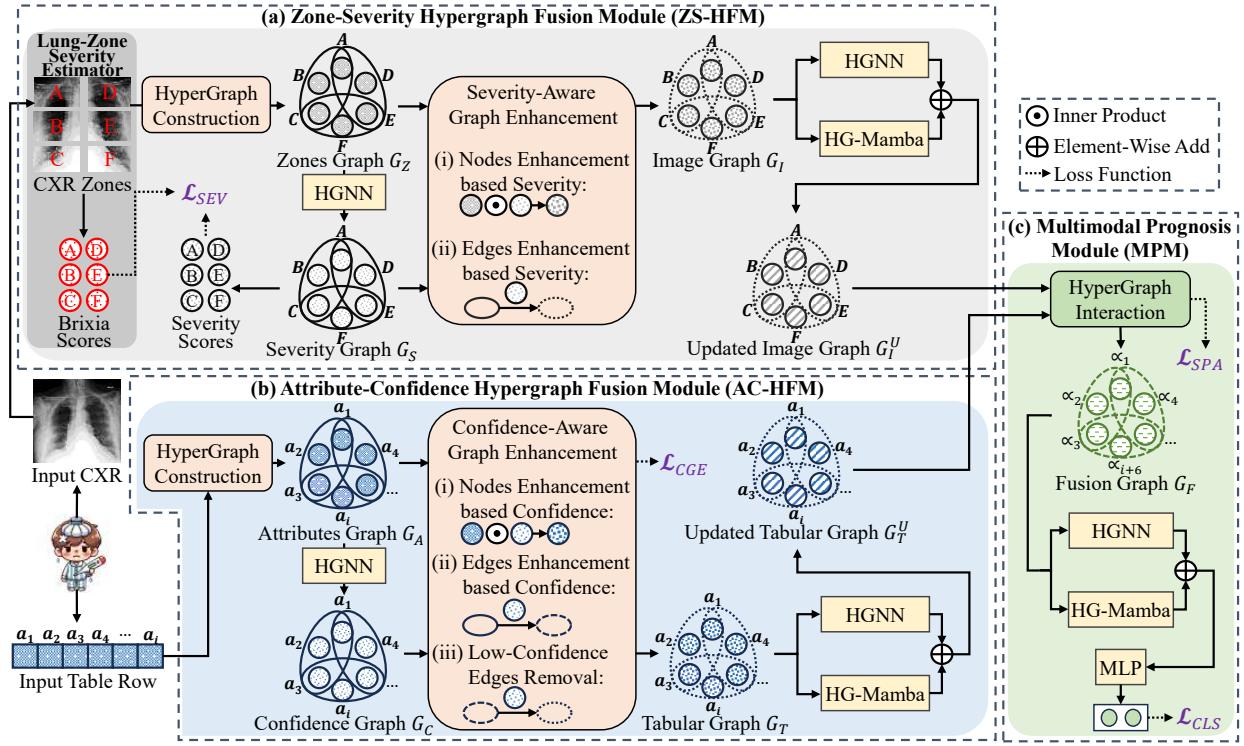


Fig. 2. Overview of our proposed HDMF network for COVID-19 prognosis, which comprises three primary modules: the Zone-Severity Hypergraph Fusion Module (ZS-HFM) for modeling CXR highlighting abnormal zones, the Attribute-Confidence Hypergraph Fusion Module (AC-HFM) for modeling table exploiting essential attributes and their interrelationship, and the Multimodal Prognosis Module (MPM) for hypergraph interaction between modalities and accurate disease prognosis.

B. Zone-Severity Hypergraph Fusion Module (ZS-HFM)

Current methods [1]–[5] mainly conduct feature extraction on the entire CXR, ignoring the lung-zone analysis. To address this issue, we introduce ZS-HFM to highlight abnormal lung zones relevant to prognosis. Specifically, we leverage a pre-trained lung-zone severity estimator [12] to encode CXR into 6 zones (A to F), resulting in zones features $Z^f \in \mathbb{R}^{6 \times 256 \times 512}$ and brixia scores $B^s \in \mathbb{R}^{6 \times 1}$. Brixia scores range from 0 to 3, representing four severity levels, with higher scores indicating increased severity.

1) HyperGraph Construction: To construct the zones hyperedges H_Z , we calculate the Euclidean distances between zones based on their features Z^f and connect each zone to its $K - 1$ nearest neighbors [21]. Subsequently, we update $G_Z = (Z^f, H_Z)$ through HGNN to acquire the severity hypergraph $G_S = (S^f, H_Z) = (HGNN(G_Z), H_Z)$, where $S^f \in \mathbb{R}^{6 \times 4}$ denotes the predictive probability of severity scores S^s . Severity scores S^s quantify the severity of lung zones like brixia scores B^s in line with clinical reality [29]. Then, the CE loss \mathcal{L}_{SEV} is calculated between S^f and brixia scores B^s to ensure that targeted modeling of lung zones reflects the correct pathology: $\mathcal{L}_{SEV} = -\sum_{i=1}^4 B_i^s \log(S_i^f)$.

2) Severity-Aware Graph Enhancement: Aiming to highlight abnormal lung zones, we apply a severity-aware graph enhancement using severity hypergraph G_S on zones hypergraph G_Z to form image hypergraph $G_I = (I^f, H_I)$: (i) Nodes Enhancement based Severity: $I^f = \hat{S}^s \odot Z^f$, where $\hat{S}^s = \text{MinMax}(S^s)$, $S^s = \arg \max(S^f, \text{dim} = 1)$. (ii) Edges

Enhancement based Severity:

$$H_I = H_Z + \lambda_I \cdot \frac{1}{|V_{H_Z}|} \sum_{i \in V_{H_Z}} S_i^s, \quad (1)$$

where V_{H_Z} is the set of nodes contained in H_Z and λ_I is an adjustable hyperparameter. Next, G_I is updated by HGNN and HG-Mamba to acquire G_I^U .

3) Hypergraph-adapted Mamba (HG-Mamba): As Mamba excels over CNNs in capturing long-range global dependencies [30], many works have used it in medical tasks [31]–[35]. But when applied to hypergraphs, its unidirectionality in recurrent scan and update may result in limited information that important nodes can access. To help important hypernodes access more contextual information, we propose Hypergraph-adapted Mamba (HG-Mamba). Here, important hypernodes are defined with high trait (severity/confidence) and degree. As depicted in Fig. 3, a hypergraph is flattened and nodes are first permuted based on their trait. As for nodes with the same trait, we continue to sort them based on their degree. Then, if nodes with the same degree appear, we swap them randomly to avoid the network's preference for a specific ordering. Important hypernodes are currently placed later in the sequence, ensuring that they can receive longer-range contextual information.

HG-Mamba processes CXR zones conducting global sequence modeling, while HGNN captures interrelationship between them performing local topology modeling. The combination of them improves the network's modeling capability, better investigating intra-modality information.

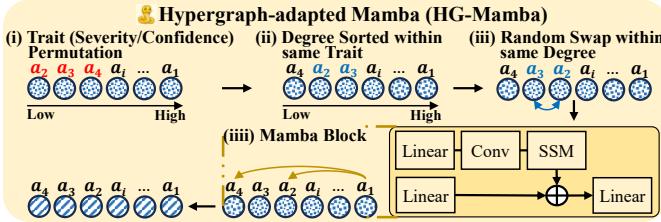


Fig. 3. Hypergraph-adapted Mamba (HG-Mamba). Hypernodes with red and blue color names have the same trait and degree.

C. Attribute-Confidence Hypergraph Fusion Module (AC-HFM)

As existing methods [1]–[5] lack sufficient modeling for attributes’ interrelationship in clinical tabular data, we devise AC-HFM to emphasize significant tabular attributes and their interrelationship pertinent to prognosis.

1) *HyperGraph Construction*: A raw table row with i attributes is represented by $T_r = (a_1, \dots, a_i)$. All attributes features are denoted as $A^f = (A_1^f, \dots, A_i^f) \in \mathbb{R}^i$. Similarly to III-B.1, we construct attribute hyperedges H_A using A^f to form attributes hypergraph G_A , which captures the complex underlying clinical interrelationship between attributes.

Then, we update G_A through HGNN to obtain the confidence hypergraph $G_C = (C^f, H_A)$, where C^f represents the confidence scores for corresponding attributes. Inspired by pruning mechanisms [36], [37], confidence scores are employed to reflect the significance of attributes for prognosis.

2) *Confidence-Aware Graph Enhancement*: To emphasize important attributes and their interrelationship, we enhance the attributes hypergraph G_A by the confidence hypergraph G_C , resulting in the tabular hypergraph $G_T = (T^f, H_T)$: (i) Nodes Enhancement based Confidence: $T^f = C^f \odot A^f$. (ii) Edges Enhancement based Confidence:

$$H_T = H_A + \lambda_T \cdot \frac{1}{|V_{H_A}|} \sum_{i \in V_{H_A}} C_i^f, \quad (2)$$

where λ_T is an adjustable hyperparameter. (iii) Low-Confidence Edges Removal: Since hyperedges are formed by centering each node in HyperGraph Construction [21], a hypergraph with i nodes contains i hyperedges. As the number of attribute nodes increases, the hyperedge structure may become redundant. Consequently, we propose a confidence-aware gated perception to avoid redundancy caused by low-confidence attribute hyperedges. For training, we apply Gumble-softmax [38] to enable the differentiation of discrete decisions on confidence scores, which allows the dynamic selection of attribute hyperedges through gating, expressed as:

$$M_l = \frac{\exp\left(\frac{\log(C_l^f) + g_l}{\tau}\right)}{\sum_{j=1}^i \exp\left(\frac{\log(C_j^f) + g_j}{\tau}\right)} \in \{0, 1\}, \quad (3)$$

$$\tilde{H}_T = H_T \odot M = H_T \odot [M_1, \dots, M_l, \dots, M_i], \quad (4)$$

where M_l is the hyperedge selector mask for the l -th attribute, C_l^f is its confidence score, and g_l is the hyperedge Gumble noise. τ is a temperature parameter to control randomness

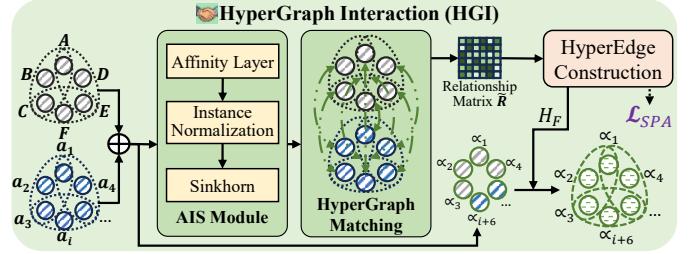


Fig. 4. The implementation of HyperGraph Interaction (HGI). Green dashed lines show the HyperGraph Matching. Darker color in \hat{R} represents the stronger correlation between hypernodes.

and follows an annealing schedule for more stable and deterministic selection: $\tau = \tau_0 \cdot \theta^{epoch}$, where τ_0 is the initial temperature set to 1, θ is the decay rate chosen as 0.95, and $epoch$ is the current training iteration. For inference, hyperedges are selected by applying argmax to the confidence scores. To regulate the selection process, we minimize the MSE loss \mathcal{L}_{CGE} between the actual and target selection rates: $\mathcal{L}_{CGE} = \left(\sum_{l=1}^i M_l / i - t \right)^2$, where t is the adjustable target selection rate. Afterwards, $G_T = (T^f, \tilde{H}_T)$ is updated by HGNN and HG-Mamba to obtain G_T^U .

D. Multimodal Prognosis Module (MPM)

Although many effective fusion modes [19], [20], [26]–[28] have been developed, the interaction between modalities is still not fully investigated. Thus, we propose MPM enabling cross-modal interaction through HyperGraph Interaction (HGI) to imitate radiologists’ decision-making behaviors for more accurate prognosis.

1) *HyperGraph Matching*: As demonstrated in Fig. 4, we facilitate hypergraph matching between G_I^U and G_T^U through the AIS module [39]. Specifically, we first concatenate G_I^U with G_T^U to form F^f and compute the affinity matrix $\mathcal{A} = (F^f)^T \mathbf{W} F^f$, where \mathbf{W} is a learnable parameter in an affinity layer. Then, \mathcal{A} is processed through instance normalization [40] to ensure positive and finite values, and then passed through Sinkhorn [41] to obtain the relationship matrix \hat{R} with all elements in the range of 0 to 1. Higher values in \hat{R} indicate stronger correlation between lung zones and tabular attributes.

2) *HyperEdge Construction*: Next, to capture high-order correlation across modalities, we convert \hat{R} into the fusion hyperedges matrix H_F connecting each node to its $K-1$ most relevant neighbor nodes based on relationship weights. The sparsification entropy loss \mathcal{L}_{SPA} is applied to enforce sparsity in H_F and focus on essential node relationships while reducing redundant connections: $\mathcal{L}_{SPA} = \sum H_F \log(H_F)$.

With the well-constructed fusion hyperedges H_F , fusion hypergraph G_F is composed as: $G_F = (F^f, H_F)$. Then, G_F is updated via HGNN and HG-Mamba to calculate the CE loss \mathcal{L}_{CLS} with ground truth: $\mathcal{L}_{CLS} = -\sum_{i=1}^2 Y_i \log(P_i)$, where P_i represents the predictive probability distribution and Y_i represents the one-hot prognosis classification label. The final loss combines multiple objectives:

$$\mathcal{L} = \omega_1 \mathcal{L}_{CLS} + \omega_2 \mathcal{L}_{SEV} + \omega_3 \mathcal{L}_{CGE} + \omega_4 \mathcal{L}_{SPA}, \quad (5)$$

TABLE II

COMPARISON OF 5-FOLD CV METRICS BETWEEN SOTA METHODS AND HDMF NETWORK. BEST RESULTS ARE HIGHLIGHTED IN **BOLD**, AND SECOND-BEST RESULTS ARE UNDERLINED. (UNIT: %)

Modality CXR Table	Method	AIforCOVID						CoCross					
		ACC	bACC	AUC	SPEC	SENS	F1-score	ACC	bACC	AUC	SPEC	SENS	F1-score
✓	ResNet-50	66.5±2.8	66.4±2.8	72.2±1.6	69.1±10.1	63.6±9.9	63.9±4.6	76.1±1.7	72.0±2.5	81.2±3.2	58.9±6.2	85.0±2.7	82.5±1.3
	DenseNet-121	67.0±1.7	67.0±1.5	71.9±0.8	66.9±5.4	67.1±4.0	65.7±1.5	77.0±1.2	71.7±1.7	82.3±1.8	54.6±9.0	88.7±6.1	83.5±1.6
	CheXpert [42]	67.7±0.7	67.7±0.7	73.2±1.2	68.9±1.9	66.4±1.5	66.0±0.7	77.8±3.4	74.6±6.3	82.5±2.4	64.3±15.6	84.8±3.4	83.4±1.6
✓	MLP	69.0±2.5	68.8±2.5	74.9±1.6	71.1±2.7	66.5±2.9	66.8±2.7	78.9±1.5	76.7±1.6	84.5±0.8	69.7±3.0	83.6±1.8	84.0±1.2
	GCN	69.7±1.9	69.5±1.6	76.2±0.7	72.5±6.5	66.6±3.8	67.4±0.9	78.9±2.1	76.3±2.0	85.7±0.8	68.1±3.5	84.4±3.0	84.1±1.7
✓	Soda <i>et al.</i> [1]	70.4±1.6	70.1±1.7	78.3±0.7	75.6±1.8	64.6±4.8	67.2±2.8	80.6±2.1	76.2±2.0	87.6±1.3	62.2±3.8	90.1±3.3	85.9±1.6
	Wu <i>et al.</i> [2]	70.2±1.3	70.0±1.4	76.7±0.9	72.8±2.5	67.2±4.9	68.0±2.5	81.1±1.5	77.1±2.4	87.1±0.9	64.3±8.8	89.9±5.3	86.2±1.5
	Guarrasi <i>et al.</i> [3]	70.1±1.6	69.9±1.6	78.4±1.6	73.7±5.4	66.1±5.5	67.4±2.2	80.0±2.4	74.7±3.1	87.4±2.8	58.4±5.9	91.1±2.1	85.8±1.6
	Wu <i>et al.</i> [4]	71.3±1.0	71.1±1.0	78.5±0.9	73.3±2.6	68.9±2.2	<u>69.2±1.1</u>	79.4±1.4	75.3±3.7	86.2±1.4	62.2±13.6	88.5±6.7	84.9±1.3
	Guarrasi <i>et al.</i> [5]	69.8±1.8	69.6±1.9	75.5±1.2	72.3±1.9	66.9±4.0	67.5±2.5	78.9±1.9	75.4±1.5	86.5±1.7	64.3±8.0	86.5±6.4	84.3±2.2
	LF [26]	70.3±0.9	70.0±0.9	78.0±0.6	74.7±4.9	65.4±5.7	67.5±2.3	80.2±2.5	73.7±3.5	86.5±3.3	53.5±7.5	93.9±2.3	86.2±1.6
✓	TMVC [27]	71.8±1.1	<u>71.5±1.2</u>	78.4±1.1	75.8±1.0	67.2±1.9	69.2±1.4	82.2±3.0	79.0±2.7	90.9±1.8	69.2±9.1	88.9±7.4	86.7±2.8
	IF [28]	70.9±1.7	70.6±1.8	78.3±0.7	<u>76.0±3.1</u>	65.1±4.0	67.8±2.4	79.6±1.5	74.8±1.7	86.2±1.8	59.5±3.8	90.1±2.2	85.3±1.1
	DCCA [19]	71.2±2.1	71.0±1.8	<u>78.7±1.2</u>	74.6±8.8	67.3±7.2	68.6±2.1	81.3±2.3	76.8±2.8	88.3±2.2	62.7±5.9	91.0±2.9	86.5±1.7
	DCCAE [20]	71.4±1.4	71.1±1.4	78.0±1.1	75.9±1.3	66.3±2.3	68.6±1.7	80.9±4.6	75.1±6.3	90.5±2.0	56.8±12.1	93.5±2.1	86.6±2.9
✓	HDMF	74.7±0.8	<u>74.3±0.7</u>	81.1±0.6	81.3±4.1	<u>67.3±3.6</u>	71.4±1.0	88.4±1.9	85.9±3.2	94.9±0.6	77.8±7.3	93.9±1.2	91.5±1.2

where ω_1 , ω_2 , ω_3 , and ω_4 are weighting parameters assigned with different values due to diverse data distributions.

IV. EXPERIMENTS AND RESULTS

A. Experimental Settings

1) *Datasets*: We evaluate **HDMF** on two public multimodal COVID-19 prognosis datasets, **AIforCOVID** [1] and **CoCross** [43], summarized in TABLE III. They are collected from distinct patients of different hospitals, resulting in varying data distributions. We apply unified data preprocessing operations to process them. CXR images are resized to **512 × 512** to meet the input request of the pretrained lung-zone severity estimator. Missing values in tabular data are imputed using Mean-Mode [44], and a min-max scaler is applied to normalize attributes into a consistent range [0, 1].

2) *Implementation Details*: Uniform training configurations are applied with PyTorch on one NVIDIA V100 32GB GPU. During **HDMF** training, the batch size is set to 16. Adam optimizer is used with an initial learning rate of 1e-3, momentum of 0.9, and weight decay of 1e-4. The training process is limited to a maximum of 300 epochs with early stopping applied after 25 epochs based on the validation loss, and the optimal model is selected at the epoch with the lowest validation loss. In all the hypergraph or hyperedge constructions, K is fixedly set to 15.

3) *Evaluation Metrics*: We set up a separate test set and adopt the 5-fold stratified cross-validation (CV) evaluation. The specific data-splitting ratios are illustrated in TABLE III. We evaluate the accuracy (ACC), balanced accuracy (bACC), area under the receiver operating characteristic curve (AUC), specificity (SPEC), sensitivity (SENS) and F1-score to assess the prognostic performance. All experimental results are presented as mean ± sd (standard deviation).

B. Comparison with Previous Methods

We compare **HDMF** with previous state-of-the-art (SOTA) methods, including unimodal methods, multimodal methods in COVID-19, and other multimodal fusion mode methods. All SOTA methods are retrained on two benchmark datasets for fair comparison. For the unimodal methods, we use ResNet-50, DenseNet-121, and CheXpert for the image modality, and

TABLE III
SUMMARY OF BENCHMARK DATASETS.

Benchmark dataset	AIforCOVID	CoCross
Sample size	1578	543
Number of tabular attributes	34	55
Category ratio (Mild:Severe)	835:743	185:358
Data-splitting ratio (Train:Val:Test)	1009:253:316	346:88:109

MLP and GCN for the tabular modality. Both ResNet-50 and DenseNet-121 are pretrained on ImageNet [14]. CheXpert [42] has been pretrained on a large number of CXR images with DenseNet-121 as its backbone. When GCN processes the clinical table, the adjacency matrix of tabular attributes is generated by random initialization. For the multimodal methods in COVID-19 [1]–[5], the comparison retains only their image encoder, table encoder, and fusion mode listed in TABLE I due to different details in technical realization. For the other multimodal fusion mode methods, we conduct the comparison with later fusion (LF) [26], trusted multi-view classification (TMVC) [27], intermediate fusion (IF) [28], deep canonical correlation analysis (DCCA) [19], and deep canonically correlated autoencoder (DCCAE) [20]. Since the above multimodal methods involve various combinations of backbone, only the greatest results are reported.

As shown in TABLE II, **HDMF** outperforms best-performing SOTA methods in most metrics across all two datasets. On AIforCOVID, **HDMF** improves bACC from 71.5% to 74.3% (+2.8%) and F1-score from 69.2% to 71.4% (+2.2%). Also, **HDMF** achieves a favorable clinical trade-off: while SENS (67.3%) is lower than the best competing method (68.9%), it secures notable gains in AUC from 78.7% to 81.1% (+2.4%) and SPEC from 76.0% to 81.3% (+5.3%), effectively reducing false-positive predictions while maintaining strong overall discriminative ability. On CoCross, **HDMF** attains the highest scores across all metrics, with its strong bACC of 85.9% (+6.9%), AUC of 94.9% (+4.0%), and F1-score of 91.5% (+4.8%), underscoring robust performance under class imbalance and small-sample setting, enabling stable adaptation to varying data distributions. Compared with unimodal methods, **HDMF** fully exploits multimodal complementarity and consistently delivers substantial performance gains. Besides, existing multimodal methods often exhibit unstable metric behaviors, such as sensitivity–specificity imbalance or large

TABLE IV

EFFECTIVENESS OF EACH MODULE IN OUR HDMF NETWORK.

Each module			AlforCOVID					
ZS-HFM	AC-HFM	MPM	ACC	bACC	AUC	SPEC	SENS	F1-score
✗	✓	✓	70.8±1.0	70.4±1.4	77.2±1.8	77.6±7.5	63.1±9.5	66.7±4.1
✓	✗	✓	70.9±1.1	70.5±0.9	76.1±0.9	77.1±5.0	63.9±3.7	67.4±0.9
✓	✓	✗	70.8±1.4	70.5±1.8	78.0±0.8	74.4±5.2	66.7±8.5	68.1±3.7
✓	✓	✓	74.7±0.8	74.3±0.7	81.1±0.6	81.3±4.1	67.3±3.6	71.4±1.0
Each module			CoCross					
ZS-HFM	AC-HFM	MPM	ACC	bACC	AUC	SPEC	SENS	F1-score
✗	✓	✓	84.0±3.0	81.3±2.4	88.6±2.2	73.0±7.4	89.7±6.8	88.0±2.7
✓	✗	✓	83.7±1.5	80.7±0.8	90.5±2.0	71.4±4.1	90.0±4.1	87.9±1.5
✓	✓	✗	84.6±2.5	80.6±3.4	89.0±1.4	68.1±7.5	93.1±3.1	88.9±1.8
✓	✓	✓	88.4±1.9	85.9±3.2	94.9±0.6	77.8±7.3	93.9±1.2	91.5±1.2

TABLE V

EFFECTIVENESS OF PRETRAINED LUNG-ZONE SEVERITY ESTIMATOR.

AlforCOVID								
Lung-zone severity estimator	ACC	bACC	AUC	SPEC	SENS	F1-score		
ImageNet [14]	71.3±1.0	70.7±1.1	78.4±0.9	81.0±5.3	60.4±6.1	66.3±2.6		
CheXpert [42]	72.8±1.0	72.5±1.1	79.4±1.0	77.6±5.9	67.3±6.7	69.7±2.3		
BS-Net [12]	74.7±0.8	74.3±0.7	81.1±0.6	81.3±4.1	67.3±3.6	71.4±1.0		
Lung-zone severity estimator	ACC	bACC	AUC	SPEC	SENS	F1-score		
ImageNet [14]	82.8±2.8	78.5±2.6	91.1±1.3	65.4±8.0	91.7±6.3	87.5±2.4		
CheXpert [42]	84.0±1.4	82.0±1.8	91.9±1.5	75.7±3.8	88.3±1.6	88.0±1.0		
BS-Net [12]	88.4±1.9	85.9±3.2	94.9±0.6	77.8±7.3	93.9±1.2	91.5±1.2		

performance variance, revealing their instability in bridging the heterogeneous gap between imaging and clinical features. In contrast, HDMF effectively addresses cross-modal heterogeneity through unified hypergraph-based interaction learning, resulting in superior performance and reduced variance.

C. Ablation Studies

To ensure the validity of our HDMF network design, we conduct ablation studies to systematically assess the contribution of each individual component.

1) *Effectiveness of each module*: As depicted in TABLE IV, we ablate each module in HDMF, i.e. Zone-Severity Hypergraph Fusion Module (ZS-HFM), Attribute-Confidence Hypergraph Fusion Module (AC-HFM), and Multimodal Prognosis Module (MPM), to individually assess its contribution. Specifically, disabling ZS-HFM indicates employing conventional ResNet-50 for CXR encoding. Likewise, removing AC-HFM corresponds to using the standard MLP for tabular data encoding. When MPM is excluded, HDMF simply concatenates hypernode features from the two modalities for prognosis. The consistent performance degradation on both datasets (i.e. deactivating ZS-HFM causes bACC decreases by 3.9%, AUC decreases by 3.9%, and F1-score decreases by 4.7% on AlforCOVID, and bACC drops by 4.6%, AUC drops by 6.3%, and F1-score drops by 3.5% on CoCross) observed in all these ablated variants confirms the necessity of each module. Both targeted intra-modal modeling modules (ZS-HFM and AC-HFM) and the deep inter-modal interaction module (MPM) play complementary and indispensable roles, jointly contributing to HDMF's superior and stable performance.

2) *Effectiveness of pretrained lung-zone severity estimator*.

To investigate the effectiveness of the pretrained lung-zone severity estimator in ZS-HFM, we compare the BS-Net [12] with other pretrained models for CXR feature extraction, including ImageNet [14] and CheXpert [42]. As illustrated in TABLE V, adopting BS-Net as the lung-zone severity estimator yields superior performance improvement across all metrics (i.e. bACC increases by +1.8%, AUC increases by 1.7%,

TABLE VI

EFFECTIVENESS OF EACH COMPONENT. 'W/O': REMOVE THE DESIGNATED COMPONENT. 'W/': USE THE SPECIFIED COMPONENT.

Each component	AlforCOVID					
	ACC	bACC	AUC	SPEC	SENS	F1-score
w/o SAGE	72.1±1.6	71.7±1.6	77.9±1.6	77.6±6.9	65.8±6.9	68.7±2.6
w/o CAGE	71.7±1.9	71.4±2.2	78.3±1.0	76.6±5.6	66.2±9.1	68.5±4.3
w/o HyperGraph	71.0±1.2	70.7±1.0	78.5±0.6	77.0±6.6	64.3±6.3	67.5±2.1
w/o HG-Mamba	71.5±1.9	71.2±2.0	78.5±2.5	75.5±9.0	66.9±9.9	68.4±4.1
w/ Mamba	72.0±1.6	71.7±1.9	78.5±1.1	76.9±5.2	66.5±8.1	68.8±3.7
HDMF	74.7±0.8	74.3±0.7	81.1±0.6	81.3±4.1	67.3±3.6	71.4±1.0

Each component	CoCross					
	ACC	bACC	AUC	SPEC	SENS	F1-score
w/o SAGE	86.1±1.2	83.7±1.5	93.6±1.7	76.2±4.4	91.1±2.7	89.6±1.0
w/o CAGE	85.5±1.4	82.9±1.9	93.7±1.3	74.6±5.3	91.1±2.9	89.2±1.0
w/o HyperGraph	84.8±1.7	81.0±4.4	89.9±1.9	69.2±12.8	92.8±4.1	89.0±0.7
w/o HG-Mamba	85.1±3.8	82.7±4.2	92.5±3.0	75.1±8.4	90.3±5.6	88.9±3.0
w/ Mamba	85.9±1.7	82.7±2.5	92.4±1.9	73.0±5.7	92.5±1.6	89.6±1.1
HDMF	88.4±1.9	85.9±3.2	94.9±0.6	77.8±7.3	93.9±1.2	91.5±1.2

TABLE VII

SENSITIVITY OF THE ADJUSTABLE HYPERPARAMETER λ_I IN SAGE.

Hyperparameter λ_I in SAGE	AlforCOVID					
	ACC	bACC	AUC	SPEC	SENS	F1-score
0.10	74.3±1.0	73.9±1.1	81.0±0.7	81.4±4.9	66.3±6.0	70.8±2.4
0.20	74.7±0.8	74.3±0.7	81.1±0.6	81.3±4.1	67.3±3.6	71.4±1.0
0.30	73.4±0.9	73.0±1.2	80.7±0.4	80.5±6.6	65.5±8.7	69.7±3.5

Hyperparameter λ_I in SAGE	CoCross					
	ACC	bACC	AUC	SPEC	SENS	F1-score
0.10	88.3±1.0	85.3±1.2	94.7±1.1	76.2±4.8	94.4±3.3	91.4±0.9
0.20	88.4±1.9	85.9±3.2	94.9±0.6	77.8±7.3	93.9±1.2	91.5±1.2
0.30	87.9±2.7	85.3±4.6	94.7±1.2	77.3±10.6	93.3±1.8	91.1±1.7

and F1-score increases by 1.7% on AlforCOVID, and bACC rises by 3.9%, AUC rises by 3.0%, and F1-score rises by 3.5% on CoCross), which benefits from incorporating severity-aware prior knowledge into ZS-HFM through explicit regional modeling. Unlike ImageNet and CheXpert producing global generic CXR representations, zone-level severity priors better capture localized pathological patterns relevant to prognosis, which subsequently enable ZS-HFM to construct more informative hypernodes and enrich the hypergraph representation.

3) *Effectiveness of SAGE in ZS-HFM*: Severity-Aware Graph Enhancement (SAGE) is designed to accentuate abnormal lung zones by structured severity scores, which is the core of targeted CXR modeling within ZS-HFM. To demonstrate that SAGE works, we remove SAGE from the overall network. It is observed in TABLE VI that the consistent performance drop across all metrics occurs (i.e. bACC descends from 74.3% to 71.7%, AUC descends from 81.1% to 77.9%, and F1-score descends from 71.4% to 68.7% on AlforCOVID, and bACC falls from 85.9% to 83.7%, AUC falls from 94.9% to 93.6%, and F1-score falls from 91.5% to 89.6% on CoCross), confirming that SAGE directs the network's focus toward clinically abnormal CXR regions. Moreover, we perform a sensitivity analysis on the adjustable hyperparameter $\lambda_I \in \{0.1, 0.2, 0.3\}$, where $\lambda_I = 0.2$ consistently achieves the best or near-optimal performance on both datasets depicted in TABLE VII.

4) *Effectiveness of CAGE in AC-HFM*: Confidence-Aware Graph Enhancement (CAGE) is devised to emphasize important attributes and their interrelationship using estimated confidence scores, which is central to specialized table modeling within AC-HFM. As seen in TABLE VI, CAGE's removal from the overall network leads to a marked performance drop across all metrics (i.e. bACC reduces from 74.3% to 71.4%, AUC reduces from 81.1% to 78.3%, and F1-score descends from 71.4% to 68.5% on AlforCOVID, and bACC lowers from 85.9% to 82.9%, AUC lowers from 94.9% to 93.7%, and F1-

TABLE VIII

SENSITIVITY OF TARGET SELECTION RATE t IN CAGE OF AC-HFM.

Selection rate t in CAGE	AlforCOVID					
	ACC	bACC	AUC	SPEC	SENS	F1-score
0.70	71.9±0.8	71.5±0.8	79.4±0.8	77.7±3.1	65.3±3.7	68.5±1.6
0.85	74.7±0.8	74.3±0.7	81.1±0.6	81.3±4.1	67.3±3.6	71.4±1.0
1.00	72.4±1.0	72.1±1.1	80.0±1.1	79.2±1.9	65.0±3.7	69.0±1.9

Selection rate t in CAGE	CoCross					
	ACC	bACC	AUC	SPEC	SENS	F1-score
0.80	85.1±1.4	82.8±2.3	92.1±1.3	75.7±7.9	90.0±4.2	88.9±1.1
0.90	88.4±1.9	85.9±3.2	94.9±0.6	77.8±7.3	93.9±1.2	91.5±1.2
1.00	86.8±2.1	83.3±3.7	94.9±2.2	72.4±9.8	94.2±3.7	90.4±1.4

TABLE IX

SENSITIVITY OF THE ADJUSTABLE HYPERPARAMETER λ_T IN CAGE.

Hyperparameter λ_T in CAGE	AlforCOVID					
	ACC	bACC	AUC	SPEC	SENS	F1-score
0.10	74.2±1.8	73.8±1.8	81.2±1.0	81.2±6.6	66.4±7.1	70.6±3.1
0.20	74.7±0.8	74.3±0.7	81.1±0.6	81.3±4.1	67.3±3.6	71.4±1.0
0.30	72.9±2.3	72.5±2.7	80.7±1.7	79.2±4.3	65.9±9.0	69.4±4.7

Hyperparameter λ_T in CAGE	CoCross					
	ACC	bACC	AUC	SPEC	SENS	F1-score
0.10	87.5±4.8	84.5±6.3	94.0±2.0	75.1±12.6	93.9±4.5	90.9±3.4
0.20	88.4±1.9	85.9±3.2	94.9±0.6	77.8±7.3	93.9±1.2	91.5±1.2
0.30	87.3±1.8	84.4±2.9	94.5±1.2	75.1±6.7	93.6±1.6	90.7±1.2

score lowers from 91.5% to 89.2% on CoCross), confirming that CAGE effectively filters and highlights prognostically relevant attributes and their interrelationship. Furthermore, we conduct a sensitivity analysis on the target selection rate t in CAGE. As shown in TABLE VIII, the optimal performances are achieved with t of 0.85 for AlforCOVID and 0.90 for CoCross, revealing that pruning hyperedges centered on low-confidence attributes reduces redundancy and improves performance. Besides, we examine the sensitivity of the adjustable hyperparameter $\lambda_T \in \{0.1, 0.2, 0.3\}$, which indicates that $\lambda_T = 0.2$ yields the best or near-optimal results on both datasets, as demonstrated in TABLE IX.

5) Effectiveness of HyperGraph: HDMF network leverages HyperGraph as its core to enable targeted intra-modal modeling and deep inter-modal interaction. To verify the effectiveness of the hypergraph, we replace the hypergraph with the regular graph in the overall network ('w/o HyperGraph'). As demonstrated in TABLE VI, the results show that the use of HyperGraph gives 3.6% of bACC, 2.6% of AUC, and 3.9% of F1-score improvements on AlforCOVID, and 4.9% of bACC, 5.0% of AUC, and 2.5% of F1-score improvements on CoCross. It is proven that hypergraph structure introduces higher-order relationship knowledge and enriches the feature representations, thereby improving prognostic performance.

6) Effectiveness of HG-Mamba: To validate the effectiveness of HG-Mamba, we substitute HG-Mamba with MLP ('w/o HG-Mamba') and standard Mamba ('w/ Mamba') in HDMF. As illustrated in TABLE VI, employing standard Mamba has already enhanced the prognostic performance (i.e. boosts F1-score from 68.4% to 68.8% on AlforCOVID, and from 88.9% to 89.6% on CoCross), benefiting from its strong global sequence modeling capability. Adapting it to hypergraphs can further optimize the prognostic performance (i.e. boosts bACC from 71.7% to 74.3%, AUC from 78.5% to 81.1%, and F1-score from 68.8% to 71.4% on AlforCOVID, and bACC from 82.7% to 85.9%, AUC from 92.4% to 94.9%, and F1-score from 89.6% to 91.5% on CoCross), indicating that HG-Mamba more effectively captures hypergraph-specific dependencies and enhances HDMF's hypergraph modeling capacity.

TABLE X

HYPERPARAMETERS TUNING FOR LOSS TERM WEIGHTS IN THE EQUATION (5) ON THE AlFORCOVID AND CoCROSS DATASETS.

ω_4	-	-	-	-	-	0.1	1
ω_3	-	-	-	1	0.1	0.1	0.1
ω_2	-	1	0.1	0.1	0.1	0.1	0.1
ω_1	1	1	1	1	1	1	1
AlforCOVID							
ACC	72.2±1.7	72.0±1.2	72.6±2.2	73.4±0.9	73.5±1.7	74.4±1.5	74.7±0.8
bACC	71.8±1.6	71.3±1.1	72.2±2.2	73.0±1.2	73.1±1.7	74.0±1.2	74.3±0.7
AUC	78.4±1.6	79.6±1.2	79.6±1.5	80.7±0.4	79.2±0.9	82.1±0.9	81.1±0.6
SPEC	77.1±6.8	81.1±7.4	79.6±3.8	80.5±6.6	79.3±4.3	81.1±9.1	81.3±4.1
SENS	66.5±5.6	61.6±6.9	64.7±4.8	65.5±8.7	67.0±4.8	67.0±7.8	67.3±3.6
F1-score	69.0±1.8	67.2±2.2	68.9±3.0	69.7±3.5	70.4±2.4	71.1±1.8	71.4±1.0
CoCross							
ACC	85.7±2.1	86.2±2.2	86.2±1.8	86.4±2.5	87.0±3.5	88.1±1.5	88.4±1.9
bACC	83.6±3.4	84.2±3.9	84.2±2.8	82.8±3.9	83.4±4.7	85.3±3.9	85.9±3.2
AUC	94.5±1.6	94.7±1.0	93.0±2.2	93.2±1.8	95.3±1.2	95.7±1.0	94.9±0.6
SPEC	77.3±11.7	77.8±10.5	78.4±7.6	71.4±10.0	72.4±10.0	76.8±13.0	77.8±7.3
SENS	90.0±6.5	90.6±3.6	90.3±3.4	94.2±4.0	94.4±4.2	93.9±5.7	93.9±1.2
F1-score	89.2±1.9	89.7±1.4	89.7±1.4	90.2±1.7	90.6±2.6	91.2±0.9	91.5±1.2

7) Effectiveness of each loss function: We analyze the effectiveness of the severity-aware CE loss \mathcal{L}_{SEV} from ZSHFM, confidence-aware MSE loss \mathcal{L}_{CGE} from AC-HFM, and sparsification entropy loss \mathcal{L}_{SPA} from MPM. TABLE X illustrates the hyperparameters tuning process on two datasets regarding each loss term weight in the equation (5). The results demonstrate that progressively incorporating these structured regularization terms leads to consistent performance gains. Specifically, \mathcal{L}_{SEV} guides the network to learn correct pathological severity in lung zones, \mathcal{L}_{CGE} constrains the dynamic selection of high-confidence tabular attribute hyperedges, and \mathcal{L}_{SPA} suppresses redundancy in the fusion hyperedges H_F . The optimal configuration utilizing all three losses achieves competitive results across most metrics, validating that each loss function effectively injects essential inductive biases, thereby enhancing the prognostic performance and robustness.

D. Interpretability Analysis

To demonstrate the interpretability of HDMF, we conduct a double-blind reader study involving 4 experienced radiologists, who independently performed prognosis assessment on 96 randomly selected patients from the AlforCOVID dataset. This study showcases the lung zones and tabular attributes that the radiologists are concerned about during prognostic diagnosis. At the following, we make the severity scores visualization in ZS-HFM, the confidence scores visualization in AC-HFM, and the fusion hyperedges matrix H_F visualization in MPM.

1) Analysis on ZS-HFM: As depicted in Fig. 5 (a), we execute the severity scores visualization on lung zones. Consistent with the pretrained lung-zone severity estimator (BS-Net [12]), HDMF assigns one of four severity levels (0 to 3, indicating increasing severity) to each of the six lung zones. The zones of interest identified by radiologists are visualized by a binarized mask. Compared to BS-Net and radiologists, it is observed that severity scores assigned by HDMF are largely consistent with those from BS-Net, and their distribution aligns with key zones annotated by radiologists. This exhibits that HDMF can provide fine-grained, localized quantification, effectively capturing critical pathological regions and achieving precise modeling of lung abnormalities in ZS-HFM.

2) Analysis on AC-HFM: In Fig. 5 (b), we perform the confidence scores visualization on tabular attributes using a

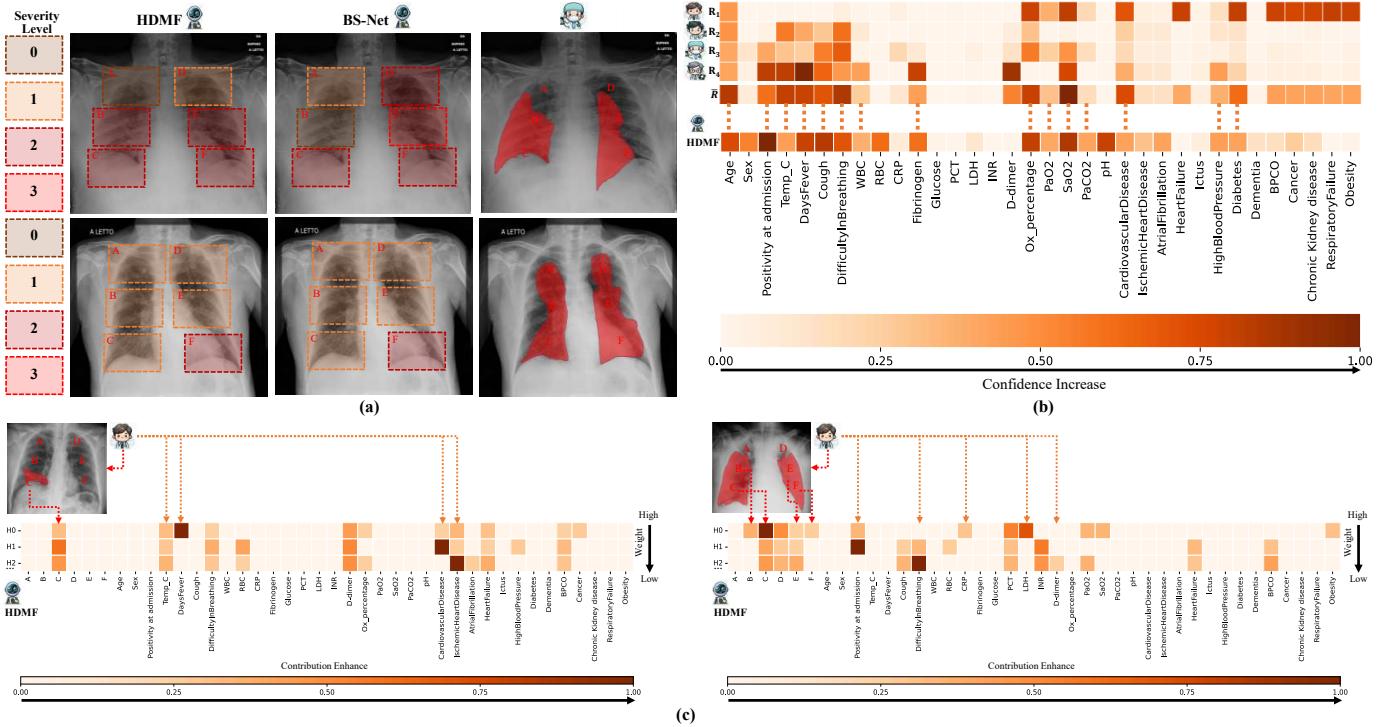


Fig. 5. Interpretability analysis for **HDMF**, including (a) Severity Scores Visualization in ZS-HFM, (b) Confidence Scores Visualization in AC-HFM, and (c) Fusion Hyperedges Matrix H_F Visualization in MPM. The color bar in (b) indicates the confidence level of each attribute, where darker shades represent higher confidence and lighter shades denote lower confidence. The color scale in (c) reflects the contribution strength of each hypernode to its associated hyperedge, with darker tones corresponding to stronger contribution and lighter tones linked to lower influence.

heatmap. For ease of comparison, the attention scores provided by the four radiologists are averaged and denoted as \bar{R} . Darker colors indicate higher confidence assigned to a given attribute, while lighter colors denote lower confidence. Compared to the attributes radiologists prioritize, it is seen that despite individual variations in radiologists' focus areas, high-confidence attributes recognized by **HDMF** closely match the aggregated opinions of radiologists, validating the credibility of confidence scores. Key indicators—such as Age, Positivity at admission, and DaysFever—demonstrate particularly strong agreement. This confirms that the confidence scores generated by **HDMF** are clinically plausible and interpretable, reinforcing the reliability of the attribute-aware modeling in AC-HFM.

3) Analysis on MPM: To further validate the capability of **HDMF** in capturing higher-order cross-modal correlation, we visualize the fusion hyperedges matrix H_F of MPM using a heatmap shown in Fig. 5 (c). Darker colors indicate the higher contribution of a hypernode within a given fusion hyperedge, while lighter colors denote lower contribution. The darkest intensity reveals that the hypernode is the center of the related hyperedge. A fusion hyperedge with a larger aggregated contribution weight implies that the involved nodes and their interactions play a dominant role in COVID-19 prognosis. H_0 to H_2 are the top 3 hyperedges ranked by node contribution, and their nodes encompass the lung zones and tabular attributes radiologists focus on, as well as other attributes strongly associated with prognosis like D-dimer, SaO_2 , and PaO_2 [16], [18]. This suggests that hypergraph interaction can simulate radiologists' decision-making for prognosis, coinciding with

clinical principles and showcasing high interpretability.

V. CONCLUSION

This paper introduces **HDMF**, a hypergraph-driven multimodal fusion network for COVID-19 prognosis. Our contributions include a ZS-HFM highlighting abnormal lung zones, an AC-HFM emphasizing key tabular attributes and their interrelationship, and a MPM enabling cross-modal interaction for high-order correlation analysis. Besides, we innovatively propose hypergraph-adapted mamba (HG-Mamba) collaborating with HGNN for comprehensive local and global hypergraph updating, enriching the network modeling capability. Extensive testing on two public datasets shows **HDMF**'s interpretability is validated against radiologist assessments, confirming its clinically relevant ability to closely mimic the decision-making process of radiologists.

REFERENCES

- [1] P. Soda, N. C. D'Amico, J. Tessadori, G. Valbusa, V. Guararsi, C. Bortolotto, M. U. Akbar, R. Sicilia, E. Cordelli, D. Fazzini *et al.*, "Aiforcovid: Predicting the clinical outcomes in patients with covid-19 applying ai to chest-x-rays. an italian multicentre study," *Medical image analysis*, vol. 74, p. 102216, 2021.
- [2] Y. Wu, A. Dravid, R. M. Wehbe, and A. K. Katsaggelos, "Deepcovid-fuse: A multi-modality deep learning model fusing chest x-rays and clinical variables to predict covid-19 risk levels," *Bioengineering*, vol. 10, no. 5, p. 556, 2023.
- [3] V. Guararsi and P. Soda, "Multi-objective optimization determines when, which and how to fuse deep networks: An application to predict covid-19 outcomes," *Computers in Biology and Medicine*, vol. 154, p. 106625, 2023.

- [4] Y. Wu, B. M. Rocha, E. Kaimakamis, G.-A. Cheimariotis, G. Petmezas, E. Chatzis, V. Kilintzis, L. Stefanopoulos, D. Pessoa, A. Marques *et al.*, “A deep learning method for predicting the covid-19 icu patient outcome fusing x-rays, respiratory sounds, and icu parameters,” *Expert Systems with Applications*, vol. 235, p. 121089, 2024.
- [5] V. Guerrasi, L. Tronchin, D. Albano, E. Faiella, D. Fazzini, D. Santucci, and P. Soda, “Multimodal explainability via latent shift applied to covid-19 stratification,” *Pattern Recognition*, vol. 156, p. 110825, 2024.
- [6] A. Khakzar, Y. Zhang, W. Mansour, Y. Cai, Y. Li, Y. Zhang, S. T. Kim, and N. Navab, “Explaining covid-19 and thoracic pathology model predictions by identifying informative input features,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2021, pp. 391–401.
- [7] F. Ruffini, L. Tronchin, Z. Wu, W. Chen, P. Soda, L. Shen, and V. Guerrasi, “Multi-dataset multi-task learning for covid-19 prognosis,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2024, pp. 251–261.
- [8] A. Chung, M. Famouri, A. Hryniowski, and A. Wong, “Covid-net clinical icu: Enhanced prediction of icu admission for covid-19 patients via explainability and trust quantification,” *arXiv preprint arXiv:2109.06711*, 2021.
- [9] Z. Wu, W. Chen, X. Li, F. Ruffini, S. Liu, L. Tronchin, D. Albano, E. Faiella, D. Fazzini, D. Santucci *et al.*, “Acgm: Attribute-centric graph modeling network for concurrent missing tabular data imputation and covid-19 prognosis,” *IEEE Journal of Biomedical and Health Informatics*, 2025.
- [10] S. R. Stahlschmidt, B. Ulfenborg, and J. Synergren, “Multimodal deep learning for biomedical data fusion: a review,” *Briefings in Bioinformatics*, vol. 23, no. 2, p. bbab569, 2022.
- [11] V. V. Danilov, D. Litmanovich, A. Proutski, A. Kirpich, D. Nefaridze, A. Karpovsky, and Y. Gankin, “Automatic scoring of covid-19 severity in x-ray imaging based on a novel deep learning workflow,” *Scientific reports*, vol. 12, no. 1, p. 12791, 2022.
- [12] A. Signoroni, M. Savardi, S. Benini, N. Adami, R. Leonardi, P. Gibellini, F. Vaccher, M. Ravanelli, A. Borghesi, R. Maroldi *et al.*, “Bs-net: Learning covid-19 pneumonia severity on a large chest x-ray dataset,” *Medical Image Analysis*, vol. 71, p. 102046, 2021.
- [13] A. Borghesi, S. Golemi, A. Scrimieri, C. M. C. Nicosia, A. Ziglani, D. Farina, and R. Maroldi, “Chest x-ray versus chest computed tomography for outcome prediction in hospitalized patients with covid-19,” *La radiologia medica*, vol. 127, no. 3, pp. 305–308, 2022.
- [14] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, “Imagenet: A large-scale hierarchical image database,” in *2009 IEEE conference on computer vision and pattern recognition*. Ieee, 2009, pp. 248–255.
- [15] F. O. Kamel, R. Magadmi, S. Qutub, M. Badawi, M. Badawi, T. A. Madani, A. Alhothali, E. A. Abozinadah, D. M. Bakhshwin, M. H. Jamal *et al.*, “Machine learning-based prediction of covid-19 prognosis using clinical and hematologic data,” *Cureus*, vol. 15, no. 12, 2023.
- [16] C. N. Hai, T. B. Duc, T. N. Minh, L. N. Quang, S. L. C. Tung, L. T. Duc, and S. Duong-Quy, “Predicting mortality risk in hospitalized covid-19 patients: an early model utilizing clinical symptoms,” *BMC Pulmonary Medicine*, vol. 24, no. 1, p. 24, 2024.
- [17] J. Chen, K. Liao, Y. Wan, D. Z. Chen, and J. Wu, “Danets: Deep abstract networks for tabular data classification and regression,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, no. 4, 2022, pp. 3930–3938.
- [18] P. Sirohiya, S. Vig, K. Pandey, J. K. Meena, R. Singh, B. K. Ratne, B. Kumar, A. Pandit, S. Bhatnagar, B. Ratne *et al.*, “A correlation analysis of peripheral oxygen saturation and arterial oxygen saturation among covid-19 patients,” *Cureus*, vol. 14, no. 4, 2022.
- [19] G. Andrew, R. Arora, J. Bilmes, and K. Livescu, “Deep canonical correlation analysis,” in *International conference on machine learning*. PMLR, 2013, pp. 1247–1255.
- [20] W. Wang, R. Arora, K. Livescu, and J. Bilmes, “On deep multi-view representation learning,” in *International conference on machine learning*. PMLR, 2015, pp. 1083–1092.
- [21] Y. Feng, H. You, Z. Zhang, R. Ji, and Y. Gao, “Hypergraph neural networks,” in *Proceedings of the AAAI conference on artificial intelligence*, vol. 33, no. 01, 2019, pp. 3558–3565.
- [22] A. S. Althenayan, S. A. AlSalamah, S. Aly, T. Nouh, B. Mahboub, L. Salameh, M. Alkubreyer, and A. Mirza, “Covid-19 hierarchical classification using a deep learning multi-modal,” *Sensors*, vol. 24, no. 8, p. 2641, 2024.
- [23] S. Sangeetha, S. K. Mathivanan, P. Karthikeyan, H. Rajadurai, B. D. Shivhare, S. Mallik, and H. Qin, “An enhanced multimodal fusion deep learning neural network for lung cancer classification,” *Systems and Soft Computing*, vol. 6, p. 200068, 2024.
- [24] T. Baltrušaitis, C. Ahuja, and L.-P. Morency, “Multimodal machine learning: A survey and taxonomy,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 41, no. 2, pp. 423–443, 2018.
- [25] C. Feichtenhofer, A. Pinz, and A. Zisserman, “Convolutional two-stream network fusion for video action recognition,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 1933–1941.
- [26] M. Glodek, S. Tschechne, G. Layher, M. Schels, T. Brosch, S. Scherer, M. Kächele, M. Schmidt, H. Neumann, G. Palm *et al.*, “Multiple classifier systems for the classification of audio-visual emotional states,” in *International Conference on Affective Computing and Intelligent Interaction*. Springer, 2011, pp. 359–368.
- [27] Z. Han, C. Zhang, H. Fu, and J. T. Zhou, “Trusted multi-view classification with dynamic evidential fusion,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 45, no. 2, pp. 2551–2566, 2022.
- [28] A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, and L. Fei-Fei, “Large-scale video classification with convolutional neural networks,” in *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, 2014, pp. 1725–1732.
- [29] A. Borghesi and R. Maroldi, “Covid-19 outbreak in italy: experimental chest x-ray scoring system for quantifying and monitoring disease progression,” *La radiologia medica*, vol. 125, no. 5, pp. 509–513, 2020.
- [30] A. Gu and T. Dao, “Mamba: Linear-time sequence modeling with selective state spaces,” *arXiv preprint arXiv:2312.00752*, 2023.
- [31] S. Bansal, S. Madisetty, M. Z. U. Rehman, C. S. Raghaw, G. Duggal, N. Kumar *et al.*, “A comprehensive survey of mamba architectures for medical image analysis: Classification, segmentation, restoration and beyond,” *arXiv preprint arXiv:2410.02362*, 2024.
- [32] Y. Yue and Z. Li, “Medmamba: Vision mamba for medical image classification,” *arXiv preprint arXiv:2403.03849*, 2024.
- [33] Z. Fang, S. Zhu, Y. Chen, B. Zou, F. Jia, L. Qiu, C. Liu, Y. Huang, X. Feng, F. Qin *et al.*, “Gfe-mamba: Mamba-based ad multi-modal progression assessment via generative feature extraction from mci,” *arXiv preprint arXiv:2407.15719*, 2024.
- [34] Z. Xing, T. Ye, Y. Yang, G. Liu, and L. Zhu, “Segmamba: Long-range sequential modeling mamba for 3d medical image segmentation,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2024, pp. 578–588.
- [35] J. Huang, L. Yang, F. Wang, Y. Nan, A. I. Aviles-Rivero, C.-B. Schönlieb, D. Zhang, and G. Yang, “Mambamir: An arbitrary-masked mamba for joint medical image reconstruction and uncertainty estimation,” *arXiv preprint arXiv:2402.18451*, 2024.
- [36] Y. Liu, M. Gehrig, N. Messikommer, M. Cannici, and D. Scaramuzza, “Revisiting token pruning for object detection and instance segmentation,” in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2024, pp. 2658–2668.
- [37] Y. Liu, Q. Zhou, J. Wang, Z. Wang, F. Wang, J. Wang, and W. Zhang, “Dynamic token-pass transformers for semantic segmentation,” in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2024, pp. 1827–1836.
- [38] E. Jang, S. Gu, and B. Poole, “Categorical reparametrization with gumble-softmax,” in *International Conference on Learning Representations (ICLR 2017)*. OpenReview. net, 2017.
- [39] K. Fu, S. Liu, X. Luo, and M. Wang, “Robust point cloud registration framework based on deep graph matching,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 8893–8902.
- [40] D. Ulyanov, A. Vedaldi, and V. Lempitsky, “Instance normalization: The missing ingredient for fast stylization,” *arXiv preprint arXiv:1607.08022*, 2016.
- [41] R. Sinkhorn, “A relationship between arbitrary positive matrices and doubly stochastic matrices,” *The annals of mathematical statistics*, vol. 35, no. 2, pp. 876–879, 1964.
- [42] J. Irvin, P. Rajpurkar, M. Ko, Y. Yu, S. Ciurea-Ilcus, C. Chute, H. Marklund, B. Haghgoor, R. Ball, K. Shpanskaya *et al.*, “Chexpert: A large chest radiograph dataset with uncertainty labels and expert comparison,” in *Proceedings of the AAAI conference on artificial intelligence*, vol. 33, no. 01, 2019, pp. 590–597.
- [43] V. Kilintzis, N. Beredimas, E. Kaimakamis, L. Stefanopoulos, E. Chatzis, E. Jahaj, M. Bitzani, A. Kotanidou, A. K. Katsaggelos, and N. Maglavera, “Cocross: An ict platform enabling monitoring recording and fusion of clinical information chest sounds and imaging of covid-19 icu patients,” in *Healthcare*, vol. 10, no. 2. MDPI, 2022, p. 276.
- [44] A. R. T. Donders, G. J. Van Der Heijden, T. Stijnen, and K. G. Moons, “A gentle introduction to imputation of missing values,” *Journal of clinical epidemiology*, vol. 59, no. 10, pp. 1087–1091, 2006.