UNLEASHING FEATURE LEARNING POTENTIAL VIA TREE-LIKE HIERARCHICAL SEMANTICS FOR PNEUMONIA LESION DATA MINING

Anonymous authors
Paper under double-blind review

ABSTRACT

Pneumonia lesion data mining inevitably deals with the low-resolution Chest X-Ray (CXR) images, posing a pivotal challenge to effective feature learning. Existing methods resort on complicated network structures to learn the latent features, while suffer from insufficient semantics due to the low resolution, resulting in performance limitations. To address this issue, this paper presents a Tree-like Hierarchical Semantic Enhancement Network (THSE-Net), aiming to unleash the feature learning potential, with an explicit Hierarchical Binary Tree (HBT) and corresponding loss functions. Specifically, THSE-Net innovatively employs the HBT to capture hierarchical semantics, and dynamically regulates the global and local feature maps to align with the semantics; and then, it dynamically computes the HBT semantic attentions based on the enhanced features to improve the feature learning capability. On the publicly available benchmark datasets, the proposed THSE-Net enhanced the IoU metric by 0.79%-1.46%, the DSC metric by 0.9%-1.39%, and the F1 score metric by 0.66-1.15% compared with the existing methods. The code will be released on GitHub upon acceptance.

1 INTRODUCTION

Pneumonia lesion data mining is compelled to enhance the feature learning capability under the low-resolution condition of Chest X-Ray (CXR) images (Madan et al., 2026)(Guo et al., 2024). Driven by the general-purpose data mining, existing methods mainly focus on devising complicated network structures to mine the latent features (Xu et al., 2025)(Yang et al., 2025), achieving certain progress because of the attention-based strategies. However, pneumonia clinical diagnosis is with its unique characteristics that could not align with the general-purpose data mining. Many clinical reports show that there are erroneous predictions like "infected areas appear in non-lung regions"(Zunaed et al., 2024), and this is already becoming a concerned issue.

In response to the above issue, various methods such as feature enhancement and multi-task learning networks are proposed in academic community. For instance, aiming to correct the erroneous predictions, Nawshad et al. (2021) proposed convolutional blocks to enhance features. Qi et al. (2022) proposed a semi-supervised learning network to enhance the associated features. Due to the indistinct boundaries between infected and non-infected areas, Bacad & Angela (2021) developed a boundary feature learning technique by multi-scale Retinex to reduce the erroneous predictions. Huang et al. (2022) designed a loss function based on Kullback-Leibler divergence consistency to enhance heterogeneous feature learning. Considering the lung segmentation and the infection classification being two tasks, Ullah et al. (2023) introduced a multi-task learning network to reduce the erroneous predictions. Tahir et al. (2021) further proposed a parallel multi-task technique named Parallel-Nets and reported better effects. Recently, the methods such as SEPM (Fang et al., 2025), PA-CXR (Dogan et al., 2026), and Segment-Anything (Zhang et al., 2024) also attracted attentions.

Although these methods have been proven effective, they fail to render sufficient semantic supports for feature learning. Regarding this, this paper presents a Tree-like Hierarchical Semantic Enhancement Network namely THSE-Net to enhance the feature learning capability. THSE-Net highlights the hierarchical binary tree (HBT), shown in Figure 1, with nodes and relationships to map the pneumonia lesion, organs, and their relations. Specifically, the first layer of the HBT is the lung

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

and non-lung regions, and the lung region is further divided into infected and non-infected areas. To enhance the feature learning capability, first, the HBT and corresponding loss functions are designed. Then, THSE-Net employs the relations of HBT nodes to correct the loss values. Further, a hierarchical spatial fusion module (HSFM) and a channel attention module (CAM) are constructed to fuse global and local feature maps regulated by the corrected loss values, and compute the HBT semantic attentions. Experimental results show that THSE-Net surpasses the existing methods.

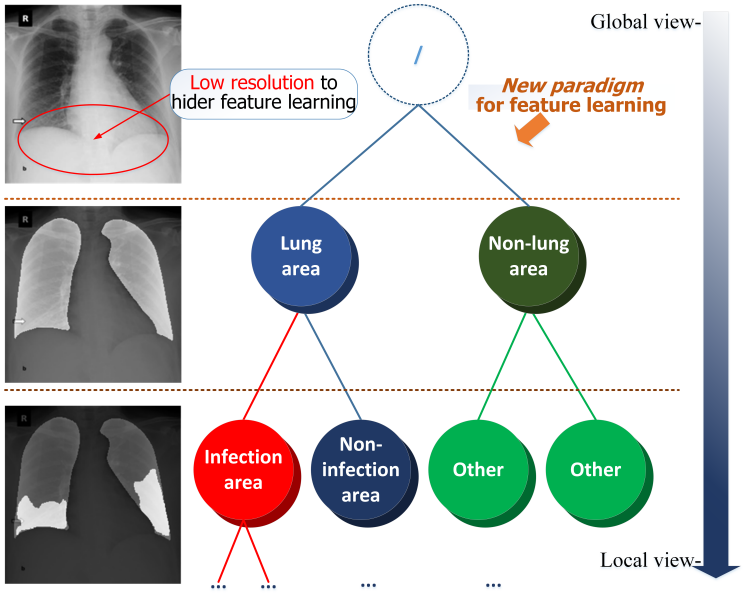


Figure 1: Tree-like hierarchical semantic structure. The tree-like hierarchical semantic structure is a Hierarchical Binary Tree (HBT) that could mitigate the semantic issue. Because the actual relations are considered within the tree, the semantic supports can be effectively enhanced.

In summary, the main contributions of this paper can be summarized as follows: **(1)** A novel feature learning paradigm namely THSE-Net is proposed to improve the performance of pneumonia lesion data mining under low-resolution image conditions, structuring the actual procedure of pneumonia diagnosis. **(2)** The HBT is designed to capture the tree-like hierarchical semantics, and the corresponding loss functions are designed to modulate the feature maps. **(3)** HSFM and CAM are designed to effectively employ the HBT with loss functions. The former introduces the multi-layer semantic fusion strategy to fuse the multi-layer feature maps regulated by the corrected loss values, and the latter further calculates the HBT semantic attentions.

2 Related Work

2.1 Single-task Feature Learning

This focuses on the feature learning to segment the infected areas. For example, the attention-based residual neural networks (Chen et al., 2021) Degerli et al. (2022) (Bruno et al., 2023), ViT-based techniques Peng et al. (2023), texture enhanced techniqures (Ding et al., 2024) are also reported. However, such method encountered a bottleneck due to the lack of task interaction.

2.2 Multi-task Feature Learning

Multi-task feature learning methods treat the pneumonia lesion data mining as multiple tasks. The superiority of such method is the mutual supports of different tasks. Traditional methods include HybridNets (Tahir et al., 2021), USegment (Chen et al., 2022), and CovidMANet (Sharma & Mishra, 2022). Recently, the frequency-aware learning (Bo et al., 2025), the channel replication strategy (Ahmed, 2025), the correlation mining techniques (Lu et al., 2022)(Sharan et al., 2024)(Pu et al., 2025) are also reported. However, the semantic relations are considerably impoverished under the low-resolution image conditions.

3 TREE-LIKE HIERARCHICAL SEMANTIC ENHANCEMENT NETWORK

3.1 PROBLEM FORMULATION

Given any image $\boldsymbol{I_i} \in \mathbb{D}_{B\times H\times W\times C}$, the task of pneumonia lesion data mining can be formulated into the following 4 steps. First, $\boldsymbol{I_i}$ is fed into the Encoder to generate the feature map $\boldsymbol{I_i^{'}}$. Then, $\boldsymbol{I_i^{'}}$ is fed into the Decoder to generate the segmented result, denoted by $\mathbb{S}^i_{H\times W\times k}$, where k denotes the number of semantic categories including non-lung region, lung region, non-infected area, and infected area. Thirdly, the softmax is imposed on $\mathbb{S}^i_{H\times W\times k}$ to constrain the values to [0, 1]. Finally, for any element $\boldsymbol{S^i_{m,n,k}} \in \mathbb{S}^i_{H\times W\times k}$, the masked image $\mathbb{Y}^i_{H\times W\times 1}$ corresponding to $\boldsymbol{I_i}$ can be obtained by the preset threshold $\mathbb{T}$, as shown in Equation (1).

$$\mathbb{Y}^i_{H\times W\times 1} \leftarrow \boldsymbol{Y^i_{m,n,a}} = \begin{cases} 0 & \boldsymbol{S^i_{m,n,k}} \leq \mathbb{T} \\ 1 & \boldsymbol{S^i_{m,n,k}} > \mathbb{T} \end{cases} (0 \leq m < H, 0 \leq n < W, k = 0, 1, 2, 3), \quad (1)$$

where $\mathbb{Y}^i_{H\times W\times 1}$ is the aggregation of $\boldsymbol{Y^i_{m,n,a}}$, and the latter represents the probability belonging to the *k*-th semantic category, indexed by *m* row and *n* column. Obviously, the Encoder and Decoder are sensitive to different feature learning strategies. Pneumonia lesion data mining makes the lung, non-lung, infected, and non-infected areas segmented; and then further predicts the pneumonia types.

3.2 MODEL FRAMEWORK

Existing methods primarily rely on complicated network structures to enhance feature learning capabilities, but fail to render sufficient semantic supports due to the low resolution. This leads to the performance limitations like "infected areas appear in non-lung regions" in clinical practice. To address this issue, this paper presents the THSE-Net. Specifically, THSE-Net leverages the HBT to characterize the relations between lesion and organs, and then modulates the feature maps to align with the semantics by the designed loss functions. To give a full play to the HBT, the HSFM in THSE-Net is constructed to fuse the global and local feature maps regulated by the corrected loss values based on the HBT and loss functions; moreover, the CAM is also constructed to compute the HBT semantic attentions to improve the feature learning capability. In short, THSE-Net provides a novel paradigm to unleash the feature learning potential, as shown in Figure 2.

3.3 HBT-BASED SEMANTIC STRUCTURE WITH LOSS FUNCTIONS

Unlike the general-purpose data mining, pneumonia clinical diagnosis is a two-stage procedure: first, it segments the lung and non-lung regions, and then it segments the infected and non-infected areas. Therefore, inspired by the policy of *a priori* knowledge guidance, if the rule "the infected areas are necessarily located in the lung region" could be considered in the feature learning, the performance might have been improved. This motivates us to ponder a tree-like hierarchical semantic structure to assist the feature learning. This is because the tree-like hierarchical semantic structure could simplify the complex semantic relations, for example, lung and non-lung areas are divided firstly, and then the lung area are further divided into infected and non-infected areas. Hence, the HBT-based semantic structure with loss functions is crucial for the proposed THSE-Net.

The HBT-based semantic structure is a strategy to correct the loss values between parent and child nodes during the feature learning (process of neural network training). By formulating related rules, it can affect the feature learning capability. Specifically, given any pixel *p* in the image and the HBT, the HBT-based semantic structure formulates the following rules.

Rule 1: For any *p*, if its node is a positive sample, then its ancestor nodes should be positive ones.
Rule 2: For any *p*, if its node is a negative sample, then its child nodes should be negative ones.
Rule 3: For any *p*, if its parent node is a negative sample, then at least one sibling node should be negative samples.

Proof. Referring to Figure 3, it is clear to showcase the reasonability of Rule 1 and Rule 2 in pneumonia lesion data mining. For Rule 3, since the corresponding parent node is a negative sample, at least one node in the next layer should be negative one because of the locality principle (Yin et al., 2020) (BIBM 2020), which include the corresponding node and the corresponding siblings, where

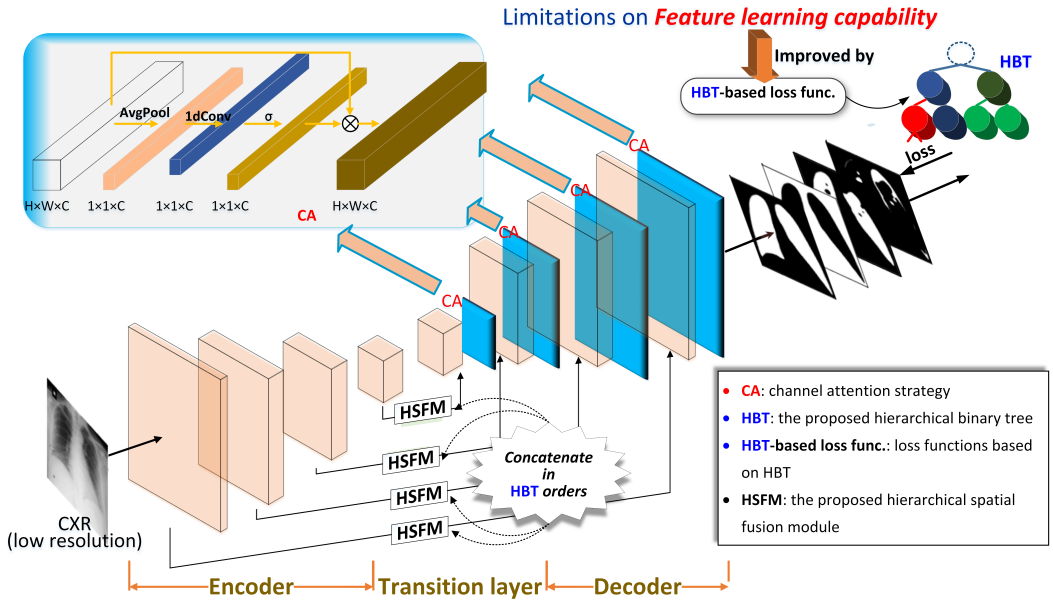
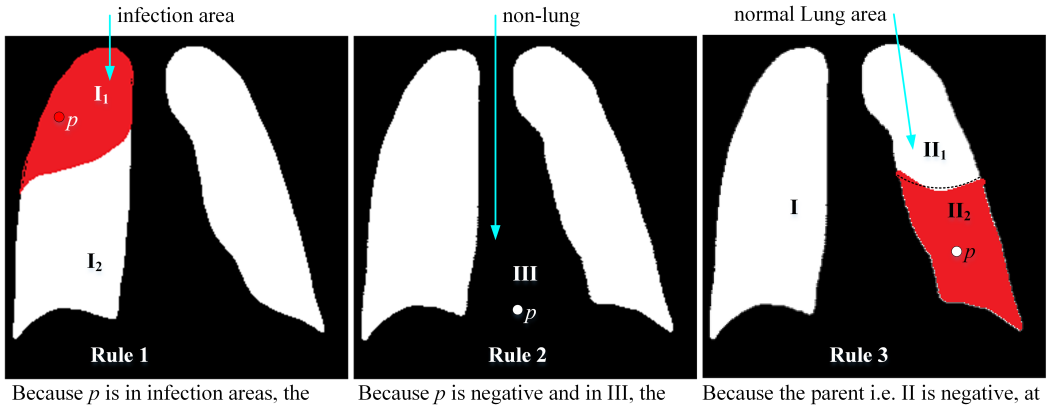


Figure 2: Illustration of the proposed THSE-Net. It exploits the HBT to add semantic guidance (Fig. 1 and Section 3.3) to the feature learning. It leverages HSFM (Section 3.4) to fuse the feature maps regulated by the corrected loss values. Then, it employs CAM (upper left corner of Figure 2 and Section 3.5) to compute the attentions. It consists of Encoder, Transition Layer, and Decoder, as shown in the figure.



Because p is in infection areas, the parent I ($I_1 + I_2$) is positive.

Because p is negative and in III, the children of III are negative samples.

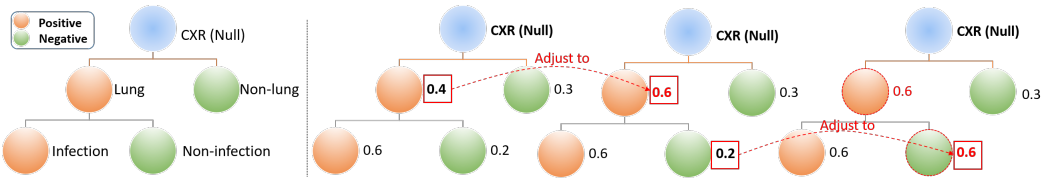
Because the parent i.e. II is negative, at least one sibling is negative, here II_1 is.

Figure 3: The designed Hierarchical semantic rules between lesion and organs.

positive and negative samples are defined according to specific domains. Therefore, Rule 3 holds. In pneumonia lesion data mining, the area of lung is the positive sample and the one of non-lung is the negative sample; further, the area of the infected is the positive sample and the one of non-infected is the negative sample.

Next, based on Rules 1, 2, and 3, the HBT-based loss functions will formulate the loss functions. Given the losses for lung, non-lung, infection, and non-infection, the HBT-based loss functions formulate the following system of equations, provided that the nodes and relations are identifiable and acquirable, shown as Equation (2).

$$\begin{cases} \boldsymbol{l_v} = \min\limits_{u \in A_v}(\boldsymbol{S_u}) & if \boldsymbol{t_v} = 1 \\ 1 - \boldsymbol{l_v} = \min\limits_{u \in C_v}(1 - \boldsymbol{S_u}) & if \boldsymbol{t_v} = 0 \; and \; \boldsymbol{t_{A_v}} = 0 \\ 1 - \boldsymbol{l_v} = 1 - \max\limits_{u \in S_v}(1 - \boldsymbol{S_u}) & if \boldsymbol{t_v} = 0 \; and \; \boldsymbol{t_{A_v}} = 1 \end{cases} \quad (Unreduced \; for \; clarity), \tag{2}$$



(a) HBT hierarchical semantics (b) Loss values (c) 1st loss correction (d) 2nd loss correction

Figure 4: Case study for feasibility: correction strategies for losses during the training process.

where $\boldsymbol{l_v}$ represents the loss of node v being classified as a positive sample, and $\boldsymbol{S_u}$ is feature probability, corresponding to the probabilities defined in Equation (1); accordingly, $1 - \boldsymbol{l_v}$ represents the loss of node v being classified as a negative sample. By the same token, this applies to A_v, where $u \in A_v$ and A_v is the ancestor set of node v. $\boldsymbol{t_v}$ represents the semantic category of node v, and $\boldsymbol{t_v} = 1$ indicates that node v is a complete positive sample, since its probability value takes the value of 1. In this circumstance, $\boldsymbol{l_v}$ should be corrected to the minimum of its ancestor node losses, corresponding to the infected area in Figure 3-Rule1. In the circumstance of $\boldsymbol{t_v} = 0$ and $\boldsymbol{t_{A_v}} = 0$, since $\boldsymbol{t_v} = 0$ and $\boldsymbol{t_{A_v}} = 0$ indicate that both v and A_v are negative samples, so $1 - \boldsymbol{l_v}$ should be corrected to the minimum of $1 - \boldsymbol{S_u}$, corresponding to the non-lung area in Figure 3-Rule2, where C_v is the child set of node v. Similarly, in the circumstance of $\boldsymbol{t_v} = 0$ and $\boldsymbol{t_{A_v}} = 0$, $\boldsymbol{l_v}$ should be corrected to the maximum of $1 - \boldsymbol{S_u}$ to rectify possible misjudgment for ancestor nodes with the assistance of sibling losses, corresponding to the normal lung area in Figure 3-Rules. S_v is the sibling set of node v, as shown in Equation (2). Obviously, the HBT-based loss functions formulate these loss corrections based on the above three rules and the characteristics of current, parent, sibling nodes, and their relations.

Case Study for feasibility: Assuming that the neural network is learning the image features associated with node *Infection*, as illustrated in Figure 4(a), the corresponding loss $\mathcal{L}_{HBT}$ at current moment is shown in Figure 4(b). According to the above Rules and Equation (2), its parent, i.e. *Lung*, should be a positive sample, because *Infection* is a positive sample. Therefore, the loss value of *Lung* should be corrected to the maximum of its child nodes to meet this rule, as shown in Figure 4(c). At next moment, assuming that the neural network will learn the image features associated with node *Non-infection*. According to Equation (2), its loss value should be corrected to 0.6, as shown in Figure 4(d). This is because it is necessary to increase the loss to approach to its parent loss, thereby enabling the neural network to make a significant adjustment, so that the infected and non-infected areas could still be correctly identified, although they exhibit low contrast.

3.4 Hierarchical Space Fusion Module: HSFM

Unlike the existing skip-connections between Encoder and Decoder, HSFM leverages a comprehensive connection method to replace the standard skip-connection to overcome the non-organized information. HSFM fuses the global and local feature maps regulated by the corrected loss values that are structured information, and then feeds the results into Decoder.

In Figure 5, HSFM showcases its structure, including the multiple semantic feature extraction, sequential concatenation, and dimensionality reduction. Inspired by the semantic feature extractions (Li et al., 2022) (CVPR 2022), HSFM exploits four-scale feature extraction: 1×1 Conv + Pooling to extract the information, 1×1 Conv + 5×5 DilatedConv (r=2) to extract the information, 1×1 Conv + 3×3 DilatedConv (r=1) to extract the information, and 1×1 Conv to retain the residual features. HSFM also performs sequential concatenation in channels, as shown in Equation (3). Finally, HSFM exploits 1×1 convolution to reduce the dimension, as shown in Equation (4). This is because the output dimension of HSFM needs to match that of the Decoder.

$$\boldsymbol{G} = ConCat(\boldsymbol{F_i^1}, \boldsymbol{F_i^2}, \boldsymbol{F_i^3}, \boldsymbol{F_i^4}) \in \mathbb{G}_{h \times w \times 2C}, \tag{3}$$

$$\boldsymbol{H} = Conv_{1\times1\times C}(\boldsymbol{G}), \tag{4}$$

where $\boldsymbol{F_i}$ represents the feature map, namely the output of the Encoder; $\boldsymbol{F_i^1}$, $\boldsymbol{F_i^2}$, $\boldsymbol{F_i^3}$, and $\boldsymbol{F_i^4}$ correspond to the feature extractions with four semantic scales. $\boldsymbol{G}$ represents the result of sequential concatenation with a channel count of 2C. h and w represent the height and width of the feature map, respectively. Finally, $\boldsymbol{H}$ is the result of dimensionality reduction.

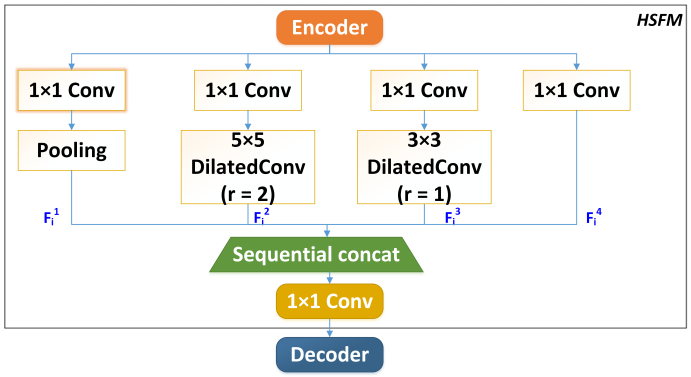


Figure 5: The designed HSFM. *r* denotes dilation rates.

3.5 Improved Channel Attention Module: CAM

For the HBT nodes, their roles are different due to their semantic layers. However, existing methods rarely consider semantic layer-related attentions. To this end, CAM is based on the HBT to enable the neural networks to adaptively change the focuses on different semantic layers, corresponding to different HBT nodes, as shown in the upper left corner of Figure 2.

The CAM consists of four parts, namely average pooling, 1d convolution attention, nonlinear activation, and residual connection. These parts are exploited to make the information differentials in different semantic layers, as shown in Equations (5) and (6).

$$\boldsymbol{\omega} = \sigma(Conv1d_k(AvgPool(\boldsymbol{z}_{h,w,c}))), \tag{5}$$

$$\boldsymbol{z}^{'}_{h,w,c} = \boldsymbol{\omega} \otimes \boldsymbol{z}_{h,w,c}, \tag{6}$$

where $\boldsymbol{z}_{h,w,c}$ represents the feature map after up-sampling, *AvgPool* represents the function of average pooling that is utilized to project the feature map into a 1d vector of length c, and $Conv1d_k$ is an 1d convolution where k=2log2c/D+1 is the size of the convolution kernel and D is the output dimensions of the linear projection layer. As c successively takes D, 2D, 4D, and 8D, the convolution kernel sizes are 3, 5, 7, and 9, respectively. σ is the sigmoid function, $\otimes$ denotes the tensor product, and $\boldsymbol{z}^{'}_{h,w,c}$ is the output of the CAM.

4 Experiments

4.1 Datasets

For pneumonia lesion data mining, COVID-QU-Ex[1] is one recognized benchmark dataset (Tahir et al., 2021). It includes 5,826 CXR images, of which 2,913 are COVID-19 images, 1,457 are common pneumonia images, and 1,456 are healthy lung images. It provides the low-resolution CXR images, the pneumonia types, lung segmentation mask labels, and infected area segmentation mask labels.

To assist the COVID-QU-Ex, a small-scale dataset, namely COVID-2020[2], is also leveraged. This dataset includes 1,195 samples where common pneumonia images are 49, heathy lung images are 1,130, and COVID-19 images are 16. This dataset is also with the lung segmentation mask labels, infected area segmentation mask labels, and pneumonia type labels.

4.2 Evaluation Metrics

As with existing methods, the Intersection over Union (IoU) and the Dice Similarity Coefficient (DSC) are utilized to evaluate the performance of segmentation, and the Accuracy (Acc), Precision (Pre), Recall (Rec), and F1 score are utilized to evaluate the performance of prediction. These evaluation metrics are provided in Appendix.

[1]https://www.kaggle.com/datasets/anasmohammedtahir/covidqu

[2]https://en.cghhospital.org/

4.3 BASELINE METHODS

The baseline methods encompass Parallel-Nets (Tahir et al., 2021), FAMNet (Bo et al., 2025), RepViT-CXR (Ahmed, 2025), PA-CXR (Dogan et al., 2026), Emb-trattunet (Bougourzi et al., 2024), and GAHFNet (Min et al., 2024). Parallel-Nets (Tahir et al., 2021) utilized a complicated collaborative network to combine the lung organ segmentation and the infected area segmentation. To align with this research, predication based on classification was supplemented. FAMNet (Bo et al., 2025) introduced complicated frequency-aware matching and multi-spectral fusion to promote the segmentation performance. To align with this research, prediction was supplemented. RepViT-CXR (Ahmed, 2025) designed a channel replication strategy to transform the single-channel CXR images to the ViT-compatible format. PA-CXR (Dogan et al., 2026) utilized large language models to segment and classify the images. Emb-trattunet (Bougourzi et al., 2024) introduced a complicated CNN-transformer fusion scheme to segment multi-classes pneumonia infection from limited CT-scans data. In this study, the original method was revised to process CXR images. GAHFNet (Min et al., 2024) introduced a global adaptive histogram feature network to automatically segment the infection regions, and it was also revised to process the CXR images.

4.4 EXPERIMENTAL SETUP

Like the baseline methods, experimental setup includes: (1) data augmentation; (2) transfer learning strategy; (3) parameter initialization; and (4) 10-fold cross validation, as shown in Appendix.

4.5 COMPARATIVE EXPERIMENTS

The comparative experiments are aimed to reflect the feature learning effects of the proposed and the baseline methods. It involves in segmentation and predication. The segmentation was evaluated by IoU and DSC metrics, while the predication was evaluated by Acc, Pre, Rec, and F1 metrics. The comparative experiments are performed on the datasets COVID-QU-Ex and COVID-2020. The ResNet18 and ResNet121 are adopted as the backbones, respectively; the U-Net++ is adopted as the inherited base model; and the CE, BCE (Zhao et al., 2024), and HBT are adopted as the loss functions.

In Table 1, the comparison showcases that the proposed THSE-Net achieves the best performance. Specifically, the proposed THSE-Net outperforms the Parallel-Nets by an average of 1.5% on the IoU metric and 1.1% on the DSC metric; outperforms the Emb-trattunet by an average of 1.4% on the IoU metric and 1.2% on the DSC metric. This is because the proposed THSE-Net can clearly represent various node semantic categories by the tree-like structure to significantly enhance the feature learning effects. THSE-Net also improves its robustness and positioning capabilities through the information confrontation of HBT-based loss functions on infected areas and non-lung areas. Therefore, the comparison shows that THSE-Net is with higher competitiveness.

Moreover, the comparison also shows the HBT-based loss function proposed in this paper not only inherits the advantages of the CE and BCE but also adds the hierarchical semantics. Therefore, better performance is achieved.

In Table 2, the comparison on pneumonia prediction also illustrates that the proposed THSE-Net significantly outperforms Parallel-Nets, ReconSegment, PA-CXR, RepViT-CXR, and Emb-trattunet. This is because THSE-Net exploits the HBT structure to assist prediction and enable the neural network be able to utilize the relations. Although Parallel-Nets achieves 100 on the Pre metric when using ResNet-18 as its backbone, the corresponding Rec metric is only 98.46. As is well known, Pre and Rec are contradictory metrics, and it is meaningful only if both reach a harmonious optimum. To sum up, the comparative experiments demonstrate that the proposed THSE-Net has obtained excellent performance and competitions.

4.6 VISUALIZATION OF RESULTS

To further confirm the effectiveness of the THSE-Net in pneumonia lesion data mining in the condition of low-resolution images, some visual examples are selected, as shown in Figure 6.

Table 1: Comparative experiments on Segmentation

Framework	Backbone	Inherited	Loss	COVID-QU-Ex		COVID-2020	
				IoU↑	DSC↑	IoU↑	DSC↑
Parallel-Nets	ResNet18	U-Net++	CE	82.90	88.06	69.50	74.20
	ResNet121	U-Net++	CE	83.05	88.21	71.61	76.74
FAMNet	ResNet18	U-Net++	CE	82.66	87.97	67.82	72.08
	ResNet121	U-Net++	CE	82.72	88.22	68.77	75.95
RepViT-CXR	ResNet18	U-Net++	CE	82.79	88.11	66.99	72.03
	ResNet121	U-Net++	CE	83.04	88.28	67.04	73.0
GAHFNet	ResNet18	U-Net++	BCE	83.20	88.20	67.27	72.14
	ResNet121	U-Net++	BCE	83.32	88.31	67.29	73.06
Emb-trattunet	ResNet18	U-Net++	BCE	83.25	88.27	68.27	72.88
	ResNet121	U-Net++	BCE	83.33	88.32	68.98	73.38
THSE-Net	ResNet18	U-Net	HBT	83.62	88.89	72.25	75.83
	ResNet121	U-Net	HBT	83.93	89.16	72.99	76.90
	ResNet18	U-Net++	HBT	83.86	89.09	72.91	76.91
	ResNet121	U-Net++	HBT	**84.24**	**89.37**	**73.11**	**77.78**

Table 2: Comparative experiments on prediction

Framework	Backbone	Inherited	Loss	COVID-QU-Ex				COVID-2020			
				Acc↑	Pre↑	Rec↑	F1↑	Acc↑	Pre↑	Rec↑	F1↑
Parallel-Nets	ResNet18	U-Net++	CE	99.23	**100**	98.46	99.22	91.99	91.95	91.78	91.86
	ResNet121	U-Net++	CE	99.23	99.14	99.31	99.22	92.48	92.46	92.13	92.29
FAMNet	ResNet18	U-Net++	CE	99.20	99.29	98.22	98.75	89.50	89.39	89.12	89.25
	ResNet121	U-Net++	CE	99.18	99.36	99.03	99.19	89.43	89.22	89.09	89.15
PA-CXR	ResNet18	U-Net++	BCE	99.28	99.31	98.24	98.77	89.54	89.44	89.15	89.29
	ResNet121	U-Net++	BCE	99.27	99.35	99.03	99.19	89.47	89.27	89.13	89.20
RepViT-CXR	ResNet18	U-Net++	CE	99.19	99.25	98.37	98.81	91.59	91.92	91.17	91.54
	ResNet121	U-Net++	CE	99.20	99.30	99.22	99.26	91.60	91.20	91.16	91.18
Emb-trattunet	ResNet18	U-Net++	BCE	98.80	99.31	98.29	98.80	88.99	88.95	88.88	88.91
	ResNet121	U-Net++	BCE	98.80	98.97	98.63	98.80	89.90	89.99	89.90	89.94
THSE-Net	ResNet18	U-Net	HBT	99.66	99.66	99.66	99.66	94.0	94.06	94.10	94.08
	ResNet121	U-Net	HBT	**99.74**	99.66	**99.83**	**99.75**	94.26	94.04	94.08	94.06
	ResNet18	U-Net++	HBT	**99.74**	99.66	**99.83**	**99.75**	94.26	94.04	94.08	94.06
	ResNet121	U-Net++	HBT	**99.74**	99.66	**99.83**	**99.75**	**94.42**	**94.21**	**94.38**	**94.29**

In Figure 6, the performance of THSE-Net is obviously better than that of the baseline for all the examples. Notably, the original CXR image shown in the second row is very dim, indicating poor image quality. For this image, the result of the baseline method suggests that infected area appears in the left and right lungs, and also the non-lung regions. However, the actual situation is only the left lung with an infected area, referring to the Ground Truth. In contrast, the result of THSE-Net is closer to the actual situation. This is because the THSE-Net enhances the information confrontation between lung infected areas and non-lung areas by the tree-like hierarchical structure and the HBT-based loss functions, thus improving the feature learning capability.

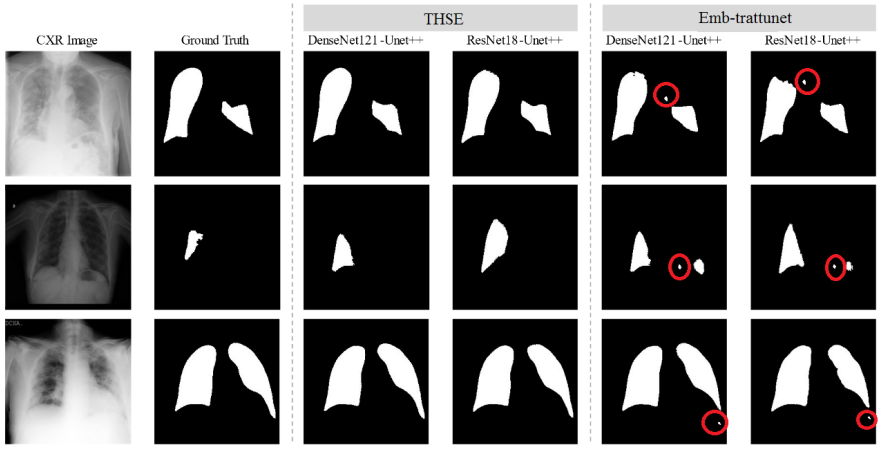


Figure 6: Visualization for THSE-Net (ours) and typical baseline method. THSE-Net has mitigated the issue of "infected areas appear in non-lung regions, " and aligns more closely with the GT.

Table 3: Ablation experiments (backbone ResNet34 + base model U-Net)

CAM	HSFM	IoU↑	DSC↑	Acc↑	Pre↑	Rec↑	F1↑
×	×	82.86	88.00	98.63	99.14	98.12	98.63
✓	×	83.84	89.06	99.66	99.66	99.66	99.66
×	✓	83.69	88.97	99.74	99.83	99.66	99.75
✓	✓	**83.96**	**89.22**	**99.83**	**99.83**	**99.83**	**99.83**
BCE		82.86	88.00	98.63	99.14	98.12	98.63
FOCAL		83.19	88.33	98.89	99.32	98.46	98.89
FTM		83.48	88.54	99.15	99.32	98.97	99.14
HBT		**83.96**	**89.22**	**99.83**	**99.83**	**99.83**	**99.83**

4.7 Ablation Studies

The related experiments include the effectiveness of CAM, HSFM, and HBT-based loss functions. All experiments were performed on the COVID-QU-Ex dataset with the evaluation metrics IoU, DSC, Acc, Pre, Rec, and F1. In Table 3, it is evident that CAM and HSFM combine their roles to achieve the best performance. If CAM removed, the performance decreases, especially the IoU and DSC metrics being significantly decreased. This is because CAM can enable the neural networks to adaptively change its focuses on different semantic layers. If HSFM removed, the performance also decreases, especially the F1 score being obviously decreased. This is because HSFM can fuse the global and local feature maps. Moreover, a total of 4 loss functions are involved in. BCE is the binary cross-entropy loss function, FOCAL is the focal loss function (Wang et al., 2024), FTM is the focal tree min loss function, and HBT is the loss function proposed in this paper. Obviously, the proposed HBT-based loss functions attain the best performance. This is because HBT-based loss functions consider the relations between nodes.

5 Conclusion

To mitigate the insufficient semantics of existing methods on pneumonia lesion data mining under low-resolution conditions, THSE-Net is proposed and evaluated on the benchmark datasets. Since the pneumonia lesion data mining is a process of progressively narrowing down the search space, the process can be naturally formulated as a hierarchical structure. Experimental results demonstrate that the THSE-Net is a competitive approach. Future work will focus on the research of sufficient semantic guidance rules, and tackle more domain-related tasks, such as the cancer prediction (Ma et al., 2025), and the thoracic and lumbar fracture assessment (Yibulayimu et al., 2026). We believe this will inspire new discoveries for feature learning.

REFERENCES

F. Ahmed. Repvit-cxr: a channel replication strategy for vision transformers in chest x-ray tuberculosis and pneumonia classification. *arXiv: 10.48550/arXiv.2509.08234, September 10*, 2025.

D.J.A. Bacad and R.A.P. Angela. Detecting covid-19 from chest x-ray images using a lightweight deep transfer learning model with improved contrast enhancement technique. *Proceedings of the 2021 International Conference on Electrical, Control and Instrumentation Engineering, ICECIE 2021*, pp. 1–8, 2021.

Y. Bo, Y. Zhu, L. Li, and H. Zhang. Famnet: frequency-aware matching network for cross-domain few-shot medical image segmentation. *Proceedings of the AAAI Conference on Artificial Intelligence, AAAI 2025*, 39:1889–1897, 2025.

F. Bougourzi, F. Dornaika, A. Nakib, and A. Taleb-Ahmed. Emb-trattunet: a novel edge loss function and transformer-cnn architecture for multi-classes pneumonia infection segmentation in low annotation regimes. *Artificial Intelligence Review*, 57(4), 2024.

A. Bruno, D.V.R. Krinski, L. Rayson, and T. Eduardo. Dacov: a deeper analysis of data augmentation on the computed tomography segmentation problem. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, 11(5):1724–1741, 2023.

H. Chen, Y. Jiang, M. Loew, and et al. Unsupervised domain adaptation based covid-19 ct infection segmentation network. *Applied Intelligence*, 52(6):6340–6353, 2022.

X. Chen, L. Yao, and Y. Zhang. Rau: An interpretable automatic infection diagnosis of covid-19 pneumonia with residual attention u-net. *Proceedings of the 22nd International Conference on Web Information Systems Engineering, WISE 2021*, pp. 122–136, 2021.

A. Degerli, S. Kiranyaz, M.E.H. Chowdhury, and et al. Osegnet: Operational segmentation network for covid-19 detection using chest x-ray images. *Proceedings of the 2022 IEEE International Conference on Image Processing (ICIP)*, pp. 2306–2310, 2022.

W. Ding, M. Abdel-Basset, H. Hawash, M. Pratama, and W. Pedrycz. Generalizable segmentation of covid-19 infection from multi-site tomography scans: a federated learning framework. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 8(1):126–139, 2024.

Y. Dogan, A. Az, O. Sogut, T. Akdemir, H. Ergenc, T.B. Umit, B.N. Armagan, and M.F. Ozturk. Domain-specific and computer-vision-driven versus general-purpose ai models in pa-cxr analysis: a comparative study with emergency-medicine specialists. *Journal of Imaging Informatics in Medicine*, 2026.

Y. Fang, J. Liang, W. Huang, H. Li, K. Su, and M. Ye. Catch your emotion: sharpening emotion perception in multimodal large language models. *Proceedings of International Conference on Machine Learning, ICML 2025*, 267:16029–16039, 2025.

Y. Guo, X. Zeng, P. Zeng, Y. Fei, L. Wen, J. Zhou, and Y. Wang. Common vision-language attention for text-guided medical image segmentation of pneumonia. *Proceedings of 27th International Conference on Medical Image Computing and Computer Assisted Intervention, MICCAI 2024*, 15009:192–201, 2024.

J. Huang, S. Yang, and X. Wang. Enhancement guidance network for classification of pneumonia in chest x-rays. *Proceedings of the 4th International Conference on Robotics, Intelligent Control and Artificial Intelligence, RICAI 2022*, pp. 940–945, 2022.

L. Li, T. Zhou, W. Wang, and et al. Deep hierarchical semantic segmentation. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2022*, pp. 1246–1257, 2022.

X.K. Lu, W.G. Wang, J.B. Shen, D.J. Crandall, and Gool Van. Segmenting objects from relational visual data. *IEEE Transactions On Pattern Analysis and Machine Intelligence*, 44(11):7885–7897, 2022.

M. Ma, T. Li, Y. Peng, L. Lin, V. Beylergil, B. Zhao, O. Akin, and X. Peng. "why is there a tumor?": tell me the reason. *show me the evidence, Proceedings of Machine Learning Research, AAAI 2025*, 267:41992–42008, 2025.

C. Madan, A. Satia, S. Basu, P. Gupta, U. Dutta, and C. Arora. Focus on texture: rethinking pre-training in masked autoencoders for medical image classification. *Proceedings of 28th International Conference on Medical Image Computing and Computer Assisted Intervention, MICCAI 2025*, 15963:207–217, 2026.

X. Min, Y. Liu, S. Zhou, H. Huang, L. Zhang, X. Gong, D. Yang, M. Wang, R. Yang, and M. Zhong. Global adaptive histogram feature network for automatic segmentation of infection regions in ct images. *Multimedia Systems*, 30(4), 2024.

M.A. Nawshad, U.A. Shami, and S. Sajid. Fraz m.m. attention based residual network for effective detection of covid-19 and viral pneumonia. *Proceedings of the 2021 International Conference on Digital Futures and Transformative Technologies, ICoDT2 2021*, pp. 1–7, 2021.

Y. Peng, T. Zhang, and Y. Guo. Cov-transnet: Dual branch fusion network with transformer for covid-19 infection segmentation. *Biomedical Signal Processing and Control*, 80:1–11, 2023.

B. Pu, L. Wang, J. Yang, X. Dong, B. Ma, Z. Chen, L. Zhao, S. Li, and K. Li. Anatomical knowledge mining and matching for semi-supervised medical multi-structure detection. *Proceedings of the AAAI Conference on Artificial Intelligence, AAAI 2025*, 39:6523–6531, 2025.

X. Qi, D.J. Foran, J.L. Nosher, and I. Hacihaliloglu. Multi-feature vision transformer via self-supervised representation learning for improvement of covid-19 diagnosis. *Proceedings of the 1st International Workshop on Medical Image Learning with Limited and Noisy Data, MILLanD 2022, Held in Conjunction with MICCAI 2022*, pp. 76–85, 2022.

R.V. Sharan, K. Qian, and Y. Yamamoto. Automated cough sound analysis for detecting childhood pneumonia. *IEEE Journal of Biomedical and Health Informatics*, 28(1):193–203, 2024.

A. Sharma and P.K. Mishra. Covid-manet: Multi-task attention network for explainable diagnosis and severity assessment of covid-19 from cxr images. *Pattern Recognition*, 131, 2022.

A.M. Tahir, M.E.H. Chowdhury, A. Khandakar, and et al. Covid-19 infection localization and severity grading from chest x-ray images. *Computers in biology and medicine*, 139, 2021.

Z. Ullah, M. Usman, and J. Gwak. Mtss-aae: Multi-task semi-supervised adversarial autoencoding for covid-19 detection based on chest x-ray images. *Expert Systems with Applications*, 216, 2023.

S.Y. Wang, Z. Qu, and L. Gao. Multi-spatial pyramid feature and optimizing focal loss function for object detection. *IEEE Transactions On Intelligent Vehicles*, 9(1):1054–1065, 2024.

M. Xu, T. Yang, T. Zhou, Q. Chen, S. Wang, and Y. Zhang. Asymmetric performance profiling using foundation models: quantifying reliability and expert capability in medical ai. *Proceedings of 2025 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2025*, pp. 1903–1908, 2025.

C. Yang, J. Shi, H. Gong, Y. Duan, and L. Yun. Biformer-seg: dual-branch visual encoding and text-guided decoding for radiological segmentation. *Proceedings of 2025 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2025*, pp. 6441–6448, 2025.

S. Yibulayimu, Y. Liu, Y. Sang, G. Zhu, H. Li, H. Lu, C. Zhao, X. Wu, and Y. Wang. Sim-to-real transformer-based shape reconstruction for automated orthopedic fracture reduction planning. *Proceedings of 28th International Conference on Medical Image Computing and Computer Assisted Intervention, MICCAI 2025*, 15962:587–597, 2026.

Y.F. Yin, H. Cheng, and F.L. Huang. Ecg pattern discovery algorithm based on local repeatability. *Proceedings of 2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM 2020)*, pp. 1206–1209, 2020.

H. Zhang, Y. Su, X. Xu, and K. Jia. Improving the generalization of segmentation foundation model under distribution shift via weakly supervised adaptation. *Proceedings of The Thirty-Fourth IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2024*, 2024.

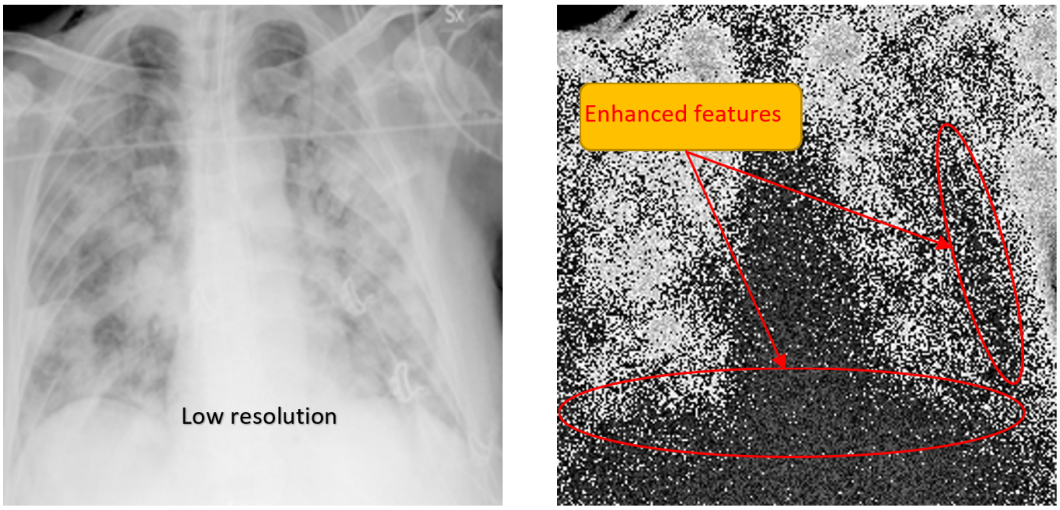
S.H. Zhao, C.R. Wang, S.Q. Liu, and H.F. Cheng. Deep feature learning for image-based kinship verification. *Proceedings of the 40th Computer Graphics International Conference, CGI 2024*, pp. 130–142, 2024.

M. Zunaed, A. Hasan, and T. Hasan. Improving pediatric pneumonia diagnosis with adult chest x-ray images utilizing contrastive learning and embedding similarity. *Proceedings of 46th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC 2024*, 2024.

A APPENDIX

A.1 EFFECTS OF TREE-LIKE HIERARCHICAL SEMANTICS

The tree-like hierarchical semantics are derived from domain-specific *a priori* knowledge. It captures associations among varying semantics to construct the HBT. HBT is employed to dynamically modulate the loss values, thereby facilitating semantic realignment. Figure A1 illustrates the effects of such hierarchical semantics on feature maps with the context of pneumonia lesion data mining.



(a) Feature maps without HBT supports (b) Feature maps with HBT supports

Figure A1: Effects of tree-like hierarchical semantics to feature maps.

A.2 EXPERIMENTAL SETUP

Like the baseline methods, experimental setup includes: (1) data augmentation on the low-resolution image sized 224×224, namely horizontal flips with the probability of 0.5 and center rotations within a random angle range of (-25, 25); (2) transfer learning strategy to get balanced parameters, namely the ImageNet-121 dataset was utilized to pre-train the model, and then the benchmark datasets are utilized to fine-tuned it; (3) model parameters, in the input layer, the image size is 224×224, in the linear projection layer, the vector dimension is 96, in the Encoder-Decoder model layer, the initial learning rate is 2e-4, and the batch size is 16; a linear warm-up method is utilized where the learning rate is warmed up to 2e-4 within the first 5 epochs, followed by a cosine annealing strategy; Adam is chosen to the optimizer; the training epochs are 100. In addition, for model parameter storing, during the training, if the model performed best on the current validation set, it would be stored; otherwise, it would be discarded and a new search would continue to find better ones.

To eliminate the bias of experimental results, a 10-fold cross validation is exploited to statistically process the experimental results, avoiding the dataset stochastic partitioning. That is, the original training dataset is randomly divided into 10 parts based on patient IDs, with 9 parts are utilized for training experiments and 1 part for validation experiments. This procedure is repeated 10 times; and the average of multiple experimental results is taken as the final result. Moreover, all the baseline methods are re-implemented in PyTorch since some of them were not incompatible with this research.

A.3 EVALUATION METRICS

As with existing methods, the Intersection over Union (IoU) and the Dice Similarity Coefficient (DSC) are utilized to evaluate the performance of segmentation, and the Accuracy (Acc), Precision (Pre), Recall (Rec), and F1 score are utilized to evaluate the performance of prediction. These evaluation metrics are formulated as follows.

$$IoU = \frac{TP}{TP + FP + FN} \tag{A1}$$

$$DSC = \frac{2TP}{2TP + FP + FN} \tag{A2}$$

$$Acc = \frac{TP + TN}{TP + FP + TN + FN} \tag{A3}$$

$$Pre = \frac{TP}{TP + FP} \tag{A4}$$

$$Rec = \frac{TP}{TP + FN} \tag{A5}$$

$$F1 = \frac{2 * Pre * Rec}{Pre + Rec} \tag{A6}$$

where *TP* (True Positive) represents that the prediction is positive and the actual result is also positive; *FP* (False Positive) represents that the prediction is positive while the actual result is negative; *TN* (True Negative) represents that the prediction is negative and the actual result is also negative; *FN* (False Negative) represents that the prediction is negative while the actual result is positive.

A.4 More Comparable Experiments

More complete comparable experiments are provided in this section, as shown in Table A1 and Table A2.

A.5 Hyper-parameter Studies

The size of CAM convolution kernel can directly affect the performance, so the corresponding studies have been conducted to get the optimal value. The experimental results are illustrated in Table A3. As shown, if the size of the convolution kernel is (3,5,7,9), the performance achieves the best. If the size is (9,9,9,9), the performance is worse, and the performance is also worse with the size of (5,5,5,5). Therefore, (3,5,7,9) is taken as the convolutional kernel size for CAM.

To determine the optimal convolution kernel size for the HSFM, the corresponding experimental studies have been also conducted. The experimental results are illustrated in Table A4. As shown, if the convolution kernel size is 5×5 and $\gamma = 2$, the performance will be the best, where γ represents the dilation rate.

A.6 Discussion

As a novel paradigm, the THSE-Net is proposed in this paper to address the feature learning capability under low-resolution conditions. To further guide the research focus and research findings, some open discussions are provided as follows.

(1) Hierarchical semantic framework. It is novel and effective, stemming from the research motivation that is to solve the issue of "infected areas appear in non-lung regions." After investigated, the associated relations between the lesion and organs should be and also be prone to formulating into a hierarchical structure to reflect the relations. This has been confirmed by experiments. That is to say, by the experiments in Table 1 and Table 2, the performance measured by DSC and F1 is illustrated being improve by (1.32%, 0.53%) compared with Tahir et al. (2021), by (1.19%, 0.96%) compared with Bougourzi et al. (2024), by (1.19%, 0.96%) compared with Min et al. (2024), by (1.3%, 0.97%) compared with Bo et al. (2025), and by (1.23%, 1.0%) compared with Ahmed (2025). To our knowledge, the performance improvement is substantial in the field of pneumonia lesion data mining.

(2) The HSFM is an effective method. HSFM is innovatively proposed inspired by the multiple semantic feature extraction and fusion, and the results are fed into Decoders. This has also been verified by experiments. From the experimental results in Table 3, it shows that the HSFM unprecedentedly improves the performance on the DSC and F1 metrics.

(3) The CAM is an effective technique. To compute the HBT semantic attentions with HBT nodes, adaptively, CAM exploits the nodes at the high layer and the nodes at the low layer to capture the

Table A1: Comparative experiments on Segmentation. "Inherited" represents the inherited base model. "Loss" represents the loss functions. IoU and DSC are evaluation metrics

Framework	**Backbone**	**Inherited**	**Loss**	**COVID-QU-Ex**		**COVID-2020**	
				IoU↑	DSC↑	IoU↑	DSC↑
Parallel-Nets	ResNet18	U-Net	CE	82.92	88.10	68.99	70.05
	ResNet50	U-Net	CE	81.73	87.02	68.92	70.02
	ResNet121	U-Net	CE	82.53	87.74	69.19	73.92
	ResNet18	U-Net++	CE	82.90	88.06	69.50	74.20
	ResNet50	U-Net++	CE	82.59	87.78	69.26	74.01
	ResNet121	U-Net++	CE	83.05	88.21	71.61	76.74
FAMNet	ResNet18	U-Net	CE	82.20	87.88	66.04	71.0
	ResNet121	U-Net	CE	82.09	87.21	66.03	71.36
	ResNet18	U-Net++	CE	82.66	87.97	67.82	72.08
	ResNet121	U-Net++	CE	82.72	88.22	68.77	75.95
RepViT-CXR	ResNet18	U-Net	CE	82.56	87.77	66.82	71.27
	ResNet121	U-Net	CE	83.01	88.0	67.04	72.91
	ResNet18	U-Net++	CE	82.79	88.11	66.99	72.03
	ResNet121	U-Net++	CE	83.04	88.28	67.04	73.0
GAHFNet	ResNet18	U-Net	BCE	82.67	87.84	66.98	71.26
	ResNet121	U-Net	BCE	83.12	88.12	67.19	72.98
	ResNet18	U-Net++	BCE	83.20	88.20	67.27	72.14
	ResNet121	U-Net++	BCE	83.32	88.31	67.29	73.06
Emb-trattunet	ResNet18	U-Net	BCE	82.78	87.97	67.01	71.76
	ResNet34	U-Net	BCE	82.86	88.00	67.03	72.0
	ResNet50	U-Net	BCE	83.06	88.15	68.24	72.31
	ResNet121	U-Net	BCE	83.01	88.05	68.11	72.03
	ResNet18	U-Net++	BCE	83.25	88.27	68.27	72.88
	ResNet34	U-Net++	BCE	82.87	88.04	67.91	72.44
	ResNet50	U-Net++	BCE	83.12	88.23	68.37	73.05
	ResNet121	U-Net++	BCE	83.33	88.32	68.98	73.38
THSE-Net	ResNet18	U-Net	HBT	83.62	88.89	72.25	75.83
	ResNet34	U-Net	HBT	83.96	89.22	72.95	76.95
	ResNet50	U-Net	HBT	83.72	89.04	72.72	76.88
	ResNet121	U-Net	HBT	83.93	89.16	72.99	76.90
	ResNet18	U-Net++	HBT	83.86	89.09	72.91	76.91
	ResNet34	U-Net++	HBT	84.08	89.26	73.0	77.63
	ResNet50	U-Net++	HBT	84.18	89.29	73.09	77.70
	ResNet121	U-Net++	HBT	**84.24**	**89.37**	**73.11**	**77.78**

global and local information, and then computes the corresponding attentions. As shown in Table 3, the CAM significantly improves the performance on the DSC and F1 metrics.

(4) Clinically, the proposed THSE-Net is promising because a vast amount of CXR images are low resolutions (Tahir et al., 2021)(Bougourzi et al., 2024)Min et al. (2024). The proposed THSE-Net met the demands. Actually, THSE-Net originates from the consideration of the actual procedure of pneumonia lesion diagnosis. Moreover, the experimental results also show that THSE-Net has exceeded the baseline methods.

(5) limitations. The three semantic guidance rules are not sufficient but constitute necessary conditions. Although the three rules are utilized to boost performance in this study, it implies that the performance is not yet the optimal and can be further improved.

Table A2: Comparative experiments on prediction. Acc, Pre, Rec, and F1 are evaluation metrics

Framework	Backbone	Inherited	Loss	COVID-QU-Ex				COVID-2020			
				Acc↑	Pre↑	Rec↑	F1↑	Acc↑	Pre↑	Rec↑	F1↑
Parallel-Nets	ResNet18	U-Net	CE	98.89	99.14	98.63	98.88	91.99	91.35	91.91	91.63
	ResNet50	U-Net	CE	98.89	98.47	99.31	98.89	91.99	91.15	91.96	91.55
	ResNet121	U-Net	CE	98.80	97.98	99.66	98.81	91.90	91.01	91.97	91.49
	ResNet18	U-Net++	CE	99.23	**100**	98.46	99.22	91.99	91.95	91.78	91.86
	ResNet50	U-Net++	CE	99.14	99.83	98.46	99.14	91.95	91.92	91.77	91.84
	ResNet121	U-Net++	CE	99.23	99.14	99.31	99.22	92.48	92.46	92.13	92.29
FAMNet	ResNet18	U-Net	CE	98.63	98.88	98.52	98.70	89.06	89.10	89.02	89.06
	ResNet121	U-Net	CE	98.77	97.98	98.61	98.29	89.21	89.31	89.0	89.15
	ResNet18	U-Net++	CE	99.20	99.29	98.22	98.75	89.50	89.39	89.12	89.25
	ResNet121	U-Net++	CE	99.18	99.36	99.03	99.19	89.43	89.22	89.09	89.15
PA-CXR	ResNet18	U-Net	BCE	98.72	98.90	98.52	98.71	89.10	89.14	89.10	89.12
	ResNet121	U-Net	BCE	98.79	98.82	98.66	98.74	89.25	89.37	89.01	89.19
	ResNet18	U-Net++	BCE	99.28	99.31	98.24	98.77	89.54	89.44	89.15	89.29
	ResNet121	U-Net++	BCE	99.27	99.35	99.03	99.19	89.47	89.27	89.13	89.20
RepViT-CXR	ResNet18	U-Net	CE	98.59	98.66	98.31	98.48	91.67	91.23	91.18	91.20
	ResNet121	U-Net	CE	98.79	98.79	98.53	98.66	91.21	91.26	91.07	91.16
	ResNet18	U-Net++	CE	99.19	99.25	98.37	98.81	91.59	91.92	91.17	91.54
	ResNet121	U-Net++	CE	99.20	99.30	99.22	99.26	91.60	91.20	91.16	91.18
Emb-trattunet	ResNet18	U-Net	BCE	98.38	98.63	98.12	98.38	88.99	89.13	88.98	89.05
	ResNet34	U-Net	BCE	98.63	99.14	98.12	98.63	88.99	89.12	88.98	89.05
	ResNet50	U-Net	BCE	98.55	99.65	97.44	98.53	88.90	89.01	88.97	88.99
	ResNet121	U-Net	BCE	98.63	99.82	96.58	98.17	88.99	89.95	88.88	89.41
	ResNet18	U-Net++	BCE	98.80	99.31	98.29	98.80	88.99	88.95	88.88	88.91
	ResNet34	U-Net++	BCE	98.63	98.63	98.63	98.63	89.48	89.46	89.13	89.29
	ResNet50	U-Net++	BCE	98.80	98.64	98.97	98.81	89.87	89.65	89.54	89.59
	ResNet121	U-Net++	BCE	98.80	98.97	98.63	98.80	89.90	89.99	89.90	89.94
THSE-Net	ResNet18	U-Net	HBT	99.66	99.66	99.66	99.66	94.0	94.06	94.10	94.08
	ResNet34	U-Net	HBT	**99.83**	99.83	**99.83**	**99.83**	**94.42**	94.02	94.11	94.06
	ResNet50	U-Net	HBT	99.74	99.66	**99.83**	99.75	94.27	94.02	94.08	94.05
	ResNet121	U-Net	HBT	99.74	99.66	**99.83**	99.75	94.26	94.04	94.08	94.06
	ResNet18	U-Net++	HBT	99.74	99.66	**99.83**	99.75	94.26	94.04	94.08	94.06
	ResNet34	U-Net++	HBT	99.74	99.83	99.66	99.75	94.26	94.04	94.08	94.06
	ResNet50	U-Net++	HBT	99.66	99.83	99.49	99.66	94.33	94.14	94.28	94.21
	ResNet121	U-Net++	HBT	99.74	99.66	**99.83**	99.75	**94.42**	**94.21**	**94.38**	**94.29**

Table A3: Convolution kernel studies of CAM

kernel	IoU↑	DSC↑	Acc↑	Pre↑	Rec↑	F1↑
(3,5,7,9)	83.96	**89.22**	**99.83**	**99.83**	**99.83**	**99.83**
(3,3,3,3)	**84.00**	89.16	99.57	99.66	99.49	99.58
(5,5,5,5)	83.96	89.18	99.57	99.83	99.32	99.58
(7,7,7,7)	83.93	89.20	99.66	99.83	99.49	99.67
(9,9,9,9)	83.84	88.97	99.66	99.83	99.49	99.67

DATA AVAILABILITY

The datasets generated during and analyzed during the current study are available from the corresponding author on reasonable request.

Table A4: Convolution kernel studies of HSFM

kernel	IoU↑	DSC↑	Acc↑	Pre↑	Rec↑	F1↑
3×3/ $\gamma = 2$	83.73	88.96	99.66	99.66	99.49	99.58
3×3/ $\gamma = 3$	83.60	88.89	99.74	99.66	99.83	99.75
5×5/ $\gamma = 1$	83.56	88.87	99.49	99.32	99.66	99.49
5×5/ $\gamma = 2$	**83.96**	**89.22**	**99.83**	**99.83**	**99.83**	**99.83**
5×5/ $\gamma = 3$	83.65	88.89	99.74	99.83	99.66	99.75

ETHICAL APPROVAL

The data used in this study were publicly available datasets on the Internet. No animals or humans were victims.

COMPETING INTERESTS

Authors declare that there are no financial and personal relationships with other people or organizations that can inappropriately influence the work, there is no professional or other personal interest of any nature or kind in any product, service and/or company that could be construed as influencing the position presented in, or the review of, the manuscript entitled.