International Journal of Computer Science and Information Technology
ISSN: 3005-9682 (Print), ISSN: 3005-7140 (Online) | Volume 5, Number 3, Year 2025
DOI: https://doi.org/10.62051/ijcsit.v5n3.12
Journal homepage: https://wepub.org/index.php/IJCSIT/index

MLFF-GNN: A Multi-Level Feature-Fusion Graph Neural Network for CXR-Image-Based Pneumonia Classification

Ruichao Tian *

School of Computer Science and Technology, Henan Polytechnic University, Jiaozuo, Henan, 454000, PR China

*Corresponding Author: trc@home.hpu.edu.cn (R. Tian)

ABSTRACT

Pneumonia does great harm to children's health. Although Graph Neural Network (GNN) has made great progress in the classification of pneumonia images, they usually use the permutation-invariant message-passing mechanism to capture the information between nodes, which is difficult to capture the interaction information between nodes. In addition, the quality of chest X-ray (CXR) images of pneumonia was uneven, and the classification model was affected by the redundant information in the pneumonia images, which could not effectively capture key information to classification. To address these problems, we proposed the multi-level feature-fusion graph neural network (MLFF-GNN) and a novel graph neural network block named the LSTM-enhanced graph network (LEGN) block. The LEGN block is designed to capture interaction information among nodes, including synergistic LEGN information within the interactions. Simultaneously, we introduce an information theory framework to explain the capture of interaction between nodes, including synergistic information within these interactions. In addition, we proposed a novel multi-level feature extraction (MLFE) module that utilizes a Convolutional Neural Network (CNN) pre-training model to extract multi-level feature maps from pneumonia CXR images and employs transfer learning to capture key features. On the two-category pneumonia classification dataset, MLFF-GNN surpasses 10 state-of-the-art (SOTA) models, with a notable 1.35% increase in accuracy. It achieves a sensitivity, precision, and accuracy of 99.30%, 99.30%, and 98.97%, respectively. On the two-category pneumonia classification dataset, MLFF-GNN maintains a high-performance standard with sensitivity, precision, and accuracy rates of 97.78%, demonstrating superior performance compared to existing models. The experimental outcomes demonstrate that the proposed model can extract the key features of CXR images of pneumonia, and can effectively capture the interaction information between nodes, which is superior to the existing methods. The proposed MLFF-GNN significantly aids clinicians by providing robust analytical support for the diagnosis of pneumonia from CXR images.

KEYWORDS

Multi-level feature-fusion graph neural network; LSTM-enhanced graph network block; Graph neural network; Transfer learning; Synergistic information

1. INTRODUCTION

Pneumonia, an infection of lung tissue, is a leading cause of death in children under five years old [1]. Traditional diagnosis heavily relies on the diagnostician's experience, making the process time-consuming and labor-intensive. CXR requires a relatively low radiation dose as a supplementary diagnostic method, making it suitable for large-scale screening and preliminary disease assessment. CXR has been widely used in pneumonia auxiliary diagnosis. However, radiologists' manual

interpretation of CXR images is time-consuming and susceptible to internal factors such as fatigue and emotions. Therefore, it is significant to introduce computer vision and artificial intelligence-assisted diagnostic techniques, such as GNN, to enhance the efficiency and accuracy of image interpretation.

As a research hotspot in artificial intelligence, GNN has the advantage of effectively capturing non-Euclidean features based on graph structures, enabling better modeling and analysis of complex relationships and dependencies [2–4]. In pneumonia diagnosis research, researchers leverage the inherent advantages of GNN to classify pneumonia CXR images by extracting multi-scale features from CXR. Meng et al. [2] proposed the Bilateral Adaptive Graph Convolution Network (BA-GCN), which utilizes both 2D and 3D features from lung CT images to identify diagnostic features in both visible and invisible environments, thus achieving diagnosis of pneumonia images. Additionally, Cui et al. [5] introduced a multi-view network for pneumonia diagnosis on pneumonia datasets. This network ranks the importance of local lung CT image areas and extracts correlations between these areas using a graph attention mechanism, allowing for the fusion of global and local information across multiple scales. Based on this, Wang et al. [6] proposed a novel method based on Topological Data Analysis (TDA), which connects nodes in the graph structure at different density scales, utilizing geometric and hierarchical topological features for image classification. However, most current GNN-based pneumonia CXR image classification models still rely on permutation-invariant aggregation functions, which fail to effectively capture interactions between neighboring nodes and lead to high feature extraction complexity. Therefore, further improving and optimizing GNN-based pneumonia CXR image classification models is necessary to enhance their accuracy and efficiency.

In GNN for pneumonia CXR images, Message Passing Neural Networks (MPNNs) [7] are widely utilized for learning node representations. Under this network structure, each node updates its hidden representation iteratively by aggregating feature vectors or hidden representations of its neighbors. Notably, the relative relationship between nodes in the graph remains unaffected by specific ordering, implying permutation invariance [8] in the aggregation function. To better comprehend the interaction information [9] between nodes, researchers decomposed the interaction information between nodes into three parts [10]: unique information from each node, redundant information, and synergistic information generated by the combination of node information. However, due to permutation invariance, traditional MPNNs fail to capture cooperative information from simultaneous observations between nodes in different orders. In this context, the core concept of synergy emphasizes that certain source variables carry more information when observed together than when observed independently. This idea emphasizes the combined effect of multiple source variables under joint action. When these source variables are observed together, their joint performance can provide more abundant and in-depth information than independent observations, thereby revealing the interrelationship and complementary characteristics between them.

Recent studies have shown that fusing multiple models can extract image feature information more effectively. Additionally, multi-level GNN can capture interactive information between nodes to a certain extent. Bai et al. [11] proposed an extensible graph-based image classification framework that combines ResNet50 [12], GNN, and Transformer [13]. While this method addresses the issue of single networks extracting medical image features to some extent, the model complexity is high, and training costs are correspondingly increased. Therefore, in response to this problem, Pati et al. [14] proposed a method that combines hierarchical GNN to map the relationship between histopathological structure and function. However, this method utilizes clustering functions such as mean clustering, which is limited in capturing feature information. Although the above method aims to extract feature information of medical images more comprehensively, it does not effectively capture synergistic information between nodes.

The extraction ability of key features from medical images is the main factor affecting the model's performance. Due to similarities in interference features, such as the distribution of blood vessels and bronchi, as well as similar interference patterns caused by shooting conditions, traditional medical

image classification models and current famous graph neural network classification models struggle to extract key features from pneumonia CXR images, such as patchy shadows and infiltration traces. Lei et al. [15] proposed a Lightweight Fusion Net (LFU-Net), which utilizes a 5-layer downsampling and 10-layer upsampling approach to obtain 15-layer features and fuse them to acquire sufficient feature information. However, this feature extraction method involves significant computational complexity. In contrast, Wang et al. [16] employed the matrix product method within the CNN to interact and concatenate the three-layer feature maps in pairs, thereby obtaining effective feature information while reducing complexity. Additionally, C.K. S et al. [17] introduced a flexible and efficient method to extract four-layer features from the CNN network. They first applied adaptive attention transformation to these features and then concatenated them to obtain effective features. The MLFE module in the MLFF-GNN model proposed in this paper utilizes the ConvNext_tiny model pre-trained on Imagenet [18] to extract multi-level feature information. Subsequently, feature mapping is performed after simple feature fusion to capture lung characteristics.

Transfer learning with CNN can extract rich and common feature representations while saving significant training time. Yu et al. [19] proposed a ResGNet framework for pneumonia images, utilizing the ResNet101 model pre-trained on the ImageNet dataset for efficient and convenient feature extraction. Furthermore, the model performance is enhanced by introducing a GNN. Additionally, Raval et al. [20] adopted a strategy to freeze all DenseNet [21] pre-training network layer weights except the last layer. They achieved better performance by freezing the network weights to prevent overfitting and updating only the weights of the feature extraction layer during the backpropagation process. Moreover, Arunesh et al. [22] used the InceptionV3 model pre-trained on the ImageNet dataset to perform data preprocessing and feature extraction on the pneumonia image dataset. This approach significantly improved the accuracy of the Ensemble-Modified Classifiers (EMC) they proposed for classifying pneumonia CXR images.

Inspired by the limitations of previous work, we propose MLFF-GNN to address unsolved challenges. This paper introduces the LEGN Block in MLFF-GNN. This module obtains the order between nodes by using spatial position sorting. It utilizes Bidirectional Long Short-Term Memory (Bi-LSTM) [23] as a special node aggregation method to aggregate and update nodes. This approach can effectively extract graph node information and node interaction information while capturing synergistic information between nodes with a simple structure. Furthermore, MLFF-GNN employs the ConvNext_tiny model pre-trained on ImageNet, which is simple in structure and does not require additional trainable parameters, enabling fast and efficient feature extraction for processing pneumonia CXR images. Additionally, MLFF-GNN combines the advantages of multi-scale feature maps and graph neural networks to further enhance model performance. This method reduces training costs and significantly improves the accuracy and robustness of pneumonia CXR image classification.

This paper will utilize MLFF-GNN to classify normal-lung CXR images, viral-pneumonia CXR images, and bacterial-pneumonia CXR images. The main contributions of this paper are as follows:

(i) We propose an efficient MLFE module that effectively extracts key feature information from pneumonia CXR images. Notably, this module has fewer trainable parameters, thereby greatly improving the training speed of the model.

(ii) We propose a LEGN Block, which facilitates the effective fusion of node features in GNN, enabling the successful extraction of interactive information between adjacent nodes, including synergistic information within these interactions.

(iii) By comparing SOTA with traditional single and hybrid networks, we validate the effectiveness of our proposed MLFF-GNN in pneumonia diagnosis.

These contributions will deepen the understanding of pneumonia diagnostic models and provide useful guidance for further enhancing the model's performance and its application in practical clinical diagnosis.

The structure of this paper is as follows: Firstly, in Section 2, we introduce our proposed MLFF-GNN in detail and explain it theoretically. Then, in Section 3, we elaborate on the dataset used, present the experimental results, and provide corresponding discussions. Finally, in Section 4, we conclude.

2. RESEARCH METHODOLOGY

Next, we elaborate on MLFF-GNN by providing a detailed explanation. As depicted in Figure 1, MLFF-GNN comprises two modules: (i) Image Feature Extraction and Fusion Module, which encompasses three components: the feature extraction part responsible for extracting features from three layers of the CNN Block, the Feature Map Transformation (FMT) consisting of feature fusion and mapping, and the adaptive neighborhood graph structure generation, tasked with generating the initial graph from FMT features. (ii) LSTM-enhanced Graph Neural Network Classification Module, composed of the GNN Block for preliminary processing of the initial graph, the LEGN Block for extracting interaction information between nodes, and the Prediction for classification. We focus on the construction of the MLFE module, including the feature extraction part and FMT in (i), and focus on the description of the theory and implementation of the LEGN Block in (ii).

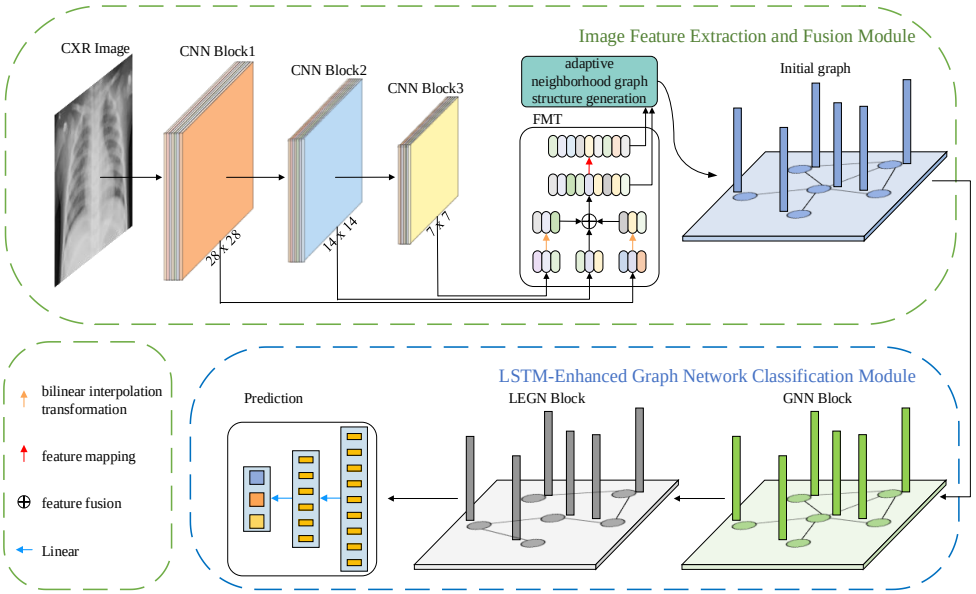

Figure 1. MLFF-GNN Model Overall Architecture Diagram

2.1. Image Feature Extraction and Fusion Module

Section 2.1.1 explains the functions and implementation details of the proposed MLFE. In section 2.1.2, we elaborate on the feature generation graph structure extracted by MLFE in the Image Feature Extraction and Fusion Module.

2.1.1. Proposed Multi-Level Feature Extraction Module

The traditional feature extraction method cannot fully consider spatial and scale information when processing medical images. This leads to relatively insufficiently rich extracted features, making capturing key feature information in pneumonia CXR images difficult. Additionally, in order to effectively extract various types of information from medical images, some methods may unnecessarily enlarge the model, resulting in excessive model parameters. To address these issues, we adopt transfer learning and multi-level feature extraction based on a CNN model pre-trained on ImageNet.

As illustrated in Figure 1, we extracted multi-level feature maps from the first CNN Block1 with a size of 28 x 28, CNN Block2 with a size of 14 x 14, and CNN Block3 with a size of 7 x 7 from the

pre-trained CNN model. Based on these three different sizes of feature maps, we propose various schemes, as detailed in Table 1. We extract feature maps from three models, ResNet18 [12], DenseNet201 [24], and ConvNext_tiny [25], to explore the best feature extraction scheme.

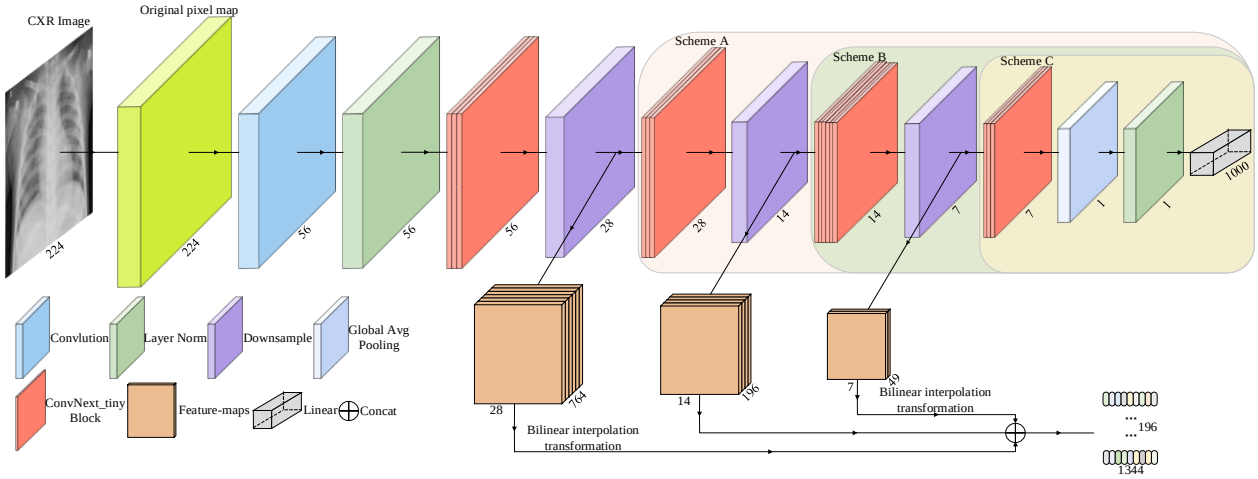

Figure 2. Different feature map extraction schemes in ConvNext_tiny model

As depicted in Figure 2, we illustrate the process of extracting pneumonia CXR image features from a pre-trained CNN model.The Scheme A denotes the extraction of the initial feature maps of size 28 x 28 in the CNN pre-training model. Subsequently, removing the orange coverage area is followed by further operation on the extracted 28 × 28 feature maps. Similarly, the Scheme B and the Scheme C entail the extraction of feature maps sized 14 x 14 and 7 x 7, respectively, from the CNN pre-training model. We designate distinct feature extraction approaches as Scheme A, B, and C.

Moreover, we propose an integrated scheme, A + B + C, which involves extracting all three feature maps followed by feature fusion. The specific methodology is delineated in Table 1. We devise a feature extraction model characterized by enhanced feature representation and reduced parameter count by comparing their efficacy across different feature extraction schemes. This augments the performance and efficiency of medical image analysis.

Table 1. Feature map extraction schemes of different CNN pre-training models.

Model	SchemeID	Feature-maps size
VGG16	A	28 x 28 x 256
	B	14 x 14 x 512
	C	7 x 7 x 512
	A+B+C	-
ResNet18	A	28 x 28 x 128
	B	14 x 14 x 256
	C	7 x 7 x 512
	A+B+C	-
DenseNet201	A	28 x 28 x 128
	B	14 x 14 x 256
	C	7 x 7 x 896
	A+B+C	-
ConvNext_tiny	A	28 x 28 x 192
	B	14 x 14 x 384
	C	7 x 7 x 768
	A+B+C	-

We utilize the characteristics of bilinear interpolation, which have a relatively smooth effect on the image or data and can help avoid jagged artifacts, thereby mitigating artifacts in pneumonia CXR images to a certain extent. Through bilinear interpolation transformation, we convert the 28 × 28 and 7 × 7 feature maps into 14 × 14 feature maps. Subsequently, we reduce the dimensionality and concatenate them to generate 196 1344-dimensional feature nodes. Next, we conduct feature mapping on the 196 1344-dimensional feature nodes to represent rich information and provide input for subsequent LSTM-Enhanced Graph Network Classification Module. We refer to the steps above as the Feature Map Transformation (FMT) module in MLFF-GNN, as depicted in Figure 1.

The algorithm corresponding to the MLFE module is described in Table 2. It accepts three feature maps, labeled as F-Map1, F-Map2, and F-Map3, with sizes designated as s1, n, and s2 respectively, arranged in descending order. Bilinear interpolation sampling is applied to F-Map1 and F-Map3 to align their dimensions with those of F-Map2. Subsequently, F-Map1, F-Map2, and F-Map3 are integrated to produce the feature map FM. Linear feature mapping is then performed on FM to derive the feature map FM_T. This procedure ensures uniformity in size across the feature maps, thus facilitating subsequent computational processes.

The above feature extraction, fusion, and mapping together constitute the MLFE module. Through the above processing, we can effectively handle pneumonia CXR images, extract key features, and provide more informative input for subsequent classifiers through such processing.

Table 2. Pseudocode of Multi-Level Feature Extraction algorithm.

Input: feature-maps F-Map1[s1][s1], F-Map2[n][n], F-Map3[s2][s2] Output: FM[n][n], FM_T[n][n] Initialization: float FM[n][n] = 0.0, FM_T[n][n] = 0.0

F-Map1 = bilinear_interpolation(F-Map1, n)
F-Map3 = bilinear_interpolation(F-Map3, n)
FM = concatenate(F-Map1, F-Map2, F-Map1)
for p in FM:
mapped_point = Linear_T(p)
FM_T.append(mapped_point)
end for

2.1.2. Features extracted by MLFE generate graph structure

As depicted in Figure 3, we use 196 1344-dimensional features in section 2.1.1 as the feature nodes to construct the graph structure. We c We utilize the characteristics of bilinear interpolation, which have a relatively smooth effect on the image or data and can help avoid jagged artifacts, thereby mitigating artifacts in pneumonia CXR images to a certain extent. Through bilinear interpolation transformation, we convert the 28 × 28 and 7 × 7 feature maps into 14 × 14 feature maps. Subsequently, we reduce the dimensionality and concatenate them to generate 196 1344-dimensional feature nodes. Next, we conduct feature mapping on the 196 1344-dimensional feature nodes to represent rich information and provide input for subsequent LSTM-Enhanced Graph Network Classification Module. We refer to the steps above as the Feature Map Transformation (FMT) module in MLFF-GNN, as depicted in Figure 1.

The algorithm corresponding to the MLFE module is detailed in Table 2. It accepts three feature maps, labeled as F-Map1, F-Map2, and F-Map3, with sizes designated as s1, n, and s2 respectively, arranged in descending order. Bilinear interpolation sampling is applied to F-Map1 and F-Map3 to align their dimensions with those of F-Map2. Subsequently, F-Map1, F-Map2, and F-Map3 are integrated to produce the feature map FM. Linear feature mapping is then performed on FM to derive the feature map FM_T. This procedure ensures uniformity in size across the feature maps, thus facilitating subsequent computational processes.

alculate the similarity between 196 feature nodes by cosine similarity. To balance model complexity and classification accuracy, we connect each feature node with its five most similar feature nodes to form initial graph. We refer to this step as adaptive neighborhood graph structure generation.

Table 3. Adaptive neighborhood graph structure generation algorithm

```
Input: feature-maps FM[n][n], FM_T, str type
Output: Graph G(V, E)
Initialization: int E[n^2][ n^2], index_array[n^2][ n^2] = 0, serial_num = 0, float V[n^2] = 0.0, similarity[n2] [n2] = 0.0,
similarity_score = 0.0
for p in FM do
V[serial_num] = p
if type = ‘similarity’ then
for i = 0 to n^2-1 do
for j = i to n^2-1 do
similarity_score = calculate_similarity(V[i], V[j])
similarity_matrix[i][j] = similarity_score
similarity_matrix[j][i] = similarity_score
end for
end for
for i = 0 to n^2– 1 do
for j = 0 to n^2-1 do
index_array[i][j] = j
end for
end for
for i = 0 to n^2-1 do
sort_indexes(index_array[i], similarity_matrix[i])
end for
for i = 0 to n^2, j = 0 to 4 do
E[i][ index_array[i][j]] = 1
end for
end if
if type = ‘4_v1’ or type = ‘8’ then
if serial_num – n >=0 then
E[serial_num][serial_num - n] = 1
end if
if serial_num % n != 0 then
E[serial_num][serial_num - 1] = 1
end if
if (serial_num + 1) % n != 0 then
E[serial_num][serial_num + 1] = 1
end if
if serial_num + n < n^2 then
E[serial_num][serial_num + n] = 1
end if
end if
if type = ‘4_v2’ or type = ‘8’ then
if serial – n >= 0 and serial_num % n != 0 then
E[serial_num][serial_num – n - 1] = 1
end if
if serial_num - n >= 0 and (serial_num + 1) % n != 0 and then
E[serial_num][serial_num – n + 1] = 1
end if
if serial_num + n < n^2 and serial_num % n != 0 then
E[serial_num][serial_num + n - 1] = 1
end if
if serial_num + n < n^2 and (serial_num + 1) % n != 0 then
E[serial_num][serial_num + n + 1] = 1
end if
end if
serial_num = serial_num + 1
end for
for i = 0 to n – 1 do
for j = 0 to n – 1 do
V[i*(n - 1) + j] = MF_T[i, j]
end for
end for
```

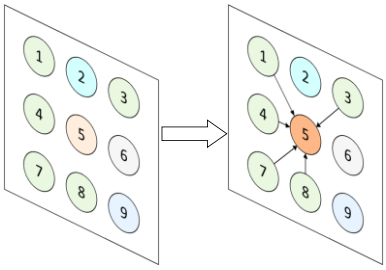

Figure 3. Adaptive neighborhood graph structure generating schematic diagram

In Table 3, we present the algorithm for generating the adaptive neighborhood graph structure. The algorithm takes the feature maps FM and FM_T as input and produces a graph $G = (V, E)$ as output, where $G = (V, E)$ represents an undirected weighted graph. Here, V denotes the set of nodes in graph G, and E denotes the set of edges. Cosine similarity and spatial position determination methods are employed to determine the neighboring nodes of each point, denoted as p, within the feature map FM.

2.2. LSTM-Enhanced Graph Network Classification Module

In this section, we elaborate on the function and implementation details of the proposed LEGN and introduce the LSTM-Enhanced Graph Network Classification Module based on LEGN. In the LSTM-Enhanced Graph Network Classification Module, we utilize the Graph Attention Network version 2 (GATv2) [26] network to process the feature nodes extracted from the MLFF-GNN feature extraction and the initial graph composed of them to obtain new feature nodes. Subsequently, we send these feature nodes to the LEGN block to extract interactive information between adjacent nodes, including synergistic information within these interactions. Specifically, the structure of the LEGN block is as follows: we sort the spatial location of medical image features in a particular manner and then input the node features sorted by spatial location into a Bi-LSTM for node aggregation. This process enables the extraction of interaction information between adjacent nodes, including synergistic information within these interactions and generating new nodes.

2.2.1. The theoretical support of LEGN Block

The representation symbols of the graph neural network are introduced as follows. Let $G = (V, E)$ be the initial graph, where $V = \{v_1, \ldots, v_N\}$ is the set of nodes in graph G, N is the number of nodes, and E is the set of edges in the graph. Let $X = [x_1, \ldots, x_N]^T \in \mathbb{R}^{N \times N}$ be the set of node features, $Y = [y_1, \ldots, y_N]^T \in \mathbb{R}^N$ be the set of feature labels. Using $\mathcal{N}(u)$ to represent the neighborhood node set of node u, we have $\mathcal{N}(u) = \{v : (v, u) \in E\}$, $\overline{\mathcal{N}}(u) = \mathcal{N}(u) \cup \{u\}$. Use $H_{\overline{\mathcal{N}}(u)} = [h_{v_1}, \ldots, h_{v_{\overline{\mathcal{N}}(u)}}]$ to denote the neighborhood containing node u. Correspondingly, there are feature sets $X_{\mathcal{N}(u)} = \{x_v : v \in \mathcal{N}(u)\}$ and $X_{\overline{\mathcal{N}}(u)} = \{x_v : v \in \overline{\mathcal{N}}(u)\}$.

GNN utilizes the graph structure and node features to infer the hidden representation of each node. It employs the classic message-passing mechanism, where each node aggregates the characteristics of its neighboring nodes and then updates itself. This process can be expressed by Equations (1) and (2).

$$t_u^{(l)} = \text{Aggregate}^{(l)}\left(\left\{h_v^{(l-1)} : v \in \mathcal{N}(u)\right\}\right) \tag{1}$$

$$h_u^{(l)} = \text{Combine}^{(l)}\left(h_u^{(l-1)}, t_u^{(l)}\right) \tag{2}$$

Where $t_u^{(l)}$ represents the representation of $\mathcal{N}(u)$ after aggregation, and the superscript of $h_u^{(l)}$ is usually hidden to simplify the expression. Taking Graph Attention Networks (GAT) [27] as an example, the node attention coefficient $e(h_v, h_u)$ is first calculated,

$$e(h_v, h_u) = LeakyReLU(\, a^T \cdot [Wh_v || Wh_u]) \tag{3}$$

Where $a \in \mathbb{R}^{2d'}$, $W \in \mathbb{R}^{d' \times d}$, d' represents the weight dimension, d represents the hidden feature dimension. The coefficient matrix a and the weight W are both learnable, and then the attention coefficient is normalized by softmax,

$$\alpha_{vu} = softmax_v(\, e(h_v, h_u)) = \frac{exp(e(h_v, h_u))}{\sum_{v' \in \mathcal{N}(u)} exp(e(h_{v'}, h_u))} \tag{4}$$

The attention coefficient α_{vu} after softmax normalization can be obtained from Equation 4. Next, GAT updates the hidden node representation aggregately,

$$h_u{}^{new} = \sigma\big(\textstyle\sum_{v \in \mathcal{N}(u)} \alpha_{vu} \cdot Wh_v\big) \tag{5}$$

From Equation (5), the hidden representation $h_u{}^{new}$ of the new node u after aggregation update can be obtained.

In the GNN block of MLFF-GNN mentioned in this paper, we utilize GATv2 as the GNN Block. Unlike the static attention mechanism used in GAT, GATv2 adopts a dynamic attention mechanism, enabling the central node to dynamically adjust its attention to its neighboring nodes to better obtain the interaction information between nodes. Specifically, GATv2 splits the coefficient matrix in Equation (3) into two parts $a = [a_1 || a_2] \in \mathbb{R}^{2d'}$, thus Equation (3) becomes the following Equation (6),

$$e(h_v, h_u) = LeakyReLU(\, a_1{}^T W h_v + a_2{}^T W h_u) \tag{6}$$

The Partial Information Decomposition (PID) proposed by Williams et al. [28] is described as follows. Let $H_{\overline{\mathcal{N}}(u)} = [h_{v_1}, \ldots, h_{v_{|\overline{\mathcal{N}}(u)|}}] \in \mathbb{R}^{|\overline{\mathcal{N}}(u)| \times d}$ denote the hidden representation set of nodes in $\overline{\mathcal{N}}(u)$. Mutual information [29] is a measure that can help us understand the interaction of omitted information. When only considering the interaction between nodes, mutual information is equivalent to interactive information. If $\overline{\mathcal{N}}(u)$ and h_u are assumed to follow the distributions $p(\overline{\mathcal{N}}(u))$ and $p(h_u)$ respectively, then the mutual information $F(h_u; H_{\overline{\mathcal{N}}(u)})$ between $\overline{\mathcal{N}}(u)$ and h_u is defined in Equation (7).

$$F(h_u; H_{\overline{\mathcal{N}}(u)}) = \iint p\big(h_u, H_{\overline{\mathcal{N}}(u)}\big)\, log\Big(\frac{p\big(h_u, H_{\overline{\mathcal{N}}(u)}\big)}{p(h_u)\, p\big(H_{\overline{\mathcal{N}}(u)}\big)}\Big) dh_u\, dH_{\overline{\mathcal{N}}(u)} \tag{7}$$

Williams et al. proposed PID, as shown in Figure 4. According to the principle of multivariate information decomposition derived by Williams et al. in information theory, as shown in Equation (8), the mutual information between nodes can be decomposed into the following three parts: (i) For any $v \in \mathcal{N}(u)$, the unique information U_v refers to the information that the node v has and the neighboring node of v does not have. (ii) Redundant information R refers to the same information in two or more neighboring nodes. (iii) Synergistic information S refers to the additional information generated when neighbor nodes are jointly observed, which is determined by a certain combination order between neighbor nodes in the neighborhood.

$$F\left(h_u; H_{\overline{\mathcal{N}}(u)}\right) = \sum_{v \in \overline{\mathcal{N}}(u)} U_v + R + S \tag{8}$$

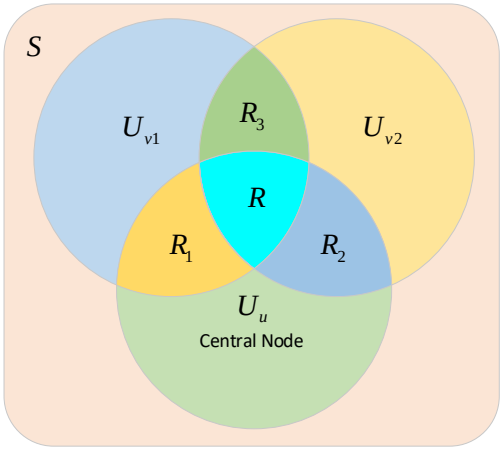

Figure 4. Part of the Partial Information Decomposition

In Figure 4, if the central node u and two neighboring nodes $v1$ and $v2$ are aggregated and updated according to the message-passing mechanism of Equations (1) and (2), mutual information can be obtained:

$$F(h_u; h_{v1}) + F(h_u; h_{v2}) = U_{v1} + U_{v2} + R_1 + R_2 + 2R_3 + 2R \tag{9}$$

According to the PID theoretical framework, the mutual information obtained is as follows:

$$F(h_u; h_{v1}, h_{v2}) = U_{v_1} + U_{v_2} + U_u + R_1 + R_2 + R_3 + R + S \tag{10}$$

The formulas indicate that GNN based on message-passing is more susceptible to interference from redundant information and cannot capture collaborative information effectively. This illustrates their incapacity to capture interaction information between neighboring nodes effectively.

2.2.2. Implementation of LEGN Block

The formulation details of the LEGN Block that can capture collaborative information are as follows. By utilizing spatial positioning, the d-dimensional disordered hidden representation set $\{h_1, \ldots, h_P\}$, where $P = \left|\overline{\mathcal{N}}(u)\right|$, is transformed into an ordered sequence $\{h_{\pi(1)}, \ldots, h_{\pi(P)}\}$. $\pi(\bullet)$is a permutation function, which is used to sort the hidden representations. Specifically, $\pi(\bullet)$considers the one-dimensional node feature as the feature graph of $n \times n$, and provides an ordered sequence representation $h_{sorted(u)} = \{h_1, \ldots, h_i, \ldots, h_u, \ldots, h_j, \ldots, h_p\}$, where $i \subset (1, u)$ and $j \subset (u, p)$, of hidden representation h_u as the ordering part (Ordering (Space)) depicted in Figure 5.

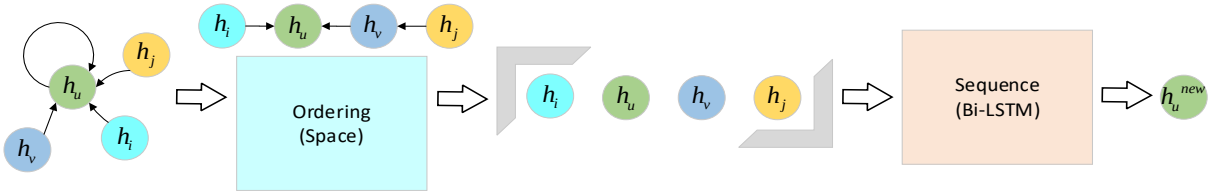

Figure 5. LEGN process diagram

After obtaining the ordered sequence $h_{sorted(u)}$ representing the hidden information h_u, we use this sequence $h_{sorted(u)}$ as input and apply an RNN to process the neighborhood information of h_u. According to PID theory, Bi-LSTM exhibits the following characteristics: a forgetting gate capable of discarding redundant information, an input gate that efficiently aggregates unique information from the hidden features, and a memory cell that retains collaborative information. Therefore, Bi-LSTM is

highly suitable as an efficient node aggregation method. The overall structure diagram is depicted in the sequential part (Sequence (Bi-LSTM)) in Figure 5.

In Figure 5, the LEGN Block is employed to aggregate and update the hidden representation of node h_u. The spatial sorting algorithm is utilized to sort its neighboring nodes, and the hidden feature representation of the node sequence is used as the input of Bi-LSTM to obtain the updated representation ${h_u}^{new}$ of h_u. This process can be expressed as Equation (11).

$$ {h_u}^{new} = \text{Bi} - \text{LSTM}(h_{sorted(u)}) \tag{11} $$

In Equation (11), the output ${h_u}^{new}$ will be fed into the Prediction, consisting of two Linear layers as depicted in Figure 1, to accomplish the classification of pneumonia CXR images.

Table 4. LSTM-Enhanced Graph Network Block algorithm.

```
Input: Graph G (V, E)
Output: h[P]
Initialization: int n = 1, E[n^2][n^2] = 0, float forward_sorted[P][] = 0.0,back_sorted[P][] = 0.0
while n^2 < P:
    n = n + 1
end while
for i = 0 to n^2 - 1 do
    E[i][i] = 1
    if (i + 1) % n != 0 then
        E[i][i + 1] = 1
    end if
    if (i + 1) % n == 0 and i + 1 < n^2 then
        E[i][i + 1] = 1
    end if
end for
for i = 0 to P - 1 do
    forward_sorted[i].append(V[i])
    back_sorted[i].append(V[i])
    j = i, k = i
    while j > 0:
        for m = 0 to Length – 1 do
            if E[j][m] !=0 and j > m then
                forward_sorted[i].append(V[m])
            end if
        end for
        j = j – 1
    end while
    reverse(forward_sorted[i])
    while k < P – 1:
        for m = 0 to P – 1 do
            if E[k][m] !=0 and j < m then
                back_sorted[i].append(V[m])
            end if
        end for
        k= k + 1
    end while
    reverse(back_sorted[i])
end for
for i = 0 to P - 1 do
    h[i] = Bi-LSTM (forward_sorted[i], back_sorted[i])
end for
```

Table 4 presents the algorithm corresponding to the LEGN block. This algorithm takes a graph $G = (V, E)$ as input and produces a list of components h, which comprises ${h_u}^{new}$. Initially, the algorithm restructures the graph into an $n \times n$ sized matrix and updates the edges E in $G = (V, E)$ based on specific spatial positions. It then retrieves the spatial position sequences forward_sorted and back_sorted for P nodes in V. Subsequently, it computes h for each node after undergoing Bi-LSTM processing.

Through this processing, we can utilize LEGN block to process pneumonia CXR image, thereby improving the performance and accuracy of the LSTM-Enhanced Graph Network Classification Module classification.

3. EXPERIMENTS

This section elaborates on the datasets and introduces the experimental environment and settings. In Section 3.2, we summarize the dataset distribution and display sample images. In Section 3.3, we propose a model evaluation index. In Section 3.4, the results of the feature extractor ablation experiment on the pneumonia CXR dataset are presented. Section 3.5 includes ablation experiments on the CXR dataset, covering the best graph structure construction algorithm, feature mapping algorithm, GNN feature fusion block, and LEGN block. Then, by comparing MLFF-GNN with SOTA CNN, Transformer, MLP, and hybrid GNN models in pneumonia diagnosis methods, we verify the excellent performance of MLFF-GNN in pneumonia diagnosis methods.

3.1. Experimental Setting

Our experiments were conducted on a server with 64 GB of RAM, an Intel Xeon Silver 4214 CPU, and an RTX8000 GPU with 48 GB of video memory. All algorithms were implemented using Python 3.9.0 and PyTorch 1.10.0.

Based on the consideration of model performance and training efficiency, we set the following training parameters on the two pneumonia CXR datasets: the feature extraction batch size is 128, the training and test batch size is 32, and the training rounds are 50. We utilize the cosine annealing function as the learning rate scheduling strategy, where each cycle corresponds to a complete training cycle, and the initial learning rate is set to 0.005. The cross-entropy loss function is chosen as the loss function. After computing the training loss, we employ the Sharpness-Aware Minimization (SAM) [30] optimization algorithm based on Stochastic Gradient Descent (SGD) to optimize the model parameters and reduce the sensitivity of the model to noise. Furthermore, for the training and testing of the two-class pneumonia dataset, we used five-fold cross-validation to reduce the risk of model overfitting. For training the three-category pneumonia dataset, we conducted training and testing based on a 7:3 ratio between the training and test sets.

3.2. Dataset

The proposed deep learning model utilizes two distinct pneumonia CXR image datasets for training and testing: the Mendeley dataset [31] and the Chest X-ray dataset [32]. The Mendeley dataset represents a two-class classification scenario with two categories: normal and pneumonia. The Chest X-ray dataset covers three different categories, reflecting a three-class classification scenario: normal, bacterial pneumonia, and viral pneumonia. Table 5 presents the sample distribution of these two datasets, while Figure 6 displays some image examples of chest X-ray images from these datasets.

Table 5. Distribution of chest X-ray data sets in two-class and three-class classification scenarios.

Dataset	Partition	Normal	Viral Pneumonia	Bacteria Pneumonia	Total/Partition	Total
Mendeley [31]: Binary Task: Two classes	Train	1260	3412		4265	5840
	Test	315	853		1575	
Chest X-ray [32]: Multiclass Task: Three Classes	Train	3500	3500	3500	10500	15000
	Test	1500	1500	1500	4500	

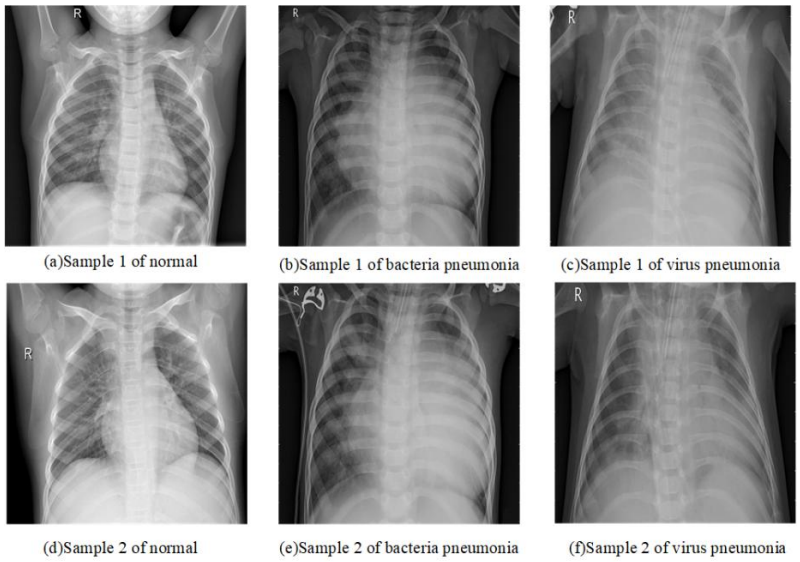

Figure 6. Some examples from the CXR dataset

3.3. Evaluation Metrics

In classification problems, the confusion matrix serves as a tool to assess the performance of classification models. Typically presented as a table, the confusion matrix features rows representing the actual categories and columns representing the categories predicted by the model. Within the confusion matrix, correct predictions of positive cases by the model are termed True Positives (TP), while incorrect predictions of negative samples as positive are termed False Positives (FP). Similarly, correct predictions of negative examples are labeled True Negatives (TN), whereas incorrect predictions of positive samples as negative are termed False Negatives (FN). Various evaluation metrics are employed to assess the performance of the proposed hybrid deep learning framework, including accuracy (Acc), precision (Pre), sensitivity (Sen), F1 value, and AUC.

Precision: It calculates the model's precision by defining the ratio of images correctly classified as positive classes to all images predicted as positive classes.

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \tag{12}$$

Sensitivity: It calculates the ratio of images correctly predicted as positive examples by the model to all actual positive examples, reflecting the model's ability to recognize true positives.

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP+FN}} \tag{13}$$

Accuracy: It is defined as the proportion of accurately classified images to the total number of actual test images.

$$\text{Accuarcy} = \frac{\text{TP+TN}}{\text{TP+FP+TN+FN}} \tag{14}$$

F1-score: It comprehensively considers precision and sensitivity, achieving a good balance between the imbalance between positive and negative examples in classification problems.

$$\text{F1} - \text{score} = \frac{2*\text{Precision}*\text{Sensitivity}}{\text{Precision+Sensitivity}} \tag{15}$$

AUC: It represents the area under the ROC curve, which reflects the model's performance under various thresholds. The ROC curve is plotted with True Positive Rate (TPR) as the vertical axis and False Positive Rate (FPR) as the horizontal axis.

$$\text{TPR} = \frac{\text{TP}}{\text{TP+FN}} \tag{16}$$

$$\text{FPR} = \frac{\text{FP}}{\text{FP+TN}} \tag{17}$$

$$\text{AUC} = \sum_{i=1}^{N-1} \frac{1}{2} (\text{TPR}_i + \text{TPR}_{i+1}) \cdot (\text{FPR}_{i+1} - \text{FPR}_i) \tag{18}$$

3.4. Contrast Experiment of Feature Extractor

Table 6. The classification performance of different feature map extraction schemes. (unit: %)

Model	Sen.	Pre.	F1	Acc.	AUC
VGG16(A)	98.78 ± 0.28	98.64 ± 0.27	98.71 ± 0.20	98.71 ± 0.20	98.26 ± 0.66
VGG16(B)	98.90 ± 0.21	98.94 ± 0.26	98.92 ± 0.16	98.42 ± 0.23	98.06 ± 0.41
VGG16(C)	98.31 ± 0.24	98.31 ± 0.39	98.31 ± 0.29	97.53 ± 0.43	96.72 ± 0.72
VGG16 (A+B+C)	99.16 ± 0.14	98.88 ± 0.21	99.02 ± 0.11	98.56 ± 0.17	98.09 ± 0.33
ResNet18(A)	98.08 ± 0.44	97.37 ± 0.64	97.72 ± 0.53	96.66 ± 0.77	95.51 ± 1.10
ResNet18(B)	98.97 ± 0.39	98.71 ± 0.47	98.84 ± 0.36	98.31 ± 0.53	97.74 ± 0.83
ResNet18(C)	98.31 ± 0.12	94.95 ± 0.33	96.60 ± 0.14	94.95 ± 0.22	98.15 ± 0.39
ResNet18 (A+B+C)	99.11 ± 0.24	99.09 ± 0.34	99.10 ± 0.23	98.68 ± 0.34	98.28 ± 0.48
DenseNet201 (A)	98.73 ± 0.36	98.60 ± 0.16	98.66 ± 0.18	98.05 ± 0.26	97.47 ± 0.29
DenseNet201 (B)	98.92 ± 0.37	98.81 ± 0.37	98.86 ± 0.29	98.34 ± 0.43	97.91 ± 0.62
DenseNet201 (C)	98.94 ± 0.40	99.25 ± 0.22	99.10 ± 0.30	98.68 ± 0.44	98.26 ± 0.66
DenseNet201 (A+B+C)	99.15 ± 0.22	99.27 ± 0.28	99.21 ± 0.20	98.85 ± 0.30	98.53 ± 0.29
ConvNext_tiny (A)	98.68 ± 0.29	98.71 ± 0.16	98.70 ± 0.11	98.10 ± 0.15	97.58 ± 0.43
ConvNext_tiny (B)	99.01 ± 0.10	99.02 ± 0.27	99.02 ± 0.12	98.56 ± 0.17	98.16 ± 0.44
ConvNext_tiny (C)	98.90 ± 0.50	98.92 ± 0.28	98.91 ± 0.17	98.41 ± 0.25	97.93 ± 0.20
ConvNext_tiny (A+B+C)(Ours)	99.30 ± 0.42	99.30 ± 0.26	99.29 ± 0.24	98.97 ± 0.35	98.60 ± 0.26

In this section, to validate the effectiveness of the MLFE module of MLFF-GNN, we assessed various feature map extraction schemes on the two-class pneumonia CXR dataset. Specifically, we extracted features using CNN models, including VGG16 [33], ResNet18, DenseNet201, and ConvNext_tiny. Under identical conditions, experiments were conducted based on the feature map extraction scheme outlined in Table 1, and the results of different feature extraction schemes are presented in Table 6.

The experimental findings reveal that when the feature extraction map comprises only one map, the F1-score is generally higher when utilizing the intermediate feature map. However, the efficacy of extracting multi-level feature maps consistently surpasses that of the single-level feature extraction scheme. This underscores that the multi-level feature extraction scheme of the pre-trained ConvNext_tiny model is capable of extracting sufficiently representative key features from pneumonia CXR images.

3.5. Model Ablation Experiment

As presented in Table 7, we conducted ablation experiments on various MLFF-GNN modules to identify the optimal module combination. Considering the constraints of prior studies and experimental apparatus, we meticulously set and compared the parameters of each module. Referring to Figure 1, we proposed distinct combinations of FMT modules and configured Linear layers with 0, 1, and 2 layers to determine the most effective feature mapping scheme. We proposed four adaptive neighborhood graph structure construction methods.As shown in Figure 3., the "Neighborhoods_4 v1" mentioned in Table 7 denotes that the neighboring nodes of node 5 are nodes 2, 4, 8, and 6. If the neighboring nodes are absent, they are excluded. Similarly, "Neighborhoods_4 v2" indicates that the neighboring nodes of node 5 are nodes 1, 3, 7, and 9, while "Neighborhoods_8" includes nodes 1, 2, 3, 4, 6, 7, 8, and 9 as the neighboring nodes. We employed a cosine similarity calculation method between nodes to determine the composition scheme of node 5 in Figure 3, connecting it with the five most similar nodes. These four adaptive graph structure construction algorithms are shown in Table 3.

In the GNN Block section, depicted in Figure 1, we compare four common and effective GNN models, GCN [34], GraphSAGE [35], GAT, and GATv2, to determine the GNN Block with the most robust node feature fusion capability. In the LEGN Block, as illustrated in Figure 1, we evaluate the performance of six different recurrent neural networks, including unidirectional LSTM, RNN [36] and GRU [37] (referred to as U-LSTM, U-RNN, U-GRU, respectively), as well as bidirectional LSTM, RNN, and GRU (referred to as Bi-LSTM, Bi-RNN, Bi-GRU, respectively), to identify the LEGN Block that achieves the optimal extraction of interactive information between adjacent nodes, including synergistic information within these interactions.

As shown in Table 7, the optimal classification performance for pneumonia CXR images is achieved by employing a double-layer Linear, costructing the graph structure based on cosine similarity, utilizing GATv2 as the GNN Block, and adopting Bi-LSTM in the LEGN Block. Therefore, this configuration is chosen as the benchmark model and compared with other proposed module combination schemes.

In the feature mapping component of the MLFF, employing a two-layer Linear structure gets a precision that is 0.26% higher compared to the absence of feature mapping. This implies that the utilization of a two-layer Linear structure for feature mapping can capture more comprehensive feature representations, consequently resulting in enhanced classification accuracy.

In the adaptive graph structure generation section of MLFF, employing a similarity-based scheme for graph structure generation results in an accuracy improvement of 0.99% compared to using the "Neighborhoods_4 v2" scheme based on spatial position sorting. This suggests that utilizing a similarity-based approach for generating the graph structure can enhance the capture of interaction information among nodes.

In the GNN Block, the utilization of various GNN methods for preliminary extraction of inter-node interaction information consistently results in high F1-score. Particularly notable is the attainment of a 99.30% F1-score when employing GATv2. This suggests that harnessing graph structures facilitates the extraction of interaction information among local feature representations of pneumonia CXR images.

In the LEGN Block, the utilization of Bi-RNN for node aggregation gets an accuracy of 97.64%, which is 1.33% lower than the accuracy of 98.97% achieved with Bi-LSTM. This suggests that Bi-RNN lacks the capability to eliminate redundant information, resulting in reduced accuracy in the classification of pneumonia CXR images. Conversely, the LEGN Block employing Bi-LSTM demonstrates the ability to eliminate redundant information among nodes and extract node interaction information, including synergistic information.

Table 7. MLFF-GNN partial module ablation experimental results (unit: %)

Module	option	Sen.	Pre.	F1	Acc.	AUC
feature mapping	-	99.18 ± 0.22	99.04 ± 0.28	99.11 ± 0.04	98.70 ± 0.06	98.50 ± 0.15
	Linear	99.04 ± 0.26	99.22 ± 0.20	99.13 ± 0.18	98.73 ± 0.25	98.51 ± 0.26
adaptive neighborhood graph structure construction	Neighborhoods_4 v1	99.06 ± 0.56	99.06 ± 0.56	98.69 ± 0.40	98.08 ± 0.58	98.42 ± 0.16
	Neighborhoods_4 v2	98.97 ± 0.62	98.28 ± 0.38	98.62 ± 0.40	97.98 ± 0.58	98.50 ± 0.16
	Neighborhoods_8	98.90 ± 0.67	98.41 ± 0.16	98.65 ± 0.39	98.03 ± 0.57	98.48 ± 0.15
GNN Block	GCN	99.23 ± 0.28	99.23 ± 0.17	99.23 ± 0.16	98.87 ± 0.22	98.34 ± 0.31
	GraphSAGE	99.25 ± 0.39	99.28 ± 0.19	99.26 ± 0.21	98.92 ± 0.30	98.37 ± 0.15
	GAT	99.09 ± 0.28	99.25 ± 0.30	99.17 ± 0.23	98.78 ± 0.33	98.30 ± 0.28
LEGN Block	U-LSTM	99.13 ± 0.22	99.34 ± 0.33	99.24 ± 0.18	98.89 ± 0.26	98.33 ± 0.22
	U-RNN	99.34 ± 0.21	99.07 ± 0.22	99.20 ± 0.20	98.84 ± 0.30	98.25 ± 0.12
	U-GRU	99.09 ± 0.43	99.16 ± 0.41	99.12 ± 0.25	98.72 ± 0.37	98.40 ± 0.19
	Bi-RNN	98.48 ± 1.59	98.29 ± 1.47	98.38 ± 1.51	97.64 ± 2.21	98.26 ± 0.35
	Bi-GRU	99.09 ± 0.43	99.16 ± 0.41	99.12 ± 0.25	98.72 ± 0.37	98.49 ± 0.15
MLFF-GNN (Ours)	-	99.30 ± 0.42	99.30 ± 0.26	99.29 ± 0.24	98.97 ± 0.35	98.60 ± 0.26

3.6. Comparison with SOTA Pneumonia Diagnosis Methods on CXR

To validate the effectiveness of MLFF-GNN in pneumonia diagnosis, we compare it with other SOTA methods on both the three-class CXR pneumonia dataset and the two-class CXR pneumonia dataset. We have standardized the dataset division as presented in Table 5. For the two-class classification tasks distinguishing between normal and pneumonia cases, the performance of SOTA methods using CNN [38], Transformer [39, 40], MLP [41, 42], and GNN [43] architectures is summarized in Table 8. For the three-class classification tasks distinguishing between normal, viral pneumonia, and bacterial pneumonia cases, the performance of SOTA methods using CNN [44], Transformer [45] and GNN architectures is summarized in Table 9. Additionally, MLFF-GNN demonstrates superior performance quickly with fewer parameters. The comparison results of specific parameters are provided in Table 10.

Table 8. Classification performance of SOTA pneumonia diagnosis model on two-category pneumonia CXR dataset (unit: %)

Model	Sen.	Pre.	F1	Acc.	AUC
VGG16	97.73 ± 0.27	97.59 ± 0.64	97.66 ± 0.31	96.58 ± 0.46	95.59 ± 0.86
ResNet18	97.63 ± 0.57	96.17 ± 0.73	96.89 ± 0.31	95.43 ± 0.47	93.55 ± 0.93
DenseNet201	98.99 ± 0.32	97.78 ± 0.76	98.38 ± 0.30	97.62 ± 0.45	96.45 ± 0.97
ConvNext_tiny	96.37 ± 0.55	95.50 ± 0.76	95.93 ± 0.37	94.02 ± 0.55	92.02 ± 1.01
Swin-T [39]	97.02 ± 0.31	89.78 ± 0.45	93.15 ± 0.36	91.78 ± 0.36	90.22 ± 0.41
VIT-B-16 [40]	94.00 ± 0.37	88.32 ± 0.41	91.15 ± 0.29	88.61 ± 0.27	87.26 ± 0.32
CycleMLP-B4 [41]	98.08 ± 0.28	92.36 ± 0.24	95.34 ± 0.33	94.08 ± 0.25	93.14 ± 0.29
GazeGNN [43]	96.96 ± 0.35	96.55 ± 0.24	96.45 ± 0.29	96.58 ± 0.37	97.45 ± 0.28
SCNet [42]	98.39 ± 0.32	92.20 ± 0.35	95.64 ± 0.24	94.75 ± 0.27	93.69 ± 0.31
EFF-PCNet [38]	98.34 ± 0.26	95.76 ± 0.29	97.18 ± 0.34	96.85 ± 0.26	96.11 ± 0.23
MLFF-GNN(Ours)	99.30 ± 0.42	99.30 ± 0.26	99.29 ± 0.24	98.97 ± 0.35	98.60 ± 0.26

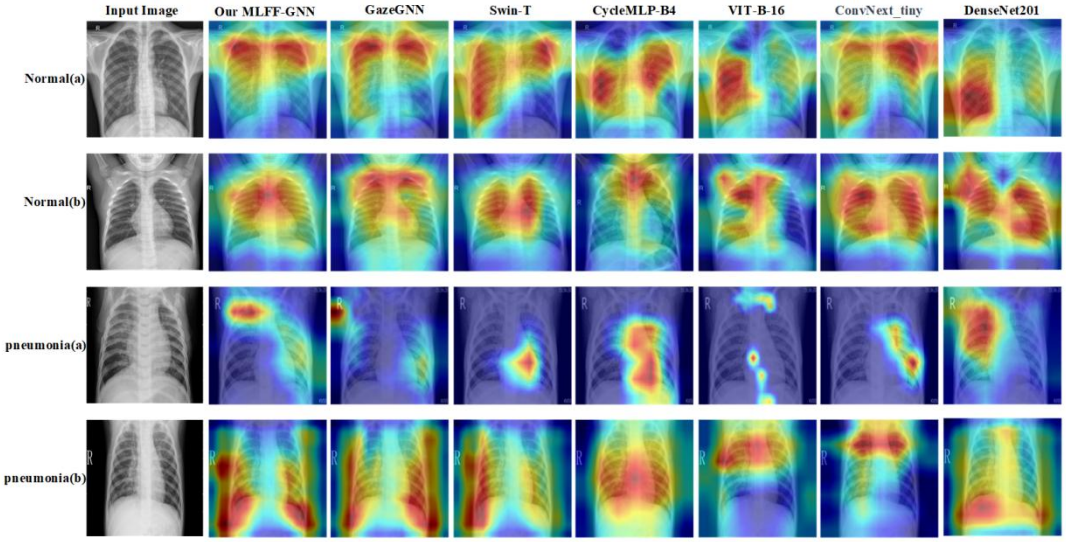

Figure 7. Comparison of classification results between MLFF-GNN and other models on a two-category pneumonia dataset with Grad-CAM visualization

As shown in Figure 7, we selected two normal CXR images, labeled as Normal (a) and Normal (b), as well as two pneumonia CXR images, labeled as Pneumonia (a) and Pneumonia (b), to demonstrate the Grad-CAM visualization effects. Compared to all the contrastive models, MLFF-GNN focuses on fewer irrelevant regions, demonstrating a strong capability to remove redundant information from features. In comparison to GazeGNN, MLFF-GNN pays attention to more regions that can be indicative of pneumonia, showcasing its ability to extract more crucial information. When compared to Transformer models Swin-T, VIT-B-16, and MLP model CycleMLP-B4, MLFF-GNN's attention is more precise, demonstrating better robustness.

Table 9. Classification performance of SOTA pneumonia diagnosis model on three-category pneumonia CXR dataset (unit: %)

Model	Sen	Pre	F1-score	Acc	AUC
VGG16	97.02	97.09	97.01	97.02	97.77
ResNet18	96.09	96.11	96.08	96.09	97.07
DenseNet201	96.16	96.22	0.9616	96.16	97.12
ConvNext_tiny	94.62	95.12	94.87	96.18	93.14
VIT-B-16	91.38	91.38	91.37	91.38	93.53
UBNetV1 [44]	89.36	88.29	88.82	88.72	87.91
UBNetV2 [44]	89.12	89.04	89.08	89.39	89.00
Transformer-E [45]	95.42	93.51	94.46	94.65	94.78
GazeGNN	96.32	96.09	96.20	95.96	97.34
MLFF-GNN (Ours)	97.78	97.80	97.78	97.78	98.33

As shown in Table 9, MLFF-GNN achieves the highest F1 score of 97.78% in the experiment of three-class classification of pneumonia CXR images. This reflects the strong stability and reliability of our proposed model in pneumonia diagnosis tasks. Furthermore, MLFF-GNN also achieves the highest accuracy of 98.33%, indicating its considerable generalization ability in pneumonia CXR image classification tasks.

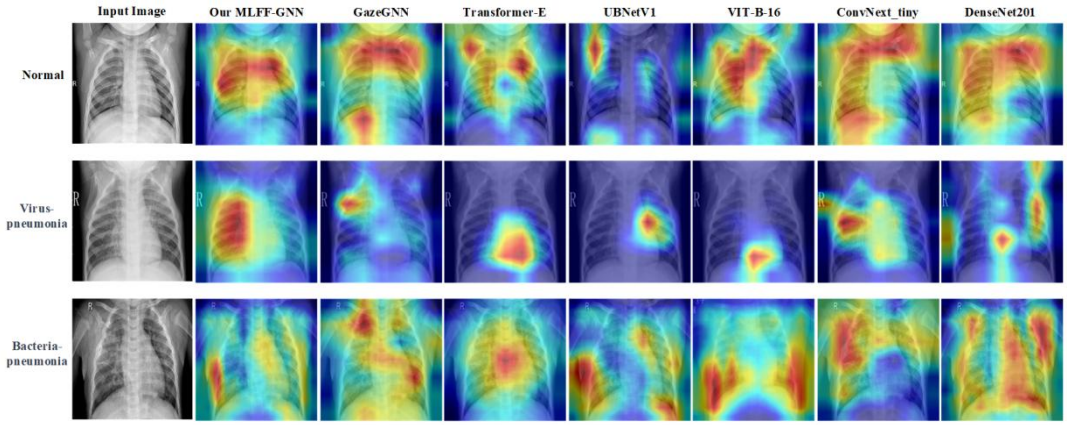

Figure 8. Comparison of classification results between MLFF-GNN and other models on a three-category pneumonia dataset with Grad-CAM visualization

As shown in Figure 8, we selected one normal CXR image labeled as "Normal", one viral pneumonia CXR image labeled as "Virus-pneumonia", and one bacterial pneumonia CXR image labeled as "Bacteria-pneumonia" to demonstrate the Grad-CAM visualization effects. Compared to all the contrastive models, MLFF-GNN still manages to focus on fewer irrelevant regions, showcasing its robust capability to remove redundant information from features. In the classification of "Virus-pneumonia" and "Bacteria-pneumonia", MLFF-GNN attends to more regions indicative of pneumonia, demonstrating its powerful ability to extract crucial information.

As depicted in Table 10, MLFF-GNN showcases the shortest training duration among the baseline models on the binary classification pneumonia CXR dataset, accomplishing five-fold cross-validation within 4722 seconds, equivalent to one-fifth of the time taken by ConvNext_tiny. However, MLFF-GNN achieves a 98.60% AUC, which surpasses ConvNext_tiny's 92.02% AUC by 6.58%. Despite its 50% parameter surplus compared to GazeGNN, MLFF-GNN exhibits shorter training time and attains a Precision of 99.30%, exceeding GazeGNN's Precision of 96.55% by 2.75%. This variance is ascribed to GazeGNN's utilization of multimodal features, while MLFF-GNN capitalizes on a pre-trained CNN model to extract pneumonia CXR image characteristics. These findings underscore

MLFF-GNN's capability to achieve superior pneumonia CXR image classification with reduced parameterization and abbreviated training periods.

Table 10. Model parameters and running time on the two-category pneumonia CXR dataset.

Model	Parameters (M)	Time (s)
VGG16	13.43	9565
ResNet18	11.18	7095
DenseNet201	18.10	8432
ConvNext_tiny	27.82	25085
Swin-T	27.50	8545
ViT-B	86.20	7495
Cycle MLP-B4	26.29	7962
GazeGNN	9.69	5685
SCNet	55.10	7526
EFF-PCNet	20.39	9252
MLFF-GNN (Ours)	16.68	2152

4. CONCLUSIONS AND FUTURE WORK

In this paper, we propose MLFF-GNN, which combines an MLFE module and a graph neural network block, the LEGN Block, to achieve effective classification of pneumonia CXR images. The experimental results demonstrate that our pneumonia diagnosis method outperforms 10 SOTA models on the two-category dataset of pneumonia mentioned in this paper and is superior to 9 SOTA models on the three-category dataset of pneumonia. There are three main reasons for the superior performance of our MLFF-GNN model: (i) Our proposed MLFE module effectively captures rich key features in pneumonia CXR images. (ii) The LEGN Block fully considers the importance of synergistic information and utilizes Bi-LSTM to remove redundant information effectively. (iii) The components of the MLFF-GNN model are relatively flexible. Combining this with transfer learning reduces the number of training parameters and training time while maintaining excellent classification ability for pneumonia CXR images.

This research exhibits two main deficiencies: (i) The paper's investigation utilized a pre-trained CNN solely to extract features from pneumonia CXR images, limiting feature acquisition to a single channel. (ii) The dataset employed comprises pneumonia CXR images, which fails to confirm whether MLFF-GNN has cross-domain generalization ability.

Future research directions entail: (i) Augmenting the credibility of MLFF-GNN for pneumonia CXR image classification by integrating multi-modal data information onto the foundation of utilizing pre-trained CNN for feature extraction. (ii) Enhancing the generalization capability of MLFF-GNN by refining its performance on alternative pneumonia datasets and potentially diverse medical image datasets.

REFERENCES

[1] Ahmed M S, Rahman A, AlGhamdi F, et al. Joint Diagnosis of Pneumonia, COVID-19, and Tuberculosis from Chest X-ray Images: A Deep Learning Approach [J]. Diagnostics, 2023, 13(15): 2562. doi: 10.3390/diagnostics13152562.

[2] Meng Y, Bridge J, Addison C, et al. Bilateral adaptive graph convolutional network on CT based Covid-19 diagnosis with uncertainty-aware consensus-assisted multiple instance learning [J]. Medical Image Analysis, 2023, 84: 102722. doi: 10.1016/j.media.2022.102722.

[3] Jie C, Jiming C, Ying S, et al. A pyramid GNN model for CXR-based COVID-19 classification [J]. The Journal of Supercomputing, 2024, 80(4): 5490-5508. doi: 10.1007/s11227-023-05633-1.

[4] Zhang W, Kong L, Lee S, et al. Detecting mental and physical disorders using multi-task learning equipped with knowledge graph attention network [J]. Artificial Intelligence in Medicine, 2024, 149: 102812. doi: 10.1016/j.artmed.2024.102812.

[5] Cui H, Ju F, Li J. Joint Multi-view Feature Network for Automatic Diagnosis of Pneumonia with CT Images [C]//International Conference on Frontier Computing. Springer, 2023: 169-180. doi: 10.1007/978-981-99-9342-0_19.

[6] Wang H, Huang G, Zhao Z, et al. CCF-GNN: A Unified Model Aggregating Appearance, Microenvironment, and Topology for Pathology Image Classification [J]. IEEE Transactions on Medical Imaging, 2023, 42(11): 3179-3193. doi: 10.1109/TMI.2023.3249343.

[7] Gilmer J, Schoenholz S S, Riley P F, et al. Neural Message Passing for Quantum Chemistry [C]//Precup D, Teh Y W. Proceedings of the 34th International Conference on Machine Learning: Vol. 70. PMLR, 2017: 1263-1272. https://proceedings.mlr.press/v70/gilmer17a.html.

[8] Shen W, Wei Z, Ren Q, et al. Rotation-Equivariant Quaternion Neural Networks for 3D Point Cloud Processing [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2024: 1-17. doi: 10.1109/TPAMI.2023.3346383.

[9] Luo T, Liu Y, Pan S J. Collaborative Sequential Recommendations via Multi-View GNN-Transformers [J]. ACM Trans. Inf. Syst., 2024. https://doi.org/10.1145/3649436. doi: 10.1145/3649436.

[10] Chatzianastasis M, Lutzeyer J, Dasoulas G, et al. Graph Ordering Attention Networks [J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2023, 37(6): 7006-7014. doi: 10.1609/aaai.v37i6.25856.

[11] Bai Y, Mi Y, Su Y, et al. A Scalable Graph-Based Framework for Multi-Organ Histology Image Classification [J]. IEEE Journal of Biomedical and Health Informatics, 2022, 26(11): 5506-5517. doi: 10.1109/JBHI.2022.3199110.

[12] He K, Zhang X, Ren S, et al. Deep Residual Learning for Image Recognition [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV, USA: IEEE, 2016: 770-778. http://ieeexplore.ieee.org/document/7780459/. doi: 10.1109/CVPR.2016.90.

[13] Li Z, Zheng Y, Shan D, et al. ScribFormer: Transformer Makes CNN Work Better for Scribble-based Medical Image Segmentation [J]. IEEE Transactions on Medical Imaging, 2024: 1-1. doi: 10.1109/TMI.2024.3363190.

[14] Pati P, Jaume G, Fernandes L A, et al. HACT-Net: A Hierarchical Cell-to-Tissue Graph Neural Network for Histopathological Image Classification [C]//Sudre C H, Fehri H, Arbel T, et al. Uncertainty for Safe Utilization of Machine Learning in Medical Imaging, and Graphs in Biomedical Image Analysis. Cham: Springer International Publishing, 2020: 208-219. doi: 10.1007/ 978-3-030-60365-6_20.

[15] Lei Z, Wang Y, Li Z, et al. Attention based multilayer feature fusion convolutional neural network for unsupervised monocular depth estimation [J]. Neurocomputing, 2021, 423: 343-352. doi: 10.1016/j.neucom.2020.11.002.

[16] Wang L, He K, Feng X, et al. Multilayer feature fusion with parallel convolutional block for fine-grained image classification [J]. Applied Intelligence, 2022, 52(3): 2872-2883. doi: 10.1007/s10489-021-02573-2.

[17] C.K. S, C.D. J, Patil N. Tomato plant disease classification using Multilevel Feature Fusion with adaptive channel spatial and pixel attention mechanism [J]. Expert Systems with Applications, 2023, 228: 120381. doi: 10.1016/j.eswa.2023.120381.

[18] Deng J, Dong W, Socher R, et al. Imagenet: A large-scale hierarchical image database [C]//2009 IEEE conference on computer vision and pattern recognition. Ieee, 2009: 248-255. doi: 10.1109/CVPR.2009.5206848.

[19] Yu X, Lu S, Guo L, et al. ResGNet-C: A graph convolutional neural network for detection of COVID-19 [J]. Neurocomputing, 2021, 452: 592-605. doi: 10.1016/j.neucom.2020.07.144.

[20] Raval D, Undavia J N. A Comprehensive assessment of Convolutional Neural Networks for skin and oral cancer detection using medical images [J]. Healthcare Analytics, 2023, 3: 100199. doi: 10.1016/j.health.2023.100199.

[21] Huang G, Liu Z, Van Der Maaten L, et al. Densely Connected Convolutional Networks [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, HI: IEEE, 2017: 2261-2269. https://ieeexplore.ieee.org/document/8099726/. doi: 10.1109/CVPR.2017.243.

[22] Rajeashwari S, Arunesh K. Enhancing pneumonia diagnosis with ensemble-modified classifier and transfer learning in deep-CNN based classification of chest radiographs [J]. Biomedical Signal Processing and Control, 2024, 93: 106130. doi: 10.1016/j.bspc.2024.106130.

[23] Hochreiter S, Schmidhuber J. Long Short-Term Memory [J]. Neural Computation, 1997, 9(8): 1735-1780. doi: 10.1162/neco.1997.9.8.1735.

[24] Chutia U, Tewari A S, Singh J P, et al. Classification of Lung Diseases Using an Attention-Based Modified DenseNet Model [J]. Journal of Imaging Informatics in Medicine, 2024. https://link.springer.com/10.1007/s10278-024-01005-0. doi: 10.1007/s10278-024-01005-0.

[25] Liu Z, Mao H, Wu C Y, et al. A ConvNet for the 2020s [C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2022: 11966-11976. doi: 10.1109/CVPR52688.2022.01167.

[26] Brody S, Alon U, Yahav E. How Attentive are Graph Attention Networks? [C]//The Tenth International Conference on Learning Representations, ICLR 2022, Virtual Event, April 25-29, 2022. OpenReview.net, 2022. https://openreview.net/forum?id=F72ximsx7C1.

[27] Velickovic P, Cucurull G, Casanova A, et al. Graph Attention Networks [C]//6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings. OpenReview.net, 2018. https://openreview.net/forum?id=rJXMpikCZ.

[28] Flecker B, Alford W, Beggs J M, et al. Partial information decomposition as a spatiotemporal filter [J]. Chaos, 2011, 21 3: 037104. doi: 10.1063/1.3638449.

[29] Wen L, Chen S, Xie M, et al. Training multi-source domain adaptation network by mutual information estimation and minimization [J]. Neural Networks, 2024, 171: 353-361. doi: 10.1016/j.neunet.2023.12.022.

[30] Foret P, Kleiner A, Mobahi H, et al. Sharpness-aware Minimization for Efficiently Improving Generalization [C]//9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021. OpenReview.net, 2021. https://openreview.net/forum?id=6Tm1mposlrM.

[31] Kermany D S, Goldbaum M, Cai W, et al. Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning [J]. Cell, 2018, 172(5): 1122-1131.e9. doi: 10.1016/j.cell.2018.02.010.

[32] Ayan E, Unver H M. Diagnosis of Pneumonia from Chest X-Ray Images Using Deep Learning [C]//2019 Scientific Meeting on Electrical-Electronics & Biomedical Engineering and Computer Science (EBBT). Istanbul, Turkey: IEEE, 2019: 1-5. https://ieeexplore.ieee.org/document/8741582/. doi: 10.1109/EBBT.2019.8741582.

[33] Ali M, Shahroz M, Akram U, et al. Pneumonia Detection Using Chest Radiographs With Novel EfficientNetV2L Model [J]. IEEE Access, 2024, 12: 34691-34707. doi: 10.1109/ACCESS.2024.3372588.

[34] Kipf T N, Welling M. Semi-Supervised Classification with Graph Convolutional Networks [C]//5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings. OpenReview.net, 2017. https://openreview.net/forum?id=SJU4ayYgl.

[35] Acharya V, Choi D, Yener B, et al. Prediction of Tuberculosis From Lung Tissue Images of Diversity Outbred Mice Using Jump Knowledge Based Cell Graph Neural Network [J]. IEEE Access, 2024, 12: 17164-17194. doi: 10.1109/ACCESS.2024.3359989.

[36] Nalluri S, Sasikala R. Pneumonia screening on chest X-rays with optimized ensemble model [J]. Expert Systems with Applications, 2024, 242: 122705. doi: https://doi.org/10.1016/j.eswa.2023.122705.

[37] Bertl M, Bignoumba N, Ross P, et al. Evaluation of deep learning-based depression detection using medical claims data [J]. Artificial Intelligence in Medicine, 2024, 147: 102745. doi: https://doi.org/10.1016/j.artmed.2023.102745.

[38] Yue W, Liu S, Li Y. Eff-PCNet: An Efficient Pure CNN Network for Medical Image Classification [J]. Applied Sciences, 2023, 13(16): 9226. doi: 10.3390/app13169226.

[39] Pan X, Ye T, Xia Z, et al. Slide-Transformer: Hierarchical Vision Transformer with Local Self-Attention [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023: 2082-2091.

[40] Xing L, Liu W, Liu X, et al. An Enhanced Vision Transformer Model in Digital Twins Powered Internet of Medical Things for Pneumonia Diagnosis [J]. IEEE Journal on Selected Areas in Communications, 2023, 41(11): 3677-3689. doi: 10.1109/JSAC.2023.3310096.

[41] Chen S, Xie E, Ge C, et al. CycleMLP: A MLP-Like Architecture for Dense Visual Predictions [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(12): 14284-14300. doi: 10.1109/TPAMI.2023.3303397.

[42] Zhang R, Wang L, Cheng S, et al. MLP-based classification of COVID-19 and skin diseases [J]. Expert Systems with Applications, 2023, 228: 120389. doi: 10.1016/j.eswa.2023.120389.

[43] Wang B, Pan H, Aboah A, et al. GazeGNN: A Gaze-Guided Graph Neural Network for Chest X-Ray Classification [C]//Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). 2024: 2194-2203.

[44] Chouhan V, Singh S K, Khamparia A, et al. A Novel Transfer Learning Based Approach for Pneumonia Detection in Chest X-ray Images [J]. Applied Sciences, 2020, 10(2): 559. doi: 10.3390/app10020559.

[45] Ukwuoma C C, Qin Z, Belal Bin Heyat M, et al. A hybrid explainable ensemble transformer encoder for pneumonia identification from chest X-ray images [J]. Journal of Advanced Research, 2023, 48: 191-211. doi: 10.1016/j.jare.2022.08.021.