Journal of Computer Science and Cybernetics, V41, Nx (2025), 1–
DOI: 10.15625/1813-9663/23278

FSGCN: FEATURE-CONNECTED GRAPH NEURAL NETWORK FOR FINE-GRAINED SILICOSIS CLASSIFICATION

NGUYEN THI TAN TIEN[1], BUI QUOC BAO[2], PHAM VAN CUONG[3], TRAN TIEN CONG[3], LE MINH DUY[3], NGUYEN VAN TAO[4]

[1] *Thai Nguyen University of Medicine and Pharmacy, 284 Luong Ngoc Quyen Street, Phan Dinh Phung Ward, Thai Nguyen Province, Viet Nam*
[2] *Hanoi University of Science and Technology, No.1 Dai Co Viet, Bach Mai Ward, Ha Noi, Viet Nam*
[3] *Posts and Telecommunications Institute of Technology, 96A Tran Phu, Ha Dong Ward, Ha Noi, Viet Nam*
[4] *Thai Nguyen University of Information and Communication Technology, Z115 Road, Quyet Thang Ward, Thai Nguyen Province, Viet Nam*

Abstract. Fine-grained classification of pulmonary diseases remains a challenging task due to subtle inter-class variations and overlapping visual patterns in Chest X-ray imaging. In this work, we propose FSGCN (Fine-grained silicosis graph-based classification network), a novel hybrid architecture that combines convolutional representation learning with relational reasoning via a graph transformer network (GTN). Specifically, image features extracted from a deep encoder are treated as nodes in a fully connected graph, where edge relationships are dynamically learned to capture semantic correlations among instances within a batch. This graph-based interaction enables the model to better distinguish silicosis from other visually similar pulmonary conditions, such as viral and bacterial pneumonia. We evaluate the proposed approach on SVBCX, a curated multiclass chest X-ray dataset collected and annotated by our research team, comprising four categories: Normal, Silicosis, Viral, and Bacterial. The experimental results demonstrate the superiority of our method over existing baselines. On average, FSGCN achieves an absolute improvement of +1.51% in accuracy and +1.85% in F1-score compared to state-of-the-art baseline models. In addition, qualitative visualizations further confirm the effectiveness of the proposed model in enhancing class discriminability. These results highlight the importance of modeling inter-instance relationships for fine-grained disease classification in medical imaging.

Keywords. Silicosis diagnosis, fine-grained image classification, graph neural networks, chest X-ray analysis.

1. INTRODUCTION

Occupational diseases result directly from prolonged exposure to hazardous work environments, significantly impacting workers' health and quality of life. According to the

E-mail addresses: nguyenthitantien@tnmc.edu.vn (N.T.T. Tien); Bao.BuiQuocM@hust.edu.vn (B.Q. Bao); cuongpv@ptit.edu.vn (P.V. Cuong); congtt@ptit.edu.vn (T.T. Cong); duylm@ptit.edu.vn (L.M. Duy); nvtao@ictu.edu.vn (N.V Tao).

Vietnam General Confederation of Labor, by the end of 2023, approximately 33.000 occupational disease cases had been recorded nationwide, with about 7.000 new cases diagnosed annually and 500 cases assessed for social insurance benefits each year [1]. Among the 35 occupational diseases covered by insurance, silicosis has a notably high incidence and severely affects the respiratory system [2]. The disease progresses silently without clear symptoms initially. Still, it gradually leads to irreversible pulmonary fibrosis, chronic respiratory failure, reduced working capacity, decreased life expectancy, and even death if not promptly diagnosed and treated [3].

Currently, chest radiography remains the primary method for diagnosing and screening silicosis. However, this technique has several limitations, such as low sensitivity in early detection, possible oversight of small lesions, and considerable dependence on the expertise of physicians [4]. Additionally, radiographic signs of silicosis can easily be confused with other respiratory conditions such as tuberculosis, pneumonia, or chronic obstructive pulmonary disease, complicating diagnosis and treatment [5].

In recent years, artificial intelligence (AI), particularly deep learning, has emerged as a promising tool for automated medical imaging analysis. Several deep learning frameworks, such as CheXNet [6], DarkNet [7], and VinDr [8], have demonstrated performance comparable to radiologists in detecting thoracic diseases. Specific to silicosis, Priego-Torres et al. [8] proposed a deep learning–based system for automated staging, while Yi et al. [9] evaluated radiologists' and AI systems' ability to diagnose silicosis from HRCT scans. These studies confirm the potential of AI in this domain but also highlight persistent challenges: (i) the difficulty in differentiating silicosis from visually similar lung conditions (e.g., bacterial or viral pneumonia), (ii) the scarcity of large, high-quality annotated datasets for occupational silicosis, and (iii) limited generalization ability when models are trained on small or heterogeneous data.

Emerging research directions, such as graph neural networks (GNNs), have shown significant promise in capturing structured relational features from medical images [10–13]. GNNs can effectively exploit inter-sample dependencies and subtle local patterns, which are crucial for fine-grained disease classification but often overlooked by standard convolutional models. However, there is still a lack of dedicated research focusing on the application of GNNs for occupational silicosis diagnosis, especially in the context of developing countries like Viet Nam, where data imbalance and annotation scarcity remain pressing issues.

To address these challenges, this paper proposes FSGCN (fine-grained silicosis graph-based classification network), a novel framework that integrates convolutional feature extraction with graph-based relational reasoning to improve silicosis detection from chest X-rays.

In summary, the main contributions of this work are as follows:

- We introduce FSGCN, a novel hybrid architecture that integrates deep convolutional encoding with graph-based relational reasoning for fine-grained classification of thoracic diseases, particularly targeting silicosis detection.
- We design a graph-based network that dynamically learns semantic relationships among instances within each batch, enabling the model to leverage inter-sample context to enhance class discriminability.
- We conduct extensive experiments demonstrating that FSGCN outperforms standard convolutional baselines in both accuracy and F1-score, and provide qualitative visualizations validating the model's ability to localize and differentiate subtle pathological patterns.

The remainder of this paper is organized as follows: Section 2 reviews related work on clinical methods, machine learning, deep learning, and GNNs for medical imaging. Section 3 introduces the proposed FSGCN framework in detail. Section 4 presents the dataset, experimental setup, and evaluation results, followed by qualitative analysis. Section 5 concludes the paper and discusses future research directions.

2. RELATED WORK

2.1. Clinical diagnostic methods for occupational silicosis

Currently, the diagnosis of occupational silicosis mainly relies on occupational exposure history and standard chest X-ray imaging. According to the guidelines of the Ministry of Health, workers must have sufficient exposure to silica dust (typically one year or more) accompanied by characteristic signs of lung damage visible on chest X-rays, such as small opacities primarily distributed in the middle and lower regions of both lungs [14]. Clinical symptoms such as shortness of breath, cough, and chest pain often appear late; thus, the disease is typically discovered incidentally during routine health checks [15].

Chest X-ray, while widely accessible and simple, has limitations in sensitivity, particularly for small or early-stage lesions. Silicosis lesions can be mistaken for other diseases like tuberculosis, pneumonia, sarcoidosis, or metastatic lung cancer [14]. Computed tomography (CT), particularly high-resolution CT (HRCT), offers greater sensitivity than X-ray, aiding in the early detection of lesions and clearer differentiation from other pulmonary conditions. However, its higher cost and increased radiation exposure limit its use to necessary cases only.

2.2. Traditional machine learning methods

Before the emergence of deep learning models, traditional machine learning methods were widely applied in medical image analysis. Techniques such as decision trees, random forest, and multi-layer perceptron (MLP) typically rely on handcrafted feature extraction, including intensity, texture, and statistical features. For instance, Chandra et al. [16] used an MLP model, achieving 95.39% accuracy in pneumonia diagnosis. Kuo et al. [17] applied decision tree methods with 94.5% accuracy in classifying pneumonia among schizophrenic patients. However, these methods are limited when handling large datasets or addressing imbalanced data [18].

Beyond imaging, some studies have explored non-image-based predictive modeling. For example, Sanchez-Morillo et al. used machine learning algorithms (LASSO, SVM and Random Forest) to integrate routine blood biomarkers for silicosis diagnosis, offering a promising non-invasive diagnostic supplement [19].

Overall, while traditional ML methods provided early progress in computer-aided diagnosis, their reliance on handcrafted features limited robustness and generalizability. This is particularly problematic for occupational silicosis, where radiographic findings are often subtle, overlapping with other lung diseases, and where datasets are typically small and imbalanced. These shortcomings motivated the shift toward deep learning approaches, which can automatically learn higher-level representations from raw imaging data.

2.3. Deep learning methods

Deep learning, particularly convolutional neural networks (CNN), has significantly advanced medical image analysis due to its ability to extract features from raw data automatically. Studies like CheXNet, utilizing a 121-layer CNN, have achieved diagnostic performance comparable to expert radiologists [7]. The Darknet model was also successfully applied for COVID-19 diagnosis from chest X-ray images [20]. Transfer learning techniques have frequently been employed to improve training performance on small or imbalanced datasets, showing promising and stable outcomes [10, 21]. For silicosis, Sharma et al. applied CNNs on 706 chest X-rays and achieved 81% accuracy, while Priego-Torres et al. extended this approach to classify silicosis stages, employing a balanced loss to mitigate class imbalance. These works confirmed the feasibility of CNNs for silicosis screening under limited data conditions. Despite these encouraging advances, CNN-based methods may still have certain limitations. They often require large annotated datasets, may struggle to capture long-range spatial dependencies, and can overlook potential inter-sample relationships. These considerations suggest that exploring more advanced approaches, such as graph neural networks (GNNs), could provide a complementary way to model relational information and potentially improve silicosis diagnosis.

2.4. Application of graph neural networks in medical imaging

Graph neural networks (GNNs) have recently emerged as powerful tools in medical imaging, offering the ability to model non-Euclidean structures and capture spatial as well as relational dependencies beyond the capability of traditional convolutional neural networks (CNNs). Their flexibility has enabled advances in tasks such as image denoising, lesion segmentation, brain tumor classification, and patient prognosis across diverse imaging modalities, including MRI, CT, ultrasound, and histopathology [22–24].

Compared to CNNs, GNN-based approaches demonstrate advantages in handling irregular spatial patterns, modeling inter-regional dependencies, and integrating multimodal information. For instance, GNNs have shown competitive or superior performance in ultrasound lesion segmentation and oncology risk stratification, often with fewer parameters and improved interpretability [25, 26].

Several architectural innovations underpin these successes. Graph convolutional networks (GCNs) are widely employed for brain tumor detection and multimodal feature fusion [27], while graph attention networks (GATs) enhance segmentation and classification through adaptive neighbor weighting [28]. Dual-branch and multimodal GNNs, which combine CNNs for local features with GNNs for global context, further enhance diagnostic accuracy [29]. Additionally, hierarchical GNNs enable multi-scale analysis, a critical feature for histopathology tisue [30].

Despite rapid progress, challenges remain regarding interpretability, data scarcity, and robust graph construction. Ongoing research is addressing these gaps through adaptive graph learning, scalable hierarchical models, and benchmarking frameworks, aiming to improve the clinical applicability and trustworthiness of GNNs in medical imaging [31, 32].

3. THE FSGCN FRAMEWORK: A GRAPH-ENHANCED MODEL FOR SILICOSIS CLASSIFICATION

3.1. Fine-grained silicosis classification problem

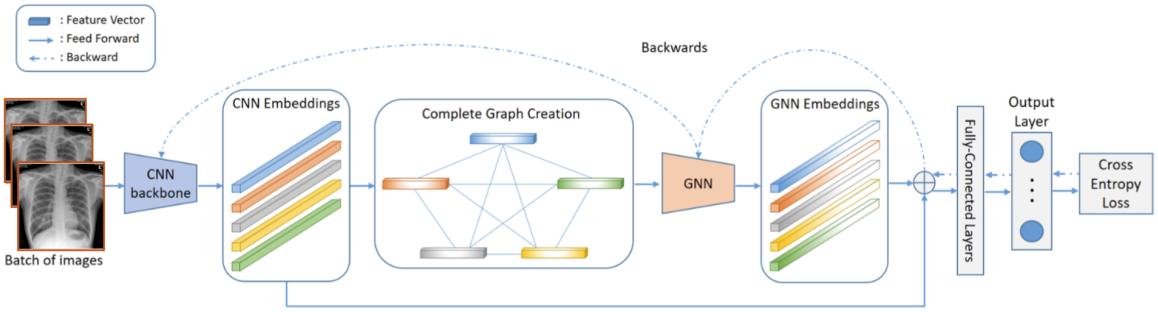

Figure 1: Overall architecture of the proposed fine-grained silicosis graph-based classification network (FSGCN). A batch of chest X-ray images is first processed by a CNN backbone to extract latent embeddings. These embeddings are then represented as nodes in a complete graph, where edges encode pairwise relationships between samples. The constructed graph is passed through a graph neural network (GNN) module to capture inter-instance dependencies. The refined GNN embeddings are subsequently fed into fully connected layers and an output layer for classification, optimized using cross-entropy loss.

Fine-grained image classification refers to the task of distinguishing among subordinate categories that share a common high-level semantic class but differ by subtle, localized visual cues. Let $\mathcal{X} \subset \mathbb{R}^{3\times H\times W}$ denote the input image space, where each sample $x \in \mathcal{X}$ corresponds to a Chest X-ray image of dimension $H \times W$ with three channels. The corresponding label space is defined as $\mathcal{Y} \subset \{0,1\}^C$, where each label $y \in \mathcal{Y}$ is a one-hot vector representing one of the C fine-grained classes under a common meta-category.

Given a training set

$$\mathcal{T} = \{(x_i, y_i)\}_{i=1}^{N} \subset \mathcal{X} \times \mathcal{Y}, \tag{1}$$

sampled from an unknown joint distribution $\mathcal{P}(x, y)$, the learning objective is to estimate a classification function $f_\theta : \mathcal{X} \to \mathcal{Y}$ parameterized by θ, such that the empirical risk is minimized

$$\mathcal{L}(\theta) = \frac{1}{N}\sum_{i=1}^{N} \ell(f_\theta(x_i), y_i), \tag{2}$$

where $\ell(\cdot,\cdot)$ is a suitable classification loss function (e.g., cross-entropy loss).

Silicosis-related pulmonary disease classification. In the specific context of this work, we define the label space as

$$\mathcal{Y} = \{y^{(1)}, y^{(2)}, y^{(3)}, y^{(4)}\} \tag{3}$$

$$= (\text{Normal}, \text{Silicosis}, \text{Viral}, \text{Bacterial}). \tag{4}$$

The discriminative features required to differentiate between classes particularly between Silicosis and other pathologies are often highly localized and of low inter-class variance, i.e.,

$$\|\phi(x_i) - \phi(x_j)\|_2 \approx 0, \quad \text{for some } i \neq j \text{ with } y_i \neq y_j, \tag{5}$$

Here, $\phi : \mathcal{X} \to \mathbb{R}^d$ denotes the latent feature extractor. This high degree of visual similarity imposes a significant challenge for conventional CNN-based classifiers, which typically rely on global feature pooling and may fail to capture fine-grained local variations.

To address this challenge, we propose a graph-based classification framework in which a set of discriminative feature nodes is constructed and explicitly connected based on spatial and semantic affinity. The model learns to propagate and aggregate information over the constructed graph structure, allowing for enhanced representation learning that is sensitive to subtle visual distinctions.

3.2. The FSGCN network architecture

Algorithm 1: Forward Pass of FSGCN

Input: Mini-batch of training samples $\{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^{b}$
Output: Final image-level representations $\{\mathbf{v}_i\}_{i=1}^{b}$

Step 1: DNN encoder
for $i = 1$ **to** b **do**
  $\mathbf{c}_i \leftarrow \texttt{DNNEncoder}(\mathbf{x}_i)$; // Extract image embedding
end
Step 2: Construct fully connected graph
$\mathcal{V} \leftarrow \{\mathbf{c}_i\}_{i=1}^{b}$; // Node features from DNN encoder
$\mathcal{E} \leftarrow \{\mathbf{e}_{ij}\}_{i,j=1}^{b}$; // Learnable edge weights
$\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{F} = \mathcal{V})$; // Graph construction
Step 3: GTN graph encoding for $i = 1$ **to** b **do**
  $\hat{\mathbf{c}}_i \leftarrow \sum_{j \in \mathcal{N}(i)} \alpha_{i,j} \left(\mathbf{W}_v \mathbf{c}_j + \mathbf{W}_e \mathbf{e}_{ij}\right)$; // Graph attention aggregation
  $\mathbf{r}_i \leftarrow \mathbf{W}_r \mathbf{c}_i$; // Residual connection
  $\beta_i \leftarrow \sigma\left(\mathbf{W}_g \left[\hat{\mathbf{c}}_i; \mathbf{r}_i; \hat{\mathbf{c}}_i - \mathbf{r}_i\right]\right)$; // Gated fusion weight
  $\mathbf{c'}_i \leftarrow \beta_i \cdot \mathbf{r}_i + (1 - \beta_i) \cdot \hat{\mathbf{c}}_i$; // Final node embedding
end
Step 4: Output projection for $i = 1$ **to** b **do**
  $\mathbf{v}_i \leftarrow \texttt{BatchNorm}(\texttt{Linear}(\mathbf{c'}_i))$; // Normalized output representation
end
return $\{\mathbf{v}_i\}_{i=1}^{b}$

The forward propagation procedure of the proposed FSGCN (fine-grained silicosis graph-based classification network), detailed in Algorithm 1, encompasses a hybrid architecture integrating deep convolutional feature extraction with graph-based representation learning. This architecture in Figure 1 is specifically designed to capture fine-grained, relational features among instances in a mini-batch of Chest X-ray images.

The process begins with the DNN encoder module, wherein each input image $\mathbf{x}_i$ from a training batch of size b is independently processed through a convolutional backbone (e.g., ResNet or DenseNet) to yield a corresponding latent feature vector $\mathbf{c}_i \in \mathbb{R}^d$. These vectors serve as node features for the subsequent graph-based reasoning component.

Next, a fully connected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is instantiated, where each node $v_i \in \mathcal{V}$ represents an image feature $\mathbf{c}_i$, and edges $\mathbf{e}_{ij}$ encode pairwise relationships between image instances. Notably, the edge features are initialized as learnable parameters and dynamically updated during training, allowing the model to adaptively infer context-sensitive affinities

between instances.

The GTN (graph transformer network) encoder then performs a relational aggregation step for each node. Specifically, for a given node i, the model aggregates information from all other nodes $j \in \mathcal{N}(i)$ via an attention-weighted message passing mechanism. The message from node j to node i incorporates both the neighbor's feature $\mathbf{c}_j$ and the associated edge embedding $\mathbf{e}_{ij}$, transformed respectively by $\mathbf{W}_v$ and $\mathbf{W}_e$. Attention weights $\alpha_{i,j}$ are computed using scaled dot-product attention between the query vector of node i and a joint key vector composed of node j's features and the edge embedding. This enables the model to selectively emphasize the most relevant nodes and edge connections in the graph.

To preserve semantic integrity and mitigate the risk of over-smoothing (a common issue in deep graph networks), the aggregated representation $\hat{\mathbf{c}}_i$ is combined with the original node embedding $\mathbf{c}_i$ through a gated residual fusion mechanism. A gating coefficient $\beta_i \in (0, 1)$, computed via a sigmoid function over concatenated inputs, adaptively balances the influence of new messages and original features. The fused representation $\mathbf{c'}_i$ thus retains essential local discriminative information while integrating global relational context.

Finally, the updated node embedding $\mathbf{c'}_i$ is projected through a fully connected layer followed by Batch Normalization to produce the final output vector $\mathbf{v}_i$. These representations $\{\mathbf{v}_i\}_{i=1}^{b}$ can then be used for downstream classification via a softmax output layer or further regularization techniques, depending on the learning objective.

This graph-based architecture enables the model to jointly reason over intra-batch feature interactions, thus enhancing its ability to detect subtle inter-class variations, crucial for accurate fine-grained classification of silicosis and other pulmonary conditions that share overlapping imaging characteristics.

3.3. Loss function

To optimize the FSGCN model for the fine-grained silicosis classification task, we adopt the standard categorical cross-entropy loss, which is widely used in multi-class classification settings. Given a predicted probability vector $\hat{\mathbf{y}}_i$ and the corresponding one-hot ground truth label $\mathbf{y}_i$ for each training sample, the loss function is computed as

$$\mathcal{L}_{\text{CE}} = -\frac{1}{b} \sum_{i=1}^{b} \sum_{c=1}^{C} y_i^{(c)} \log \hat{y}_i^{(c)}, \tag{6}$$

where b is the batch size and C is the total number of fine-grained categories. This loss encourages the model to assign high probability to the correct class and penalizes incorrect predictions accordingly. The entire FSGCN architecture is trained in an end-to-end manner using this objective.

4. EXPERIMENTAL RESULTS

4.1. The SVBCX dataset

In this study, we utilize the SVBCX dataset, a curated and ethically approved collection of chest X-ray (CXR) images designed to support research on pulmonary disease classification, with a specific focus on silicosis. The dataset includes both healthy and pathological cases and

Table 1: Comparison of classification accuracy and F1-score among baseline models and their FSGCN-enhanced counterparts.

Method	Inference time	# params	Acc (%)	F1-score
MobilenetV3-S	0.013	1.6M	90.12	80.25
MobilenetV3-S + FSGCN (ours)	0.016	17.4M	92.01(+1.89)	82.23(+1.98)
Densenet201	0.28	18.3M	91.55	81.2
Densenet201 + FSGCN (ours)	0.29	73.7M	93.72(+2.17)	84.48(+3.28)
ResNet-50	0.51	22.5M	93.75	83.27
ResNet-50 + FSGCN (ours)	0.52	91.7M	95.29(+1.54)	85.35(+2.08)
SwinT-Big	0.82	87.0M	96.11	85.96
SwinT-Big + FSGCN (ours)	0.84	102.8M	97.06(+0.95)	87.9(+1.94)
ConvNeXtLarge	1.22	197.9M	96.71	86.74
ConvNeXtLarge + FSGCN (ours)	1.23	231.8M	97.79(+1.08)	87.81(+1.07)
TransFG	0.85	86.3M	93.96	85.14
TransFG + FSGCN (ours)	0.86	95.2M	94.82(+0.86)	85.66(+0.52)
P2P-Net	0.46	63.4M	91.95	84.46
P2P-Net + FSGCN (ours)	0.48	110.6M	94.03(+2.08)	86.57(+2.11)
Avg. Improvement			+1.51	+1.85

is structured to enable fine-grained classification among four categories: Normal, Silicosis, Bacterial, and Viral.

Specifically, the dataset comprises 535 cases diagnosed with silicosis, 2772 cases of bacterial pneumonia, 1493 cases of viral pneumonia, and 1554 normal cases, amounting to a total of 6,354 chest X-ray images. The silicosis subset includes X-rays of workers exposed to silica dust, evaluated in accordance with the International labour organization (ILO) standards. All data collection procedures adhered strictly to research ethics and patient data confidentiality guidelines. The study was approved by the Ethics Committees of Hanoi Medical University and Thai Nguyen University of Medicine and Pharmacy. Participation was entirely voluntary, and all participants were informed of the study objectives and their rights, including the right to withdraw at any point.

The dataset was compiled from two primary sources:

- Domestic data source: The domestic dataset was obtained from the national research project KC.10.33/16-20, which collected chest X-ray images of workers exposed to silica dust across five provinces in Vietnam: Thai Nguyen, Hai Duong (North), Binh Dinh, Phu Yen (Central), and Dong Nai (South). All X-rays were assessed by certified radiologists and occupational health professionals.
- International data source: In addition to domestic data, we incorporated publicly available chest X-ray images from the well-known open-access dataset introduced by Kermany et al. and hosted on Mendeley Data [33]: To ensure sufficient representation of viral and bacterial pneumonia cases, additional CXR images were sourced from a well-established open-access dataset annotated by medical professionals. These images were used in both training and testing phases, and the dataset remains accessible via Mendeley Data [33].

The SVBCX dataset thus provides a comprehensive and diverse foundation for evaluating fine-grained classification methods in pulmonary imaging, with particular relevance to occupational lung diseases.

4.2. Experimental setup

Training configuration. All experiments were conducted on a workstation equipped with an NVIDIA RTX 3090 GPU (24GB VRAM), an Intel Core i9 CPU, and 128 GB RAM, running Ubuntu 20.04 with PyTorch 1.13 and CUDA 11.6. The input images were resized to 224×224, and data augmentation techniques such as random horizontal flipping, cropping, and brightness adjustment were applied during training to enhance model generalization.

We used the Adam optimizer with a starting learning rate of 1e-4, decayed by a factor of 0.1 every 10 epochs. The batch size was set to 32, and training was performed for 50 epochs. For classification, we used the cross-entropy loss function. All models were trained using the same settings to ensure fair comparison.

Baseline models. To validate the effectiveness and generalizability of our proposed fine-grained silicosis graph-based classification Network (FSGCN), we benchmark it against a diverse set of representative baseline architectures, spanning from lightweight CNNs to advanced transformer-based and fine-grained visual recognition models:

- CNN-based models: *MobileNetV3-Small* [34], *DenseNet201* [35], and *ResNet-50* [36].
- Transformer-based and hybrid models: *Swin Transformer* [37], *ConvNeXt-Large* [38], *TransFG* [39], and *P2P-Net* [40].

Performance is reported using both accuracy and F1-score, which jointly capture the model's predictive quality, particularly under class imbalance.

4.3. Comparison with baseline models

Table 1 presents a comprehensive comparison between our proposed FSGCN-enhanced framework and several representative baseline models across multiple backbone architectures. We evaluate each model based on accuracy, F1-score, inference time, and number of parameters.

Across all backbones from lightweight CNNs like MobileNetV3-S to large-scale transformer-based models such as ConvNeXtLarge and SwinT-Big, the integration of FSGCN consistently improves performance. On average, the inclusion of FSGCN leads to a +1.51% increase in accuracy and a +1.85 improvement in F1-score, highlighting its effectiveness in boosting both classification precision and recall, particularly for fine-grained silicosis categories.

Specifically, on the lightweight MobileNetV3-S, FSGCN improves accuracy from 90.12% to 92.01% and F1-score from 80.25 to 82.23, with only a modest increase in inference time (from 0.013s to 0.016s). This demonstrates that the proposed approach maintains efficiency while significantly enhancing feature discrimination, even in resource-constrained settings.

Among heavier backbones, Densenet201+FSGCN achieves a notable +2.17% accuracy and +3.28 F1-score gain compared to its original counterpart. Similarly, ResNet-50 sees a +2.08 improvement in F1-score, which confirms that FSGCN complements existing deep architectures by exploiting feature-wise relationships beyond what convolutional layers can capture.

Transformer-based baselines such as SwinT-Big and ConvNeXtLarge already achieve strong performance; however, FSGCN still brings additional gains, e.g., SwinT-Big's F1-score increases from 85.96 to 87.9, while ConvNeXtLarge reaches the highest accuracy (97.79%) and F1-score (87.81) overall. Importantly, the added inference time remains negligible.

Table 2: Comparison of FSGCN model performance under sequential and shuffled data sampling during evaluation (the reported results represent classification accuracy, measured in %).

Method	Stanford Dogs	
	Sequential	Shuffle
MobilenetV3-S + FSGCN	92.01	92.17 ± 0.21
Densenet201 + FSGCN	93.72	93.20 ± 0.15
ResNet-50 + FSGCN	95.29	95.27 ± 0.08
SwinT-Big + FSGCN	97.06	96.97 ± 0.13
ConvNeXtLarge + FSGCN	97.79	97.27 ± 0.10
TransFG + FSGCN	94.82	94.97 ± 0.16
P2P-Net + FSGCN	94.03	93.52 ± 0.12

Finally, both TransFG and P2P-Net, which are designed specifically for fine-grained visual tasks, also benefit from our method. P2P-Net+FSGCN shows a remarkable +2.08% accuracy and +2.11 F1-score gain, underscoring the adaptability and broad applicability of FSGCN across both general and fine-grained classification architectures.

These results collectively affirm that the FSGCN module serves as a powerful, plug-and-play enhancement, yielding consistent performance improvements while maintaining scalability across a variety of model sizes and families.

4.4. The impact of batch configurations

Shuffling the validation dataset during evaluation. Unlike conventional deep learning models, the proposed FSGCN architecture leverages a fully connected graph structure to refine image representations based on inter-sample relationships within a batch. This design makes the model potentially sensitive to the intra-batch sample diversity, especially during evaluation when graph edges are constructed without label supervision. Therefore, it is essential to examine the stability of the model under different batch sampling strategies.

To this end, we compare the performance of several FSGCN-based models on the validation set under two batch construction scenarios:

- Sequential sampling, where data is grouped by class labels (i.e., samples from the same class are batched together), resulting in low variation within batches;
- Shuffled sampling, where data is randomly sampled across all classes, forming high intra-batch variation, a more realistic evaluation setting.

Table 2 presents the comparative results for models including MobileNetV3-S, DenseNet201, ResNet-50, SwinT-Big, ConvNeXtLarge, TransFG, and P2P-Net, all equipped with the FSGCN module. We observe that:

- The performance differences between the two sampling modes are generally small, with variations mostly within $\pm 0.5\%$, indicating that FSGCN is robust against batch shuffling.
- For certain models, such as MobileNetV3-S and TransFG, the shuffled setting even leads to slightly better results, likely due to the richer inter-class relational cues within each batch.
- Some high-capacity models, such as ConvNeXtLarge and P2P-Net, experience marginal performance drops (e.g., -0.52 and -0.51 in accuracy, respectively), which may be attributed to overfitting to structured patterns in sequential batches.

Table 3: Performance comparison for GPHs using various batch sizes (the reported results represent classification accuracy, measured in %).

	1	4	8	16	32	48	64	120
MobilenetV3-S + FSGCN	87.54	89.57	91.22	91.54	92.01	92.11	92.32	92.61
Densenet201 + FSGCN	92.19	92.35	92.54	92.67	92.72	92.94	92.86	-
TransFG + FSGCN	92.33	93.05	93.44	94.20	94.82	94.96	-	-
P2P-Net + FSGCN	92.25	93.47	93.68	93.76	94.03	94.09	-	-

Overall, these results demonstrate that while FSGCN's performance can be influenced by intra-batch diversity, it remains remarkably stable across different batch configurations. This suggests that our method is suitable for real-world applications where evaluation data is typically presented in a shuffled manner.

Batch size. In our framework, the model constructs a graph using both the inference image and a batch of support images, enabling it to capture pairwise feature relationships for fine-grained classification. This design inherently depends on batch-level relational information, motivating an analysis of how batch size impacts model performance.

As shown in Table 3, we observe a consistent trend across all four backbones: the accuracy improves significantly when increasing the batch size from 1 to 4 and 8. For instance, with MobileNetV3-S + FSGCN, accuracy rises sharply from 87.54% (batch = 1) to 91.22% (batch = 8), while TransFG + FSGCN improves from 92.33% to 93.44% over the same range. This early-stage improvement indicates that a relatively small number of contextual samples is sufficient to enrich the graph-based reasoning module.

Beyond batch size 8, the performance increases more gradually, suggesting that the model becomes more stable when the relational context is sufficiently represented. In many cases, the accuracy plateaus after batch size 32. This observation suggests that using a batch size of 32 is a reliable choice that balances both performance and computational efficiency.

It is also worth noting that MobileNetV3-S, due to its lightweight architecture, allows experiments to be run with very large batch sizes (up to 120) without memory constraints. However, for larger models like DenseNet201, TransFG, and P2P-Net, we were unable to test with the largest batch sizes due to GPU memory limitations, hence the missing values at the end of those rows in Table 3.

These results demonstrate the importance of batch-wise relational modeling in our architecture and highlight the practical considerations when selecting batch size for deployment. To further complement Table 3, Figure 2 provides a visual comparison of model performance across different batch sizes. The chart clearly illustrates the sharp initial improvements, the gradual stabilization beyond batch size 8, and the performance plateau after batch size 32, offering an intuitive understanding of the observed trends.

4.5. Qualitative analysis

Feature distribution via t-SNE. To further explore the discriminative capacity of the learned representations, we visualize the high-dimensional image features extracted by the final FSGCN layer using t-SNE, a non-linear dimensionality reduction technique. As illustrated in Figure 3, the samples belonging to four categories form well-separated clusters in the 2D projection space, with only slight overlap between Silicosis and Bacterial cases, an expected ambiguity given their clinical similarity in chest X-rays.

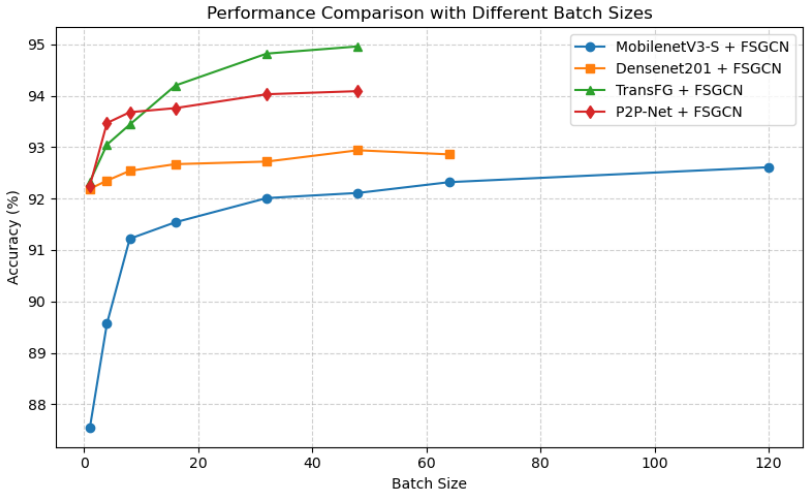

Figure 2: Accuracy (%) comparison of different FSGCN-based models across varying batch sizes. The chart illustrates the performance trends of MobileNetV3-S, DenseNet201, TransFG, and P2P-Net when integrated with the proposed FSGCN framework.

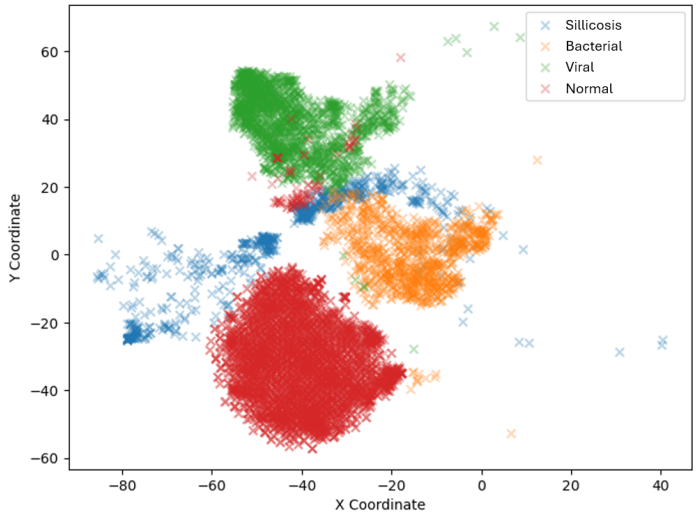

Figure 3: t-SNE visualization of the learned feature representations extracted from the final layer of the FSGCN model.

The compactness of intra-class clusters and the clear inter-class boundaries, especially for the Normal and Viral categories, indicate that FSGCN learns highly discriminative and structured feature representations. This structured embedding space is crucial for fine-grained classification tasks where subtle visual cues distinguish between categories.

Overall, the t-SNE plot provides strong qualitative evidence that the relational modeling in FSGCN enhances the network's ability to organize feature spaces in a semantically meaningful way, facilitating more reliable and interpretable decision-making.

Attention visualization via grad-CAM. To qualitatively assess the discriminative capability and interpretability of the proposed FSGCN model, we present Grad-CAM visualizations of representative test samples across four categories: Silicosis, Viral pneumonia,

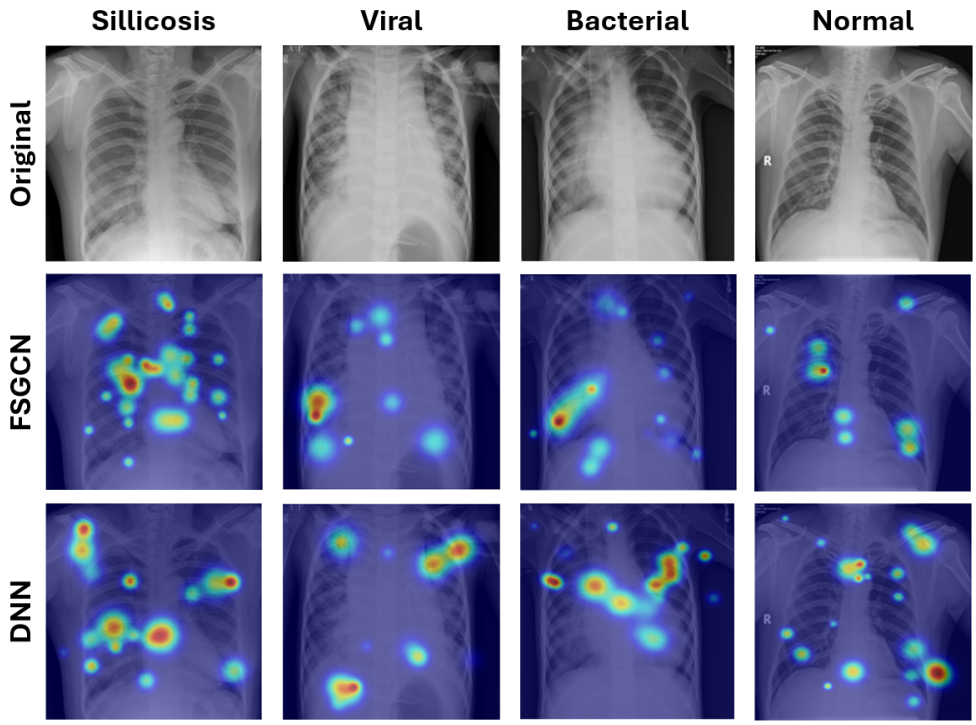

Figure 4: Visual comparison of Grad-CAM heatmaps between the proposed MobileNetV3-S + FSGCN model (middle row) and baseline MobileNetV3-S (bottom row) across four classes: Silicosis, Viral, Bacterial, and Normal (top row: original chest X-ray images).

Bacterial pneumonia, and Normal, as illustrated in 4. The visual comparison is made between the FSGCN-enhanced MobileNetV3-S (middle row) and the baseline MobileNetV3-S (bottom row), with the corresponding original chest X-ray images shown on the top row.

For the three disease classes (Silicosis, Viral, and Bacterial), we observe that FSGCN consistently produces more concentrated and semantically meaningful activation regions, often aligning with clinically relevant pathological areas such as localized opacities, interstitial patterns, or consolidation zones. These focused heatmaps suggest that the graph-based reasoning module in FSGCN helps the model to attend to subtle and spatially distributed features, a hallmark of fine-grained pulmonary conditions, more effectively than the standard DNN baseline. In contrast, the baseline model tends to generate more scattered or diffuse activations, sometimes highlighting irrelevant regions, which could lead to reduced diagnostic precision or increased inter-class confusion.

For the Normal class, the FSGCN model yields minimal and spatially neutral activations, reflecting high model confidence and lower false-positive sensitivity. Meanwhile, the baseline model occasionally exhibits unnecessary activations, indicating uncertainty or overfitting to noise.

These qualitative results substantiate the effectiveness of FSGCN in learning richer, more interpretable features by integrating inter-sample relationships, ultimately enhancing its fine-grained classification performance in thoracic disease diagnosis.

5. CONCLUSION

This paper introduces FSGCN, a hybrid deep learning framework that integrates convolutional encoding with graph-based relational reasoning for fine-grained classification of pulmonary diseases. Evaluated on the curated SVBCX dataset, FSGCN demonstrates improved accuracy and F1-score over baseline models, particularly in distinguishing silicosis from other similar conditions.

While the results are promising, one limitation lies in the class imbalance within the dataset, which may affect the model's sensitivity to minority classes. Moreover, the current framework does not incorporate clinical metadata, such as patient history or exposure duration, which could provide valuable complementary information. Future work will focus on addressing data imbalance, integrating multi-modal clinical inputs, and evaluating the model on larger, more diverse datasets to improve robustness and generalizability.

REFERENCES

[1] Vietnam Ministry of Health, "7,000 workers diagnosed with occupational diseases annually," https://moh.gov.vn/, 2024, accessed on April 7, 2025.

[2] V. Nguyen, "Epidemiological and molecular epidemiological characteristics of silicosis in workers from several provinces in Vietnam," Ph.D. dissertation, Hanoi Medical University, 2023.

[3] D. V. Ham, *Occupational Hygiene and Occupational Diseases*. Ha Noi: Medical Publishing House, 2007.

[4] E. K. Austin, C. James, J. J. Tessier, and P. Health, "Early detection methods for silicosis in australia and internationally: a review of the literature," *International Journal of Environmental Research and Public Health*, vol. 18, no. 15, p. 8123, 2021.

[5] A. Durairaj, P. Howlett, and J. Feary, "S17 the diagnostic accuracy of chest X-ray for the diagnosis of silicosis and how this relates to silica exposure," in *BMJ Open Respiratory Research*. BMJ Publishing Group Ltd, 2024.

[6] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[7] P. Rajpurkar, J. Irvin, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Ding, A. Bagul, C. Langlotz, K. Shpanskaya, M. P. Lungren, and A. Y. Ng, "Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning," 2017. [Online]. Available: https://arxiv.org/abs/1711.05225

[8] B. Priego-Torres *et al.*, "Automated engineered-stone silicosis screening and staging using deep learning with X-rays," *Computers in Biology and Medicine*, vol. 191, p. 110153, 2025.

[9] J. Yi, K. Tapia, J. W. Robinson, Z. Gandomkar, M. E. Suleiman, P. C. Brennan, N. Sommerfeld, and S. T. Taba, "Radiologists' performance in diagnosing silicosis on high-resolution computed tomography (HRCT) scans: an online platform," in *Medical Imaging 2024: Image Perception, Observer Performance, and Technology Assessment*, C. R. Mello-Thoms and Y. Chen, Eds., vol. 12929, International Society for Optics and Photonics. SPIE, 2024, p. 1292908. [Online]. Available: https://doi.org/10.1117/12.3006114

[10] J. Zhang, Y. Xie, G. Pang, Z. Liao, J. Verjans, W. Li, Z. Sun, J. He, Y. Li, C. Shen, and Y. Xia, "Viral pneumonia screening on chest x-rays using confidence-aware anomaly detection," *IEEE Transactions on Medical Imaging*, vol. 40, no. 3, pp. 879–890, 2021.

[11] A. Sharma, A. Sharma, and K. J. Guo, "Intelligent medical diagnosis model based on graph neural networks for medical images," *Computers and Their Applications in Intelligent Technologies*, vol. 10, no. 4, pp. 1201–1216, Aug. 2025.

[12] J. Bai, B. Li, and S. Nabavi, "Semi-supervised classification of disease prognosis using CR images with clinical data structured graph," in *Proceedings of the 13th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics*, 2022, pp. 1–9.

[13] D. Ahmedt-Aristizabal, M. A. Armin, S. Denman, C. Fookes, and L. J. S. Petersson, "Graph-based deep learning for medical diagnosis and analysis: past, present and future," *Sensors*, vol. 21, no. 14, p. 4758, 2021.

[14] R. F. Álvarez, C. M. González, A. Q. Martínez, J. J. B. Pérez, L. C. Fernández, and A. P. Fernández, "Recommendations of separ: Guidelines for the diagnosis and monitoring of silicosis," *Archivos de Bronconeumología (English Edition)*, vol. 51, no. 2, pp. 86–93, Feb. 2015.

[15] CDC – National Institute for Occupational Safety and Health, "Silicosis reporting," https://www.cdc.gov/niosh/surveillance/respiratorydisease/silicosis-reporting.html, 2025, accessed on April 7, 2025.

[16] T. B. Chandra and K. Verma, "Pneumonia detection on chest x-ray using machine learning paradigm," in *Proceedings of 3rd International Conference on Computer Vision and Image Processing (CVIP 2018), Volume 1*. Springer, 2020, pp. 21–33.

[17] K. M. Kuo, P. C. Talley, C. H. Huang, and L. C. J. Cheng, "Predicting hospital-acquired pneumonia among schizophrenic patients: a machine learning approach," *BMC Medical Informatics and Decision Making*, vol. 19, pp. 1–8, 2019.

[18] V. Rajinikanth, S. Kadry, R. Damaševičius, D. Taniar, and H. T. Rauf, "Machine-learning-scheme to detect choroidal-neovascularization in retinal oct image," in *2021 Seventh International conference on Bio Signals, Images, and Instrumentation (ICBSII)*, 2021, pp. 1–5.

[19] G.-k. Sun, Y.-h. Xiang, L. Wang, P.-p. Xiang, Z.-x. Wang, J. Zhang, and L. Wu, "Development of a multi-laboratory integrated predictive model for silicosis utilizing machine learning: a retrospective case-control study," *Frontiers in Public Health*, vol. Volume 12 - 2024, 2025. [Online]. Available: https://www.frontiersin.org/journals/public-health/articles/10.3389/fpubh.2024.1450439

[20] T. Ozturk *et al.*, "Automated detection of COVID-19 cases using deep neural networks with X-ray images," *Computers in Biology and Medicine*, vol. 121, p. 103792, 2020.

[21] R. E. Al Mamlook, S. Chen, and H. F. Bzizi, "Investigation of the performance of machine learning classifiers for pneumonia detection in chest X-ray images," in *2020 IEEE International Conference on Electro Information Technology (EIT)*. IEEE, 2020, pp. 98–104.

[22] J. van der Laak, F. Ciompi, and G. Litjens, "Graph neural networks in histopathology: Emerging trends and challenges," *Patterns*, vol. 2, no. 7, p. 100365, 2021.

[23] A. UMC, "Graph neural networks in histopathology," Technical Report, 2021, available online: https://amsterdamumc.org.

[24] K. Zhao, Y. Wang, and J. Zhang, "Graph neural networks in cancer and oncology research: A comprehensive review," *Briefings in Bioinformatics*, vol. 23, no. 5, p. bbac297, 2022.

[25] Z. Zeng, Y. Li, Q. Xu, and W. Chen, "A multimodal graph neural network framework for cancer molecular subtyping," *BMC Bioinformatics*, vol. 24, p. 355, 2023.

[26] S. Ahmed, A. Hassan, and H. Kim, "Enhanced brain tumor classification using CNN-based graph neural network," *Scientific Reports*, vol. 13, no. 1, p. 12945, 2023.

[27] M. Nguyen, Q. Tran, and D. Pham, "Graph neural network-based breast cancer diagnosis using ultrasound imaging," *Frontiers in Oncology*, vol. 13, p. 1123456, 2023.

[28] C. Li, H. Zhang, and R. Wang, "A graph neural network framework for mapping histological topology," *BMC Bioinformatics*, vol. 23, no. 1, p. 487, 2022.

[29] L. He, W. Zhang, and X. Chen, "Graph neural networks for image-guided disease diagnosis: A review," *WIREs Computational Molecular Science*, vol. 11, no. 6, p. e1528, 2021.

[30] G. Gong, Y. Wang, and L. Zhang, "Graph neural networks in brain connectivity studies: methods and applications," *Frontiers in Neuroscience*, vol. 18, p. 1278912, 2024.

[31] F. Liu, H. Yang, and Y. Zhao, "Multimodal data integration for oncology in the era of deep neural networks," *Cancer Letters*, vol. 543, p. 215812, 2022.

[32] Z. Sun, X. Jiang, and F. Wang, "Graph neural networks in healthcare: A systematic review," *Artificial Intelligence in Medicine*, vol. 137, p. 102455, 2023.

[33] D. J. M. Kermany, "Labeled optical coherence tomography (OCT) and chest X-ray images for classification," https://data.mendeley.com/datasets/rscbjbr9sj/3, 2018.

[34] A. Howard, M. Sandler, B. Chen, W. Wang, L.-C. Chen, M. Tan, G. Chu, V. Vasudevan, Y. Zhu, R. Pang, H. Adam, and Q. Le, "Searching for MobileNetV3," in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019, pp. 1314–1324.

[35] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 4700–4708.

[36] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.

[37] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin transformer: Hierarchical vision transformer using shifted windows," in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021, pp. 9992–10 002.

[38] Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, "A convNet for the 2020s," in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022, pp. 11 966–11 976.

[39] X. He, H. Fan, Y. Wu, Y. Zhang, J. Liang, and L. Sun, "TransFG: A transformer architecture for fine-grained recognition," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, no. 1, 2022, pp. 1096–1104.

[40] Z. Huang and Y. Li, "Interpretable and accurate fine-grained recognition via region grouping," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 8659–8669.

Received on August 08, 2025
Accepted on September 16, 2025