International Journal on Advanced Science Engineering Information Technology

Vol.15 (2025) No. 1
ISSN: 2088-5334

An Efficient Lung Disease Classification from X-ray Image Using Graph Neural Network and Transformer

Jiang Zhihao [a,b], Noridayu Bt Manshor [b,*], Hazlina Hamdan [b], Chen Limi [a]

[a] *Hainan Vocational University of Science and Technology, Hainan, China*
[b] *Faculty of Computer Science and Information Technology, Universiti Putra Malaysia, Malaysia*
Corresponding author: [*]ayu@upm.edu.my

***Abstract*—This research addresses the need for improved accuracy and efficiency in lung disease classification from X-ray images by developing a novel approach that integrates Graph Neural Networks (GNNs) with Transformer models. Using the extensive ChestX-ray14 dataset, which includes a wide array of lung disease cases, this study introduces and validates the TransGNN model. TransGNN harnesses the complementary strengths of GNNs and Transformers through advanced attention mechanisms and a dual-branch architecture, effectively capturing the complex and variable characteristics inherent in medical imaging data. The methodological framework includes rigorous data preprocessing, the application of weighted focus loss functions to address significant class imbalances, and extensive data augmentation techniques to bolster the model's robustness during testing. Results show that TransGNN surpasses conventional models by achieving superior classification accuracy across multiple lung diseases, demonstrating substantial improvements in diagnostic precision and reliability. Furthermore, the model's continuous learning mechanisms enable it to adapt seamlessly to new data and advances in medical imaging technology, making it highly versatile. This study highlights the potential of GNNs and Transformers to revolutionize the diagnostic landscape, offering a powerful, precise, and efficient tool for lung disease diagnosis. Future research should aim to incorporate additional patient-specific data, explore more advanced neural network architectures, and validate the model across diverse patient populations to enhance diagnostic accuracy further and expand its practical applicability in clinical settings, paving the way for a new standard in AI-driven medical diagnostics.**

***Keywords*—Lung disease classification; X-ray images; Graph Neural Networks (GNNs); transformers.**

Manuscript received 7 Mar. 2024; revised 19 Jun. 2024; accepted 25 Aug. 2024. Date of publication 28 Feb. 2025.

I. INTRODUCTION

Lung diseases pose significant global health challenges, requiring prompt and accurate diagnosis for effective treatment. Traditional diagnostic methods, such as radiologists' manual interpretation of X-ray images, are often time-consuming and prone to variability and errors. The increasing prevalence of lung diseases and the growing volume of medical imaging data demand more efficient and reliable diagnostic tools. Recent advancements in artificial intelligence (AI) have introduced powerful techniques like Graph Neural Networks (GNNs) and Transformers, which have shown promise in enhancing medical image analysis.

The concept of graph neural networks (GNNs) was first introduced by Gori et al. [1] in 2005 and further detailed by Scarselli et al. [2]. in 2009. A Graph Neural Network (GNN) directly operates on graph structures. Graphs, consisting of nodes (vertices) connected by edges, represent complex relational data effectively. GNNs capture dependencies between nodes, making them highly effective for graph-representative data in applications like social networks, molecular structures, and communication networks [3]. Graph Neural Networks, known for their ability to handle relational data, offer a novel approach to medical image analysis. Unlike traditional convolutional neural networks (CNNs) that process images as a grid of pixels, GNNs can interpret images as a series of interconnected nodes (or regions of interest). This perspective is particularly advantageous in understanding the complex spatial relationships and patterns present in lung X-rays, which are crucial for accurate disease classification [4].

Vaswani [3], in their seminal paper "Attention Is All You Need", presented the Transformer model, which significantly diverged from traditional sequence modeling techniques such as Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks. This model is a deep learning framework that utilizes a multi-head attention mechanism. In

2023, Peebles developed the Diffusion Transformer (DiT), adapting the Transformer for diffusion-based image generation. Initially developed for natural language processing, Transformers have recently gained traction in computer vision [4]. Their self-attention mechanisms allow them to focus on relevant input parts, making them suitable for analyzing medical images. Transformers can efficiently handle sequential pixel intensity variations in X-rays, providing a complementary approach to GNNs for understanding and classifying lung diseases.

GNNs are particularly effective in capturing complex spatial relationships within medical images, making them suitable for tasks like lung disease classification from X-ray images. Initially developed for natural language processing, transformers have been successfully adapted for image analysis due to their ability to handle sequential data and long-range dependencies. By integrating GNNs and Transformers, it is possible to leverage their complementary strengths, leading to improved feature extraction and classification accuracy.

This study aims to develop an efficient hybrid model combining GNNs and Transformers for lung disease classification from X-ray images. The proposed models, TransGNN, incorporate attention mechanisms and dual-branch architectures to enhance diagnostic performance. This research addresses the limitations of conventional methods, such as class imbalance and insufficient feature representation, by utilizing advanced AI techniques and continuous learning mechanisms. The ultimate goal is to provide a robust, accurate, and efficient diagnostic tool to improve patient outcomes and significantly support radiologists in clinical practice.

For the literature review, from 2021 to 2024, the researchers applied traditional machine learning and deep learning to image recognition and classification of medical diseases. However, it has some disadvantages, such as low precision, low efficiency, and limited ability to process complex data.

A. Shortcomings of Conventional Machine Learning Approaches

Traditional machine learning methods such as SVM and decision trees have been widely used in medical image analysis due to their relatively low computational requirements and ease of implementation. However, these methods often rely on handcrafted features, which can limit their ability to capture the subtle and complex patterns inherent in medical images [5].This reliance on manual feature engineering can introduce bias and subjectivity, potentially leading to inconsistent diagnostic outcomes [6]. Additionally, conventional ML techniques struggle with the high dimensionality of medical images, which can result in overfitting and poor generalizability to new, unseen data [7].

B. Inefficiencies of Current Diagnostic Techniques

Current diagnostic techniques for lung diseases heavily rely on manual interpretation of chest X-rays by radiologists, which is time-consuming and subject to variability and error [8]. This manual approach faces challenges such as difficulty in detecting subtle early-stage pathologies, which can lead to delayed treatments [3]. Moreover, traditional diagnostic methods often lack the flexibility to integrate and analyze complex data from modern electronic health records, resulting in incomplete patient assessments and impaired diagnostic accuracy [9].

C. Advancements in Graph Neural Network (GNN)

Graph Neural Networks (GNNs) have emerged as a powerful tool in medical image analysis due to their ability to capture complex spatial relationships within data [10]. GNNs are particularly effective in interpreting intricate patterns in lung X-rays, making them superior to traditional Convolutional Neural Networks (CNNs) in specific contexts [11]. Recent studies have demonstrated the effectiveness of GNNs in identifying spatial patterns and disease stages, with some models integrating clinical data for more personalized diagnoses [12]. However, GNNs still face challenges such as data scarcity and the need for extensive annotated datasets for practical training [13].

D. Adoption of Transformers in Medical Imaging

Initially developed for natural language processing [4], transformers have recently been adapted for medical imaging, including lung disease classification from X-ray images. Transformers excel in capturing long-range dependencies and global context within images, enhancing the accuracy of disease diagnosis [4]. Hybrid models combining Transformers with CNNs have shown promise in effectively distinguishing between similar pathologies by focusing on critical areas of the images [14] Despite their potential, the high computational demands of Transformers pose challenges for their integration into clinical workflows [15].

E. Integration of GNNs and Transformers

The integration of GNNs and Transformers in medical image analysis leverages the strengths of both models, offering a robust solution to the limitations of conventional methods [16]. GNNs' ability to capture complex spatial relationships complements Transformers' proficiency in handling sequential data and long-range dependencies [17]. This combination enhances feature extraction and classification accuracy, as evidenced by the superior performance of the proposed TransGNN models in lung disease classification [18].

In summary, the advancements in GNNs and Transformers and strategies to address data imbalance represent significant progress in medical image analysis. These developments provide a strong foundation for further research and the development of more accurate and efficient diagnostic tools for lung disease classification [19].

II. MATERIALS AND METHOD

We first collect a lung dataset in the methodological section with X-ray images of public-sector diseases.

A. Materials

In this section, we conduct experiments on a publicly available chest X-ray 14 dataset [20], the most extensive available collection of chest radiographs. It comprises 112,120 X-ray images from 30,805 patients, each with a resolution of 1024×1024 pixels, and stored in 8-bit grayscale PNG format. The dataset includes labels for 14 chest diseases derived from radiologists' reports using Natural Language

Processing (NLP) with an accuracy exceeding 90%. Of all the chest radiographs, 51,759 contain one or more diseases, while 60,361 are labeled as "no findings." Table 1 lists the pulmonary disease types in the dataset. The dataset includes multi-label classification with a sample size large enough for deep learning. In this experiment, the entire dataset is divided into a hospital-scale training set (totaling 75,708 images), a validation set (totaling 10,816 images), and a test set (totaling 25,596 images). All images of the same patient appear only once across the training, validation, and test sets. Chest X-ray 14 dataset statistical table of lung disease types shown in Table 1.

TABLE I
CHEST X-RAY 14 DATASET STATISTICAL TABLE OF LUNG DISEASE TYPES

Pulmonary Disease	Number of cases of the disease	Number of non-local disease cases	Ratio
Atelectasis	11559	100561	10.31
Cardiomegaly	2776	109344	2.48
Effusion	13317	98803	11.88
Infiltration	19894	92226	17.74
Mass	5782	106338	5.16
Nodule	6631	105789	5.65
Pneumonia	1431	110689	1.28
Pneumothorax	5302	106818	4.73
Consolidation	4667	107453	4.16
Edema	2303	109817	2.05
Emphysema	2516	109604	2.24
Fibrosis	1686	110434	1.50
Pleural Thickening	3385	108735	3.02
Hernia	227	111893	0.20
No finding	60361	51759	53.84

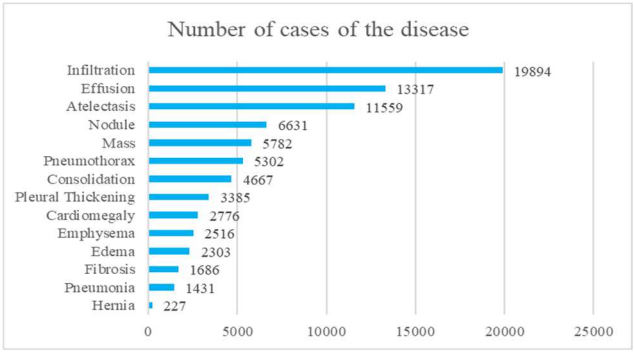

Fig. 1 Quantitative distribution of 14 lung diseases [19]

The bar graph presents the distribution of lung disease cases within a dataset, showcasing the prevalence of various conditions. Infiltration is the most common disease, evidenced by the highest count of 19,894 cases, followed by Effusion and Atelectasis with 13,317 and 11,559 cases, respectively. Diseases like Nodules, Masses, and Pneumothorax also display significant occurrences. Hernia is the least common condition towards the lower end, with only 227 cases. This graph helps understand the relative frequencies of these diseases, which could inform healthcare priorities and resource allocation in medical settings.

B. Data Description

Pre-trained models require normalized input images, typically small batches of three-channel RGB images (3×H×W), with H or W at least 224 [21]. Chest X-ray 14 images are 1024×1024 with 8-bit grayscale values. The dataset is split into training, validation, and test sets. Grayscale images are converted to three-channel RGB, center-cropped to 224×224, and normalized using the mean ([0.485, 0.456, 0.406]) and standard deviation ([0.229, 0.224, 0.225]). The model was trained with the Adam optimizer, an initial learning rate of 1.0e-4, and a batch size of 32 [22]. Training iterations were completed, and the model was validated, tested, and saved after each cycle for optimal performance. ROC curves and AUC values evaluated multi-class classification performance, saving model weights corresponding to the best AUC value for feature extraction. To avoid confusion, the family name must be written as the last part of each author's name [23]. Each affiliation must include, at the very least, the name of the company and the name of the country where the author is based [24].

To solve the problem of data imbalance in the sample of 14 diseases, we use focus loss in the weighted loss function, which adjusts the contribution of each sample to the overall loss according to the classification difficulty. By increasing the weight of underrepresented categories, models can focus more on learning the characteristics of those categories and improving their performance on less common diseases.

C. Lung Disease X-ray Image Processing

The images come from various datasets with differing sizes and conditions. Preprocessing is applied to minimize performance impacts. Different X-ray types are used, each with distinct technologies and resolutions. This represents the dataset normalization process:

$$M_i = \frac{y_i - min(y)}{max(y) - min(y)} \quad (1)$$

where, y is for original images and M is a normalized image. X represents the standardized image. The mean and standard deviation from the training set is used to standardize the data distribution for the test and validation sets.

$$x_i = \frac{M_i - mean(M)}{S.D(M)} \quad (2)$$

Histogram equalization extends pixel intensity from 0 to 255, producing an image with higher contrast and intensity. Gaussian blur with a filter size of 55 minimizes noise, while a bilateral filter with values of 5, 75, and 75 preserves edge information. To enhance classification accuracy, the diaphragm is removed using adaptive masking, which calculates pixel intensity's maximum and minimum and applies binary thresholding.

D. Methods

The algorithm enhances the Mask GNN architecture for segmenting lung disease areas from chest X-ray (CXR)images. Mask GNN, introduced in 2018 [19], generates pixel-level masks for detected objects, and showed high accuracy in the COCO segmentation challenge (see Fig .2) [19].

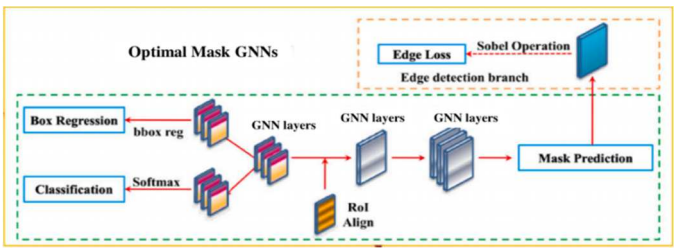

Fig. 2 The overall architecture of GNN [19]

In graph representation learning, graph neural networks (GNNs) have become the dominant choice, achieving state-of-the-art performance in tasks such as node classification, link prediction, and graph classification [25]. Although GNNs and their variants have made significant progress, some limitations remain. On the one hand, the message-passing mechanism relies on edges to fuse graph structure and node attributes, resulting in strong bias and noise. If there are no connections between nodes, some helpful information cannot be obtained.

Furthermore, in real graph data, many noisy edges connect irrelevant nodes. These biases and noise caused by the message-passing mechanism cause the representation learned by the GNN to contain incomplete information, thus harming the performance of downstream tasks [26]. On the other hand, the receptive field of GNN is also limited due to the over-smoothing problem. It turns out that as the architecture of GNNs becomes deeper and reaches a certain extent, the model no longer responds to the training data, and the node representations obtained by such deep models tend to be over-smoothed and become indistinguishable. These limitations lead to frustrating compromises in GNNs, where shallow GNNs with limited receptive fields can only aggregate incomplete information within the neighborhood, while deep GNNs suffer from over-smoothing [24]. With the help of Transformer, the receptive field of GNN can be expanded to more relevant nodes, which may be far away from the central node. On the other hand, GNN can help the Transformer capture complex graph topology information and efficiently aggregate more relevant nodes from the neighborhood.

The framework of TransGNN is shown in Fig.3. The framework consists of three essential components: (1) attention sampling module, (2) position encoding module, (3) TransGNN module. First, each central node's most relevant nodes are sampled by considering the semantic similarity and graph structure information in the attention sampling module. Then in the positional encoding module, the positional encoding is calculated to help the Transformer capture the graph topology information. After these two modules, we use the TransGNN module, which contains three sub-modules in sequence: (i) Transformer layer, (ii) GNN layer, and (iii) Sample update sub-module. Among them, the Transformer layer is used to expand the receptive field of the GNN layer and efficiently aggregate attention sample information, while the GNN layer helps the Transformer layer perceive graph structure information and obtain more relevant information about neighboring nodes. The sample update sub-module is used to update attention samples upon new representations efficiently.

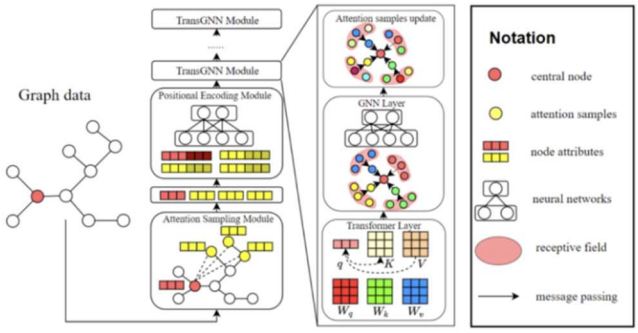

Fig. 3 TransGNNs Framework [24]

First, relevant node sampling is performed on the central node, then the position encoding is calculated, and the original attributes are enhanced by combining structural information. In the TransGNN module, the Transformer layer and the GNN layer improve each other, followed by the sample update sub-module [27].

1) Transformer layer: Use Transformer layers to improve GNN layers, extending the receptive field to more relevant nodes that may be far away from the neighborhood [28].

$$\begin{cases} q = h_i W_q \\ K = H_i^{Samp} W_k \\ V = H_i^{Samp} W_v \\ a_t = \frac{qK^T}{\sqrt{d_{out}}} \end{cases} \quad (3)$$

2) Multi-head attention:

$$MultiHead(h_i) = Concat(head_1, \cdots, head_m)W_m \quad (4)$$

3) GNN layer: Use the GNN layer to fuse representation and graph structure to help the Transformer layer make better use of the graph structure [29].

$$\begin{cases} h_{M(v_i)} = Messag(h_k, \forall v_k \in N(v_i)) \\ h_i = Combine(h_i, h_{M(v_i)}) \end{cases} \quad (5)$$

*4) Sample update submodule***:** After the Transformer layer and GNN layer, the attention samples should be updated according to the new representation [30].

E. Methods

Various crucial metrics are employed when evaluating medical image analysis, especially within machine learning and deep learning models. These metrics are instrumental in assessing the models' performance, accuracy, and reliability in classifying, segmenting, or detecting medical conditions from images. Image classification is a fundamental computer vision task categorizing images into distinct classes with minimal error. In single-label classification, where each image belongs to one category, key performance indicators include accuracy, ROC curve, and AUC [31].

$$Accuracy = (TP + TN)/(TP + FP + TN + FN) \quad (6)$$

The accuracy rate and the recall rate in the machine learning classification algorithm contradict each other. As the judgment threshold increases, the accuracy rate will continue to increase while the recall rate will continue to decrease. Generally, the higher the Recall rate, the lower the Precision rate. The precision-recall curve can be drawn according to different values, as shown in Fig.4.

ROC curves evaluate a classifier's performance at various thresholds and are typically shown in Fig.5. In a ROC curve, each point's horizontal coordinate is the False Positive Rate (FPR), and the vertical coordinate is the True Positive Rate (TPR), illustrating the balance between true positives and false positives [32].

$$TPR = TP/(TP + FN) \quad (7)$$

$$FPR = FP/(FP + TN) \quad (8)$$

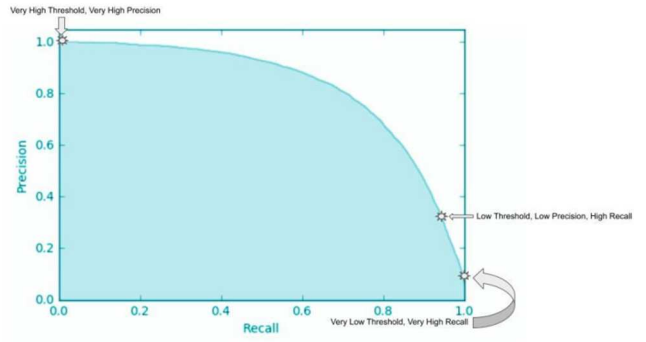

Fig. 4 Precision-Recall curve [19]

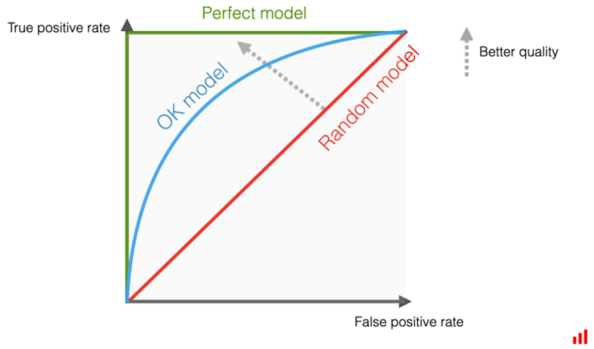

Fig. 5 ROC curve [19]

The ROC curve can remain unchanged when the distribution of positive and negative samples in the test set varies, showing insensitivity to sample imbalance. For instance, a tenfold increase in negative samples does not impact TPR, while FPR will also increase proportionally with little overall change. Therefore, ROC is usually used as the evaluation criterion for unbalanced sample problems. The optimal classifier performance is indicated when its ROC curve approaches the top left corner. The superiority of one classifier over another is established if its ROC curve entirely encompasses that of the other [33].

III. Results and Discussion

A. Results

Different models, including MobileLungNetV2, GNN, and TransGNN models, were selected for comparison to verify the generalization performance of the TransGNN method. Experiments were carried out using the same method on the same data set [26]. Each ROC value of the generalization experiment is shown in Fig.6.

The comparison of the classification performance of each model for 14 chest diseases is shown in Fig.6. Here, each bar graph corresponds to each disease, and the text on the horizontal axis of the bar graph represents MobileLungNetV2, GNN, TransGCN and TransGNN and the average performance of the six models. The values on the vertical axis are the AUC values, the red bar is the model with the highest AUC value, and the green bar is the arithmetic mean of the AUC values of the six models. The AUC values of the six models for the classification of 14 chest diseases and their average AUC values are shown in Table 2.

Table II compares the classification accuracy of each model for 14 chest diseases. The classification accuracies of MobileLungNetV2, GNN, and TransGNN were 0.7930, 0.9697, and 0.9766.

TABLE II
ACCURACY COMPARISON WITH EXISTING METHODS ON CHESTXRAY14 DATASET

Type	MobileLungNetV2[26]	GNN [18]	TransGNN
Atelectasis	0.9949	0.8356	0.9951
Cardiomegaly	0.9963	0.9683	0.9968
Effusion	0.9694	0.7911	0.9752
Infiltration	0.9635	0.9040	0.9759
Mass	0.9371	0.9135	0.9689
Nodule	0.9697	0.7681	0.9689
Pneumonia	0.9968	0.8140	0.9952
Pneumothorax	0.9596	0.6852	0.9764
Consolidation	0.9459	0.6472	0.9552
Edema	0.9692	0.7748	0.9723
Emphysema	0.9779	0.8189	0.9781
Fibrosis	0.9562	0.7222	0.9646
Thickening	0.9686	0.6918	0.9755
Hernia	0.9705	0.7666	0.9736
Average	0.9697	0.7930	0.9766

Table III shows that TransGNN's classification of 14 chest diseases is superior to two advanced methods such as MobileLungNetV2 and GNN, with better AUC value and stronger classification and generalization ability.

TABLE III
AUC COMPARISON WITH EXISTING METHODS ON CHESTX-RAY14 DATASET

Type	MobileLungNetV2[26]	GNN [18]	TransGNN
Atelectasis	0.936	0.816	0.941
Cardiomegaly	0.855	0.926	0.929
Effusion	0.931	0.958	0.959
Infiltration	0.912	0.749	0.910
Mass	0.883	0.806	0.898
Nodule	0.962	0.799	0.964
Pneumonia	0.904	0.774	0.916
Pneumothorax	0.967	0.891	0.963
Consolidation	0.889	0.873	0.931
Edema	0.958	0.893	0.962
Emphysema	0.935	0.947	0.943
Fibrosis	0.891	0.805	0.899
Thickening	0.872	0.817	0.882
Hernia	0.975	0.931	0.971
Average	0.923	0.856	0.933

B. Discussion

TransGNN achieves the highest AUC value of 0.933, outperforming MobileLungNetV2 by 0.010 due to its enhanced feature extraction capabilities, attributable to its deep network layers. The complexity of GNN training frameworks, especially when using deeper transformers, complicates training and increases the risk of overfitting, which may hinder achieving optimal classification performance with the TransGCN model. Suggesting that its key technologies significantly enhance classification performance.

Fig. 6 and Table III depict the model's overall performance, with an accuracy of 97.66% and an AUC value of 0.933, reflecting its excellent classification and generalization abilities. The study utilizes graph neural networks and transformers to classify multiple lung lesions effectively. The TransGNN model and two other transfer learning models

were trained on the ChestX-ray14 dataset, which is widely recognized for its size and has been used in numerous prior studies for benchmarking and technique comparison.

Fig.6 and Table III present a comparative analysis of the classification performance across 14 lung diseases using three different models: MobileLungNetV2, GNN, and TransGNN. The figure uses bar graphs to depict the AUC values for each disease, allowing for a clear comparison of the models' effectiveness. The results demonstrate that TransGNN consistently achieves the highest AUC values across most diseases, indicating its superior classification capability.

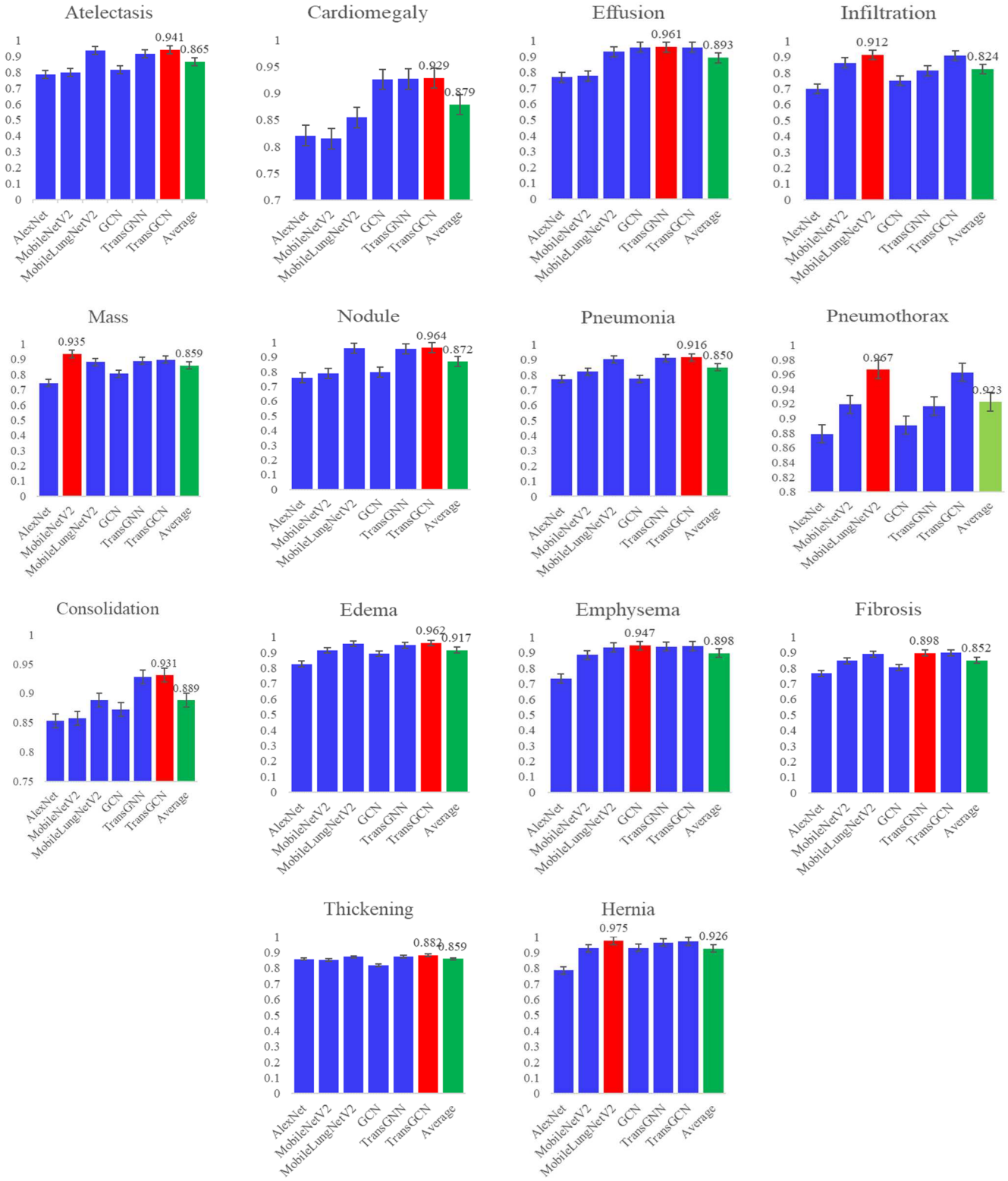

Fig. 6 Comparison of classification performance of each model for 14 diseases

For example, in diseases such as Atelectasis, Cardiomegaly, and Pneumonia, TransGNN significantly outperforms the other models, reaching near-perfect AUC values close to 1.0. This suggests that TransGNN has a strong ability to distinguish between classes in these specific diseases. In contrast, MobileLungNetV2 and GNN show lower performance, particularly in complex diseases like Pneumothorax and Consolidation, where their AUC values are notably lower. This underperformance highlights their limitations in capturing intricate patterns in medical imaging data.

The figure also shows that, on average, TransGNN has a superior generalization ability, achieving higher overall AUC scores than the other models. This superiority is likely due to the enhanced feature extraction and attention mechanisms employed by TransGNN, allowing it to capture the nuances in lung disease images better. Overall, Fig.6 and Table III illustrate the effectiveness of TransGNN in lung disease classification, showcasing its potential as a robust diagnostic tool in medical imaging.

IV. Conclusion

This research has demonstrated that the integration of Graph Neural Networks (GNNs) and Transformers significantly enhances the accuracy and efficiency of lung disease classification from X-ray images. The proposed models, TransGNN, leverage advanced AI techniques, including attention mechanisms and dual-branch architectures, to address and overcome the limitations of traditional diagnostic methods. These models offer superior performance in recognizing various lung conditions and adapting continuously to new data, making them highly effective in clinical settings.

Through extensive validation of the ChestX-ray14 dataset, these models have shown a marked improvement over conventional methods, particularly in managing the inherent class imbalance in medical datasets. Applying a weighted focus loss function has proved instrumental in this success, ensuring that the models prioritize learning from underrepresented classes without sacrificing accuracy.

For future work, exploring the integration of additional contextual patient data, such as demographic and clinical histories, is recommended to further personalize and enhance the diagnostic processes. Moreover, developing methods to reduce the computational demands of these AI models could facilitate their integration into existing medical infrastructure, making advanced diagnostics more accessible to a broader range of healthcare providers. Further exploration into multi-modal learning, combining X-ray imaging with other diagnostic modalities, could also provide a more comprehensive understanding of lung diseases, potentially leading to earlier and more accurate diagnoses.

References

[1] M. Gori, G. Monfardini, and F. Scarselli, "A new model for learning in graph domains," in *Proc. IEEE Int. Joint Conf. Neural Netw.*, vol. 2, 2005, pp. 729–734, doi: 10.1109/ijcnn.2005.1555942.

[2] F. Scarselli, M. Gori, A. C. Tsoi, M. Hagenbuchner, and G. Monfardini, "The graph neural network model," *IEEE Trans. Neural Netw.*, vol. 20, no. 1, pp. 61–80, Jan. 2009, doi: 10.1109/tnn.2008.2005605.

[3] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, et al., "Attention is all you need", *Proc. 31st Int. Conf. Neural Information Processing Systems*, pp. 6000-6010, 2017.

[4] W. Peebles and S. Xie, "Scalable diffusion models with transformers," in *2023 IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, 2023, pp. 4172–4182, doi: 10.1109/iccv51070.2023.00387.

[5] D. Ahmedt-Aristizabal, M. A. Armin, S. Denman, C. Fookes, and L. Petersson, "Graph-based deep learning for medical diagnosis and analysis: Past, present and future," *Sensors*, vol. 21, no. 14, p. 4758, Jul. 2021, doi: 10.3390/s21144758.

[6] L. Zhang, Y. Zhao, T. Che, S. Li, and X. Wang, "Graph neural networks for image-guided disease diagnosis: A review," *iRadiolog*,vol. 1, no. 2, pp. 151–166, Jun. 2023, doi:10.1002/ird3.20.

[7] J. Zhou et al., "Graph neural networks: A review of methods and applications," *AI Open*, vol. 1, pp. 57–81, 2020, doi:10.1016/j.aiopen.2021.01.001.

[8] A. Zanfei, B. M. Brentan, A. Menapace, M. Righetti, and M. Herrera, "Graph convolutional recurrent neural networks for water demand forecasting," *Water Resour. Res.*, vol. 58, no. 7, Jul. 2022, doi:10.1029/2022wr032299.

[9] R. Yamashita, M. Nishio, R. K. G. Do, and K. Togashi, "Convolutional neural networks: An overview and application in radiology," *Insights Imaging*, vol. 9, no. 4, pp. 611–629, Jun. 2018, doi: 10.1007/s13244-018-0639-9.

[10] L. Wu et al., "Graph neural networks for natural language processing: A survey," *Found. Trends Mach. Learn.*, vol. 16, no. 2, pp. 119–328, 2023, doi: 10.1561/2200000096.

[11] L. Wu, P. Cui, J. Pei, L. Zhao, and X. Guo, "Graph neural networks: Foundation, frontiers and applications," in *Proc. 29th ACM SIGKDD Conf. Knowl. Discov. Data Min.*, 2023, pp. 5831–5832, doi:10.1145/3580305.3599560.

[12] S. Thirumalaisamy et al., "Breast cancer classification using synthesized deep learning model with metaheuristic optimization algorithm," *Diagnostics*, vol. 13, no. 18, p. 2925, Sep. 2023, doi:10.3390/diagnostics13182925.

[13] R.-K. Sheu and M. S. Pardeshi, "A survey on medical explainable AI (XAI): Recent progress, explainability approach, human interaction and scoring system," *Sensors*, vol. 22, no. 20, p. 8068, Oct. 2022, doi:10.3390/s22208068.

[14] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998, doi: 10.1109/5.726791.

[15] M. Mamalakis et al., "DenResCov-19: A deep transfer learning network for robust automatic classification of COVID-19, pneumonia, and tuberculosis from X-rays," *Comput. Med. Imaging Graph.*, vol. 94, p. 102008, Dec. 2021, doi: 10.1016/j.compmedimag.2021.102008.

[16] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Med. Image Comput. Comput.-Assist. Interv. (MICCAI)*, 2015, pp. 234–241, doi:10.1007/978-3-319-24574-4_28.

[17] Y. Li, X. Wu, P. Yang, G. Jiang, and Y. Luo, "Machine learning for lung cancer diagnosis, treatment, and prognosis," *Genomics, Proteomics Bioinf.*, vol. 20, no. 5, pp. 850–866, Oct. 2022, doi:10.1016/j.gpb.2022.11.003.

[18] J. Hanson et al., "SPOT-Disorder2: Improved protein intrinsic disorder prediction by ensembled deep learning," *Genomics, Proteomics Bioinf.*, vol. 17, no. 6, pp. 645–656, Dec. 2019, doi: 10.1016/j.gpb.2019.01.004.

[19] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," *arXiv*, 2017. [Online]. Available: https://arxiv.org/abs/1706.05587.

[20] T. Dhar, N. Dey, S. Borra, and R. S. Sherratt, "Challenges of deep learning in medical image analysis—Improving explainability and trust," *IEEE Trans. Technol. Soc.*, vol. 4, no. 1, pp. 68–75, Mar. 2023, doi: 10.1109/tts.2023.3234203.

[21] K. He et al., "Transformers in medical image analysis," *Intell. Med.*, vol. 3, no. 1, pp. 59–78, Feb. 2023, doi: 10.1016/j.imed.2022.07.002.

[22] M. H. Al-Sheikh et al., "Multi-class deep learning architecture for classifying lung diseases from chest X-ray and CT images," *Sci. Rep.*, vol. 13, no. 1, Nov. 2023, doi: 10.1038/s41598-023-46147-3.

[23] M. Hajij et al., "Topological deep learning: Going beyond graph data," *arXiv*, 2022. [Online]. Available: https://arxiv.org/abs/2206.00606

[24] F. J. M. Shamrat, S. Azam, A. Karim, K. Ahmed, F. M. Bui, and F. De Boer, "High-precision multiclass classification of lung disease through customized MobileNetV2 from chest X-ray images," *Comput. Biol. Med.*, vol. 155, p. 106646, Mar. 2023, doi:10.1016/j.compbiomed.2023.106646.

[25] C. Metta, A. Beretta, R. Pellungrini, S. Rinzivillo, and F. Giannotti, "Towards transparent healthcare: Advancing local explanation methods in explainable artificial intelligence," *Bioengineering*, vol. 11, no. 4, p. 369, Apr. 2024, doi: 10.3390/bioengineering11040369.

[26] J. Shi, R. Wang, Y. Zheng, Z. Jiang, H. Zhang, and L. Yu, "Cervical cell classification with graph convolutional network," *Comput. Methods Programs Biomed.*, vol. 198, p. 105807, Jan. 2021, doi:10.1016/j.cmpb.2020.105807.

[27] Y. Huang et al., "Transformer-based factorized encoder for classification of pneumoconiosis on 3D CT images," *Comput. Biol. Med.*, vol. 150, p. 106137, Nov. 2022, doi:10.1016/j.compbiomed.2022.106137.

[28] D. Yao et al., "A mutual multi-scale triplet graph convolutional network for classification of brain disorders using functional or structural connectivity," *IEEE Trans. Med. Imaging*, vol. 40, no. 4, pp. 1279–1289, Apr. 2021, doi: 10.1109/tmi.2021.3051604.

[29] R. Najjar, "Redefining radiology: A review of artificial intelligence integration in medical imaging," *Diagnostics*, vol. 13, no. 17, p. 2760, Aug. 2023, doi: 10.3390/diagnostics13172760.

[30] W. Zhang et al., "HyGAnno: Hybrid graph neural network–based cell type annotation for single-cell ATAC sequencing data," *Brief. Bioinform.*, vol. 25, no. 3, Mar. 2024, doi: 10.1093/bib/bbae152.

[31] F. Shamshad et al., "Transformers in medical imaging: A survey," *Med. Image Anal.*, vol. 88, p. 102802, Aug. 2023, doi:10.1016/j.media.2023.102802.

[32] S. Bharati, P. Podder, and M. R. H. Mondal, "Hybrid deep learning for detecting lung diseases from X-ray images," *Inform. Med. Unlocked*, vol. 20, p. 100391, 2020, doi: 10.1016/j.imu.2020.100391.

[33] B. Hu, C. Zhou, H. Wang, and S. Chen, "Nonlinear tribo-dynamic model and experimental verification of a spur gear drive under loss-of-lubrication condition," *Mech. Syst. Signal Process.*, vol. 153, p. 107509, May 2021, doi: 10.1016/j.ymssp.2020.107509.