

Multi-class pneumonia detection using fine-tuned vision transformer model

Khushboo Trivedi, Chintan B. Thacker

Computer Science and Engineering Department, Parul University, Vadodara, India

Article Info

Article history:

Received Sep 12, 2024

Revised Mar 5, 2025

Accepted Mar 20, 2025

Keywords:

Chest X-Ray classification

Fine-tuning

Medical imaging

Multi-class pneumonia

Vision transformer

ABSTRACT

Distinguishing between the various forms of pneumonia (bacterial, viral, fungal, and normal) using chest X-rays is a major problem in global health. Conventional approaches to pneumonia identification frequently depend on laborious and error-prone manual interpretation. Current machine learning (ML) models, like convolutional neural networks (CNNs), have demonstrated some success, but they frequently fail on jobs requiring multi-class classification or generalization. The potential of vision transformer (ViT) models, fine-tuned to address these limitations, is explored. The approach enhances the accuracy of pneumonia classification into four distinct classes by leveraging the attention mechanism in vision transformers (ViTs). Fine-tuning with a tagged chest X-ray dataset improves the algorithm's ability to detect subtle variations in pneumonia types. The findings demonstrate the model's effectiveness in multi-class pneumonia diagnosis, achieving a significant performance improvement with 98% accuracy across the four classes. This work highlights the promise of vision transformers in medical imaging, enabling the development of improved and scalable pneumonia classification methods.

This is an open access article under the [CC BY-SA](#) license.



Corresponding Author:

Khushboo Trivedi

Computer Science and Engineering Department, Parul University

Vadodara, Gujarat, India

Email: khushboo.trivedi21305@paruluniversity.ac.in

1. INTRODUCTION

Pneumonia remains a significant global health challenge, infecting millions annually and causing a substantial number of deaths. Early and accurate diagnosis is critical for improving patient outcomes and reducing complications [1], [2]. Chest X-rays are widely used for diagnosing pneumonia, but interpreting these images is a complex task, especially when distinguishing between bacterial, viral, fungal, and mycoplasma pneumonia. Variability in how pneumonia manifests across patients further complicates diagnosis, making manual interpretation by radiologists both time-intensive and prone to errors [3]–[5].

Recent advancements in deep learning have led to promising developments in automating medical image analysis. Convolutional neural networks (CNNs) have been particularly successful in image-based disease diagnosis, including pneumonia classification [6]–[8]. However, CNNs often struggle with multi-class classification tasks that require recognizing subtle variations between categories. Additionally, their performance can degrade when applied to unseen or diverse datasets, highlighting a need for more adaptable and robust solutions [9], [10].

Vision transformers (ViTs) offer a compelling alternative to CNNs by leveraging self-attention mechanisms to capture long-range dependencies within images. This capability makes them well-suited for complex classification problems. The goal of this research is to explore the effectiveness of fine-tuned ViTs

in the multi-class classification of pneumonia, aiming to overcome the limitations of existing methods and achieve high accuracy across different pneumonia types. This study aspires to contribute to the development of scalable, reliable, and efficient diagnostic tools for pneumonia detection.

2. LITERATURE STUDY

The use of artificial intelligence (AI) and deep learning has revolutionized medical imaging, particularly in pneumonia diagnosis. ViTs have improved efficiency and accuracy in chest X-ray analysis for pneumonia [1]. Transfer learning algorithms, such as those applied to detect COVID-19 pneumonia, have been key in enhancing diagnosis [2], [3]. Genetic algorithms refining models like DCGANs with CNN architectures (*e.g.*, VGG-16) have further improved pneumonia categorization [4]. DenseNet-121 has been effective in pediatric pneumonia classification, even with imbalanced datasets [5], [6]. Systematic reviews have supported pneumonia diagnosis in regions with high comorbidities, like India [7], while models like the LACE index predict 30-day hospital readmissions for pneumonia [8], [9].

The coronavirus disease 2019 (COVID-19) pandemic has amplified the importance of deep learning in pneumonia detection. Emerging techniques, including explainable models and fuzzy-enhanced deep learning, have shown promise in early pneumonia prediction, especially in COVID-19 cases [10], [11]. Graph-based deep learning with diffusion pseudo-labeling has enhanced explainability in diagnoses [12], [13], and ViT-based models for COVID-19 screening provide strong diagnostic justifications [14]. Interpretable pneumonia algorithms that integrate multisource data have been developed [15]. Other diagnostic approaches, such as antigen and nucleic acid amplification tests, have also contributed to pneumonia research [16], [17]. CNNs with LIME have improved the interpretability of pneumonia diagnoses [18], and self-supervised learning has enhanced generalizability [19]. Pseudo-labeling has further refined COVID-19 diagnosis accuracy [20]. The World Health Organization has stressed the need to improve detection protocols during the pandemic [21]. Deep residual networks combined with transfer learning have optimized pediatric pneumonia diagnosis [22].

GANs are increasingly used in medical image analysis, including bone surface segmentation and breast ultrasound imaging [23], [24]. GAN-based augmentation has helped overcome data shortages in pneumonia diagnosis [25], [26] and has increased model generalizability in other areas, such as hip fracture detection and prostate cancer grading [27], [28]. Deep learning models like CheXNeXt have demonstrated effectiveness in diagnosing chest illnesses, including pneumonia, compared to radiologists [29]. The availability of open chest X-ray datasets has accelerated research by providing critical training data for models [30]. CNNs and GANs are particularly useful for pneumonia diagnosis, especially in data-limited scenarios [31], [32]. Fine-tuning pre-trained CNN models has enabled accurate localization and classification of lung illnesses in chest X-rays [33]. AI-based screening systems have played a crucial role in improving the accuracy and speed of pneumonia diagnosis during the COVID-19 pandemic [34]. The role of AI extends beyond pneumonia, influencing diagnostics in other diseases, such as early lung cancer detection [35], [36]. Pneumonia diagnosis systems based on CNNs have successfully processed large X-ray datasets [37], [38]. Advances in computational algorithms continue to drive the evolution of AI-driven medical image diagnosis [39], [40], and comparisons between CT scans and PCR tests have highlighted the importance of prompt COVID-19 pneumonia diagnosis [41].

3. METHOD

The method section begins with an overview of the datasets used for pneumonia detection, as detailed in Table 1. These publicly available datasets include a variety of chest X-ray images labelled to identify pneumonia, with details about their size, source, and specific use cases provided in the table. The preprocessing techniques, outlined in Table 2, describe steps taken to prepare the medical images for deep learning models, such as resizing to a consistent image size, normalizing pixel values, and augmenting the dataset to improve model generalization. Finally, Table 3 presents the deep learning architectures employed to classify pneumonia from chest X-rays, including advanced models like ViT, VGG-16, and ResNet-50, each chosen for their proven performance in image classification tasks. These models are trained using the processed datasets and are evaluated based on their accuracy and ability to detect pneumonia from the X-ray images.

Table 1. List of datasets

Dataset	Number of Images/Patients	Categories
NIH dataset [38]	108,948 chest radiographs from 32,717 patients	8
CHE dataset [39]	224,316 chest X-rays from 65,240 patients	14
KAG dataset [40]	5,863 chest X-ray images (pediatric patients only)	2

Table 2. List of pre-process methods

Method	Methods/Details
Resize [2], [3], [13], [14], [16], [17], [19], [21], [24]	<ul style="list-style-type: none"> - Nearest-neighbor interpolation: Uses the closest pixel to the new pixel location (fast but may appear blocky). - Bilinear interpolation: Averages the four closest pixels for a smoother result. - Bicubic interpolation: Uses cubic polynomials for smoother images but is computationally expensive. - Aspect ratio: Preserves aspect ratio during resizing to avoid distortion. - Padding: Adds borders to maintain aspect ratio during resizing.
Normalization [1], [2], [4], [6], [18], [22]	<ul style="list-style-type: none"> - Min-max normalization: Scales pixel values to a specified range, usually [0, 1] using the formula: $\text{normalized_value} = (\text{pixel_value} - \text{min_value}) / (\text{max_value} - \text{min_value})$ - Z-score normalization: Adjusts pixel values to have a mean of 0 and standard deviation of 1 using: $\text{normalized_value} = (\text{pixel_value} - \text{mean}) / (\text{standard_deviation})$ - Standardization: A subset of normalization, it rescales pixel values to have a mean of 0 and variance of 1 using: $\text{standardized_value} = (\text{pixel_value} - \text{mean}) / (\text{standard_deviation})$

Table 3. List of model

Method	Description
Vision transformer (ViT) [1]	A transformer-based architecture for image classification that uses self-attention mechanisms instead of CNNs. Divides images into fixed-size patches, processes them as tokens, and captures global dependencies.
VGG-16 [4]	A deep CNN architecture with 16 weight layers: 13 convolutional and 3 fully connected layers. It uses small 3x3 filters to extract fine details, with max-pooling layers for spatial downsampling.
DenseNet-121 [6]	A CNN where each layer is connected to all previous layers. Reduces feature redundancy and enhances feature reuse. Includes four dense blocks and transition layers for downsampling.
ResNet-50 [22]	A 50-layer architecture using residual connections (skip connections) to combat vanishing gradients. The network learns residual mappings instead of direct transformations.
InceptionNetV3 [38]	An advanced version of the Inception architecture with multiple filter sizes in a single layer. Uses inception modules to capture features at different scales and factorized convolutions to reduce computational complexity.
Fine-Tuning [1], [4], [6], [22], [38]	Involves applying a pre-trained model to a new task by continuing training on a different dataset, typically with a smaller learning rate. Helps adapt the model to new tasks without retraining from scratch.

4. PROPOSED SYSTEM

Figure 1 demonstrates a proposed three-stage procedure for training, testing, and assessing a model. This model is likely to be used for a medical imaging image classification problem with the following classes: normal, viral, bacteria, and fungal. The following is an exhaustive description of the steps and their constituent parts:

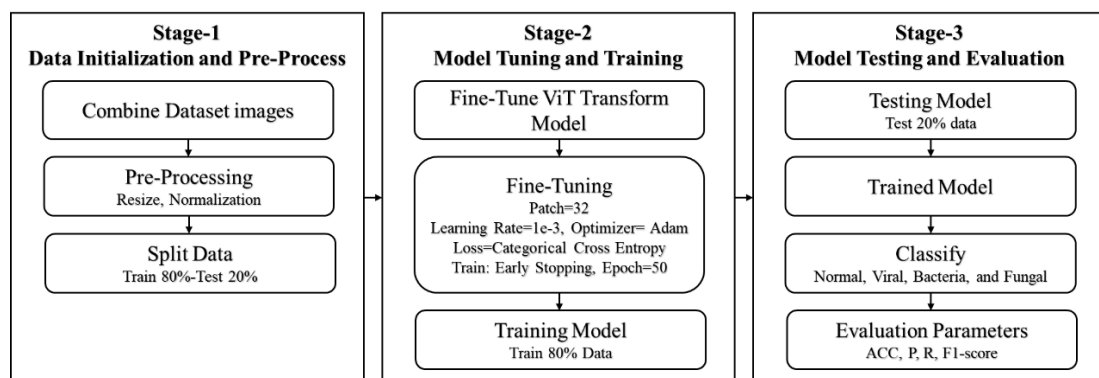


Figure 1. Proposed system flow diagram

Stage 1 data initialization and pre-processing: This stage involves combining images from various sources into a single dataset, sorted into categories like normal, viral, bacterial, and fungal. The data undergoes pre-processing, including resizing images to fit the ViT model and normalizing pixel values to a standard range. Finally, the dataset is split into training (80%) and testing (20%) sets for model training and evaluation.

Stage 2 model tuning and training: In this stage, the ViT model is fine-tuned using a pre-trained model and medical image data. The model processes images by dividing them into 32×32 patches. The learning rate is set to 1e-3, and the Adam optimizer adjusts weights during training. The categorical cross-entropy loss function is used, and early stopping prevents overfitting by halting training if performance plateaus before 50 epochs.

Stage 3 model testing and evaluation: The trained model is evaluated using the 20% test set, where its ability to generalize to unseen data is assessed. The model classifies images as normal, viral, bacterial, or fungal. Performance is evaluated using metrics such as accuracy (ACC), precision (P), recall (R), and F1-score to determine classification accuracy and model effectiveness.

5. RESULTS AND DISCUSSION

The findings of this study highlight the performance of the proposed ViT model, fine-tuned for pneumonia classification. A concise comparison with current transfer learning models, supported by detailed tables and figures, demonstrates the model's effectiveness across four pneumonia categories. The results are discussed in the context of the study's objectives, existing hypotheses, and related research while addressing potential interpretations and limitations. Figure 2 summarizes the dataset distribution across the four classes: 148 images for viral pneumonia, 242 for bacterial, 23 for fungal, and 232 for normal cases. This imbalance highlights the challenges in achieving robust performance across all categories, particularly for the minority class (fungal).

Figure 3 provides an overview of the fine-tuned ViT architecture. The model comprises over 85 million parameters, with only the final linear layer (3,076 trainable parameters) optimized during training. This design leverages the pre-trained weights of the frozen layers while fine-tuning the output layer to adapt to the pneumonia classification task.

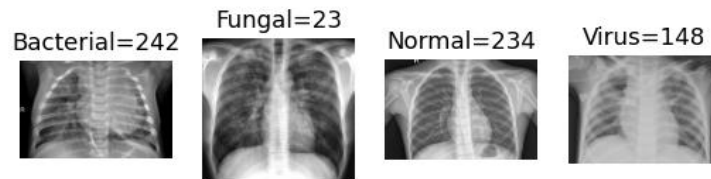


Figure 2. Dataset reading

Layer (type (var_name))	Input Shape	Output Shape	Param #	Trainable
VisionTransformer (VisionTransformer)	[32, 3, 224, 224]	[32, 4]	768	Partial
Conv2d (conv_proj)	[32, 3, 224, 224]	[32, 768, 14, 14]	(590,592)	False
Encoder (encoder)	[32, 197, 768]	[32, 197, 768]	151,296	False
Dropout (dropout)	[32, 197, 768]	[32, 197, 768]	--	--
Sequential (layers)	[32, 197, 768]	[32, 197, 768]	--	False
EncoderBlock (encoder_layer_0)	[32, 197, 768]	[32, 197, 768]	(7,087,872)	False
EncoderBlock (encoder_layer_1)	[32, 197, 768]	[32, 197, 768]	(7,087,872)	False
EncoderBlock (encoder_layer_2)	[32, 197, 768]	[32, 197, 768]	(7,087,872)	False
EncoderBlock (encoder_layer_3)	[32, 197, 768]	[32, 197, 768]	(7,087,872)	False
EncoderBlock (encoder_layer_4)	[32, 197, 768]	[32, 197, 768]	(7,087,872)	False
EncoderBlock (encoder_layer_5)	[32, 197, 768]	[32, 197, 768]	(7,087,872)	False
EncoderBlock (encoder_layer_6)	[32, 197, 768]	[32, 197, 768]	(7,087,872)	False
EncoderBlock (encoder_layer_7)	[32, 197, 768]	[32, 197, 768]	(7,087,872)	False
EncoderBlock (encoder_layer_8)	[32, 197, 768]	[32, 197, 768]	(7,087,872)	False
EncoderBlock (encoder_layer_9)	[32, 197, 768]	[32, 197, 768]	(7,087,872)	False
EncoderBlock (encoder_layer_10)	[32, 197, 768]	[32, 197, 768]	(7,087,872)	False
EncoderBlock (encoder_layer_11)	[32, 197, 768]	[32, 197, 768]	(7,087,872)	False
LayerNorm (ln)	[32, 197, 768]	[32, 197, 768]	(1,536)	False
Linear (heads)	[32, 768]	[32, 4]	3,076	True
Total params: 85,801,732				
Trainable params: 3,076				
Non-trainable params: 85,798,656				
Total mult-adds (G): 5.52				
Input size (MB): 19.27				
Forward/backward pass size (MB): 3330.74				
Params size (MB): 229.21				
Estimated Total Size (MB): 3579.21				

Figure 3. Fine-tune ViT architecture

Figure 4 depicts the accuracy and loss curves over 50 training iterations. Both training and testing accuracies converge at approximately 0.98, with steadily decreasing loss values and closely aligned curves for training and testing. These results indicate strong model performance without overfitting.

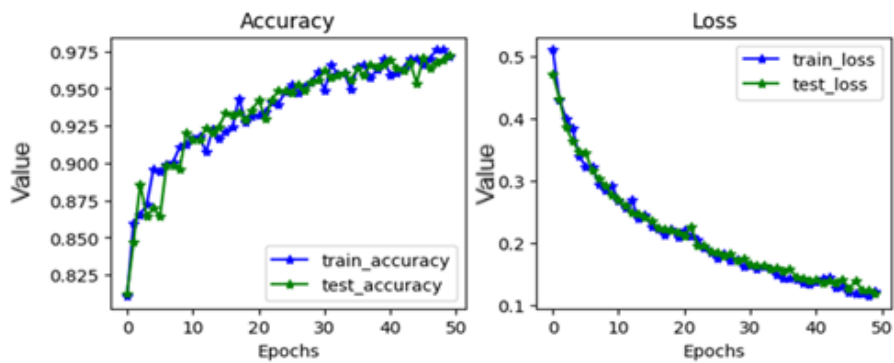


Figure 4. Fine-tune ViT accuracy/loss plots

Figure 5 presents the confusion matrix and classification report, showcasing the model's performance across the four pneumonia classes. High diagonal values in the confusion matrix indicate accurate predictions for all categories. The classification report further confirms these results, with an overall accuracy of 0.98 and high precision, recall, and F1-scores for all classes. Notably, fungal pneumonia predictions achieve flawless scores, underscoring the model's ability to handle imbalanced datasets effectively.

Table 4 compares the proposed model's performance with other deep-learning approaches for pneumonia detection. While most prior studies address only two- or three-class problems, the fine-tuned ViT achieves 98% accuracy across four classes, matching or exceeding the performance of existing methods. This demonstrates the ViT's capability to generalize and classify pneumonia more comprehensively.

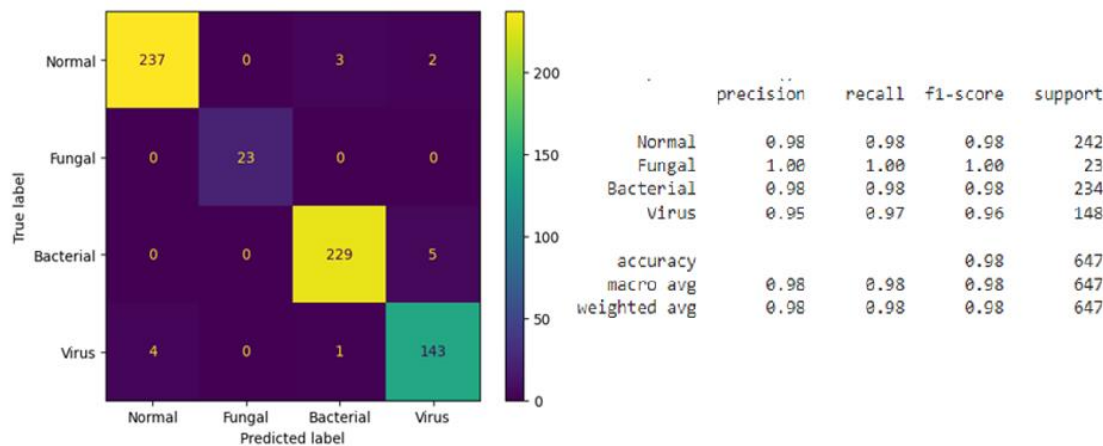


Figure 5. Confusion matrix and classification report

Table 4. Assessment of deep learning strategies				
Model	ACC (%)	P (%)	R (%)	F1 (%)
Singh <i>et al.</i> [1] [2-class]	0.97	0.96	0.97	0.97
Ali <i>et al.</i> [2] [3-class]	0.95	0.94	0.95	0.95
Gu and Lee [3] [2-class]	0.98	0.97	0.98	0.98
Putri and Al Maki [4] [3-class]	0.96	0.95	0.96	0.96
Asnake <i>et al.</i> [5] [2-class]	0.94	0.93	0.94	0.94
Proposed fine-tune ViT [4-class]	0.98	0.98	0.98	0.98

6. CONCLUSION

This research demonstrated that the fine-tuned ViT model achieved an impressive 98% accuracy, alongside 98% precision, recall, and F1-scores, across the four pneumonia classes—normal, fungal, bacterial, and viral. The ViT model’s ability to effectively differentiate between various forms of pneumonia highlights its improved capacity to capture complex patterns in chest X-ray images, made possible by its self-attention mechanism. These findings suggest that the fine-tuned ViT model is a promising tool for clinical application, potentially accelerating diagnosis and significantly improving accuracy, ultimately leading to better patient outcomes. In summary, the findings of this research have significant implications for both the research field and the healthcare community, providing a robust framework for the development of advanced, scalable, and reliable diagnostic tools in medical imaging.

Future research should focus on further enhancing the generalizability of the ViT model by utilizing larger and more diverse datasets that incorporate variations in demographics, imaging conditions, and disease stages. To improve performance, hybrid models combining the strengths of ViTs and CNNs could be explored, with a particular focus on better capturing both local and global information. Additionally, incorporating explainability techniques into the model’s decision-making process could increase physicians’ trust and facilitate wider adoption in healthcare. Finally, expanding the use of fine-tuned ViT models to other medical imaging modalities, such as CT and MRI, could open up new avenues for diagnosing a broader range of pulmonary and other diseases.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Khushboo Trivedi	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓			✓	
Chintan B. Thacker		✓				✓		✓	✓	✓	✓	✓		

- C : Conceptualization
- M : Methodology
- So : Software
- Va : Validation
- Fo : Formal analysis
- I : Investigation
- R : Resources
- D : Data Curation
- O : Writing - Original Draft
- E : Writing - Review & Editing
- Vi : Visualization
- Su : Supervision
- P : Project administration
- Fu : Funding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

INFORMED CONSENT

We have obtained informed consent from all individuals included in this study.

ETHICAL APPROVAL

This research did not involve the use of human participants or animals. Therefore, ethical approval was not required.

DATA AVAILABILITY

Data availability is not applicable to this paper as no new data were created or analyzed in this study.

REFERENCES

[1] S. Singh, M. Kumar, A. Kumar, B. K. Verma, K. Abhishek, and S. Selvarajan, “Efficient pneumonia detection using vision transformers on chest X-rays,” *Scientific Reports*, vol. 14, no. 1, Jan. 2024, doi: 10.1038/s41598-024-52703-2.





[2] A. M. Ali, K. Ghafoor, A. Mulahuwaish, and H. Maghdid, “COVID-19 pneumonia level detection using deep learning algorithm and transfer learning,” *Evolutionary Intelligence*, vol. 17, no. 2, pp. 1035–1046, Sep. 2022, doi: 10.1007/s12065-022-00777-0.

- [3] C. Gu and M. Lee, "Deep transfer learning using real-world image features for medical image classification, with a case study on pneumonia X-ray images," *Bioengineering*, vol. 11, no. 4, p. 406, Apr. 2024, doi: 10.3390/bioengineering11040406.
- [4] K. A. Putri and W. Fawwaz Al Maki, "Enhancing pneumonia disease classification using genetic algorithm-tuned DCGANs and VGG-16 integration," *Journal of Electronics, Electromedical Engineering, and Medical Informatics*, vol. 6, no. 1, Dec. 2023, doi: 10.35882/jeeemi.v6i1.349.
- [5] N. W. Asnake, A. O. Salau, and A. M. Ayalew, "X-ray image-based pneumonia detection and classification using deep learning," *Multimedia Tools and Applications*, vol. 83, no. 21, pp. 60789–60807, Jan. 2024, doi: 10.1007/s11042-023-17965-4.
- [6] T. S. Arulananth, S. W. Prakash, R. K. Ayyasamy, V. P. Kavitha, P. G. Kuppusamy, and P. Chinnasamy, "Classification of paediatric pneumonia using modified DenseNet-121 deep-learning model," *IEEE Access*, vol. 12, pp. 35716–35727, 2024, doi: 10.1109/access.2024.3371151.
- [7] C. J. Ghia and G. S. Rambhad, "Systematic review and meta-analysis of comorbidities and associated risk factors in Indian patients of community-acquired pneumonia," *SAGE Open Medicine*, vol. 10, Jan. 2022, doi: 10.1177/20503121221095485.
- [8] V. Rajaguru, T. H. Kim, J. Shin, S. G. Lee, and W. Han, "Ability of the LACE index to predict 30-day readmissions in patients with acute myocardial infarction," *Journal of Personalized Medicine*, vol. 12, no. 7, p. 1085, Jun. 2022, doi: 10.3390/jpm12071085.
- [9] M. O. Lewis, P. T. Tran, Y. Huang, R. A. Desai, Y. Shen, and J. D. Brown, "Disease severity and risk factors of 30-day hospital readmission in pediatric hospitalizations for pneumonia," *Journal of Clinical Medicine*, vol. 11, no. 5, p. 1185, Feb. 2022, doi: 10.3390/jcm11051185.
- [10] C. Ieracitano *et al.*, "A fuzzy-enhanced deep learning approach for early detection of Covid-19 pneumonia from portable chest X-ray images," *Neurocomputing*, vol. 481, pp. 202–215, Apr. 2022, doi: 10.1016/j.neucom.2022.01.055.
- [11] M. Rostami and M. Oussalah, "A novel explainable COVID-19 diagnosis method by integration of feature selection with random forest," *Informatics in Medicine Unlocked*, vol. 30, p. 100941, 2022, doi: 10.1016/j.imu.2022.100941.
- [12] A. I. Aviles-Rivero, P. Sellars, C.-B. Schönlieb, and N. Papadakis, "GraphXCOVID: Explainable deep graph diffusion pseudo-Labeling for identifying COVID-19 on chest X-rays," *Pattern Recognition*, vol. 122, p. 108274, Feb. 2022, doi: 10.1016/j.patcog.2021.108274.
- [13] A. Malhotra *et al.*, "Multi-task driven explainable diagnosis of COVID-19 using chest X-ray images," *Pattern Recognition*, vol. 122, p. 108243, Feb. 2022, doi: 10.1016/j.patcog.2021.108243.
- [14] A. K. Mondal, A. Bhattacharjee, P. Singla, and A. P. Prathosh, "xViTCOS: explainable vision transformer based COVID-19 screening using radiography," *IEEE Journal of Translational Engineering in Health and Medicine*, vol. 10, pp. 1–10, 2022, doi: 10.1109/jtehm.2021.3134096.
- [15] H. Ren *et al.*, "Interpretable pneumonia detection by combining deep learning and explainable models with multisource data," *IEEE Access*, vol. 9, pp. 95872–95883, 2021, doi: 10.1109/access.2021.3090215.
- [16] F. Colavita *et al.*, "COVID-19 rapid antigen test as screening strategy at points of entry: experience in lazio region, Central Italy, August–October 2020," *Biomolecules*, vol. 11, no. 3, p. 425, Mar. 2021, doi: 10.3390/biom11030425.
- [17] M. Mustafa Hellou *et al.*, "Nucleic acid amplification tests on respiratory samples for the diagnosis of coronavirus infections: a systematic review and meta-analysis," *Clinical Microbiology and Infection*, vol. 27, no. 3, pp. 341–351, Mar. 2021, doi: 10.1016/j.cmi.2020.11.002.
- [18] M. Toğaçar, N. Muzoğlu, B. Ergen, B. S. B. Yarman, and A. M. Halefoğlu, "Detection of COVID-19 findings by the local interpretable model-agnostic explanations method of types-based activations extracted from CNNs," *Biomedical Signal Processing and Control*, vol. 71, 2022, doi: 10.1016/j.bspc.2021.103128.
- [19] D. Mahapatra, Z. Ge, and M. Reyes, "Self-supervised generalized zero shot learning for medical image classification using novel interpretable saliency maps," *IEEE Transactions on Medical Imaging*, vol. 41, no. 9, pp. 2443–2456, Sep. 2022, doi: 10.1109/tmi.2022.3163232.
- [20] M. Li *et al.*, "Explainable COVID-19 infections identification and delineation using calibrated pseudo labels," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 7, no. 1, pp. 26–35, Feb. 2023, doi: 10.1109/tetci.2022.3189054.
- [21] "WHO director-general's opening remarks at the media briefing on COVID-19," *World Health Organisation*. Accessed: Mar. 11, 2020. [Online]. Available: <https://www.who.int/director-general/speeches/detail>.
- [22] G. Liang and L. Zheng, "A transfer learning method with deep residual network for pediatric pneumonia diagnosis," *Computer Methods and Programs in Biomedicine*, vol. 187, Apr. 2020, doi: 10.1016/j.cmpb.2019.06.023.
- [23] T. Fujioka *et al.*, "Efficient anomaly detection with generative adversarial network for breast ultrasound imaging," *Diagnostics*, vol. 10, no. 7, p. 456, Jul. 2020, doi: 10.3390/diagnostics10070456.
- [24] A. Zaman, S. H. Park, H. Bang, C. Park, I. Park, and S. Joung, "Generative approach for data augmentation for deep learning-based bone surface segmentation from ultrasound images," *International Journal of Computer Assisted Radiology and Surgery*, vol. 15, no. 6, pp. 931–941, May 2020, doi: 10.1007/s11548-020-02192-1.
- [25] P. Dimitrakopoulos, G. Sfikas, and C. Nikou, "ISING-GAN: annotated data augmentation with a spatially constrained generative adversarial network," in *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, Apr. 2020, pp. 1600–1603, doi: 10.1109/isbi45749.2020.9098618.
- [26] I. Arvidsson, N. C. Overgaard, K. Astrom, and A. Heyden, "Comparison of different augmentation techniques for improved generalization performance for gleason grading," in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, Apr. 2019, pp. 923–927, doi: 10.1109/isbi.2019.8759264.
- [27] Y.-J. Lin and I.-F. Chung, "Medical data augmentation using generative adversarial networks: X-ray image generation for transfer learning of hip fracture detection," in *2019 International Conference on Technologies and Applications of Artificial Intelligence (TAAI)*, Nov. 2019, pp. 1–5, doi: 10.1109/taai48200.2019.8959908.
- [28] V. Bhagat and S. Bhaumik, "Data augmentation using generative adversarial networks for pneumonia classification in chest Xrays," in *2019 Fifth International Conference on Image Information Processing (ICIIP)*, Nov. 2019, pp. 574–579, doi: 10.1109/ICIIP47207.2019.8985892.
- [29] P. Rajpurkar *et al.*, "Deep learning for chest radiograph diagnosis: a retrospective comparison of the CheXNeXt algorithm to practicing radiologists," *PLOS Medicine*, vol. 15, no. 11, Nov. 2018, doi: 10.1371/journal.pmed.1002686.
- [30] H. Q. Nguyen *et al.*, "VinDr-CXR: an open dataset of chest X-rays with radiologist's annotations," *Scientific Data*, vol. 9, no. 1, p. 429, Jul. 2022, doi: 10.1038/s41597-022-01498-w.
- [31] S. Islam *et al.*, "Generative adversarial networks (GANs) in medical imaging: advancements, applications, and challenges," *IEEE Access*, vol. 12, pp. 35728–35753, 2024, doi: 10.1109/ACCESS.2024.3370848.
- [32] J. Gui, Z. Sun, Y. Wen, D. Tao, and J. Ye, "A review on generative adversarial networks: algorithms, theory, and applications," *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 4, pp. 3313–3332, Apr. 2023, doi: 10.1109/TKDE.2021.3130191.





- [33] G. M. M. Alshmrani, Q. Ni, R. Jiang, H. Pervaiz, and N. M. Elshennawy, "A deep learning architecture for multi-class lung diseases classification using chest X-ray (CXR) images," *Alexandria Engineering Journal*, vol. 64, pp. 923–935, Feb. 2023, doi: 10.1016/j.aej.2022.10.053.
- [34] R. Arora *et al.*, "AI-based diagnosis of COVID-19 patients using X-ray scans with stochastic ensemble of CNNs," *Physical and Engineering Sciences in Medicine*, vol. 44, no. 4, pp. 1257–1271, Dec. 2021, doi: 10.1007/s13246-021-01060-9.
- [35] Y. Guan *et al.*, "Comparison of deep-learning and radiomics-based machine-learning methods for the identification of chronic obstructive pulmonary disease on low-dose computed tomography images," *Quantitative Imaging in Medicine and Surgery*, vol. 14, no. 3, pp. 2485–2498, Mar. 2024, doi: 10.21037/qims-23-1307.
- [36] Q. An, W. Chen, and W. Shao, "A deep convolutional neural network for pneumonia detection in X-ray images with attention ensemble," *Diagnostics*, vol. 14, no. 4, p. 390, Feb. 2024, doi: 10.3390/diagnostics14040390.
- [37] S. prasad Koyyada and T. P. Singh, "An explainable artificial intelligence model for identifying local indicators and detecting lung disease from chest X-ray images," *Healthcare Analytics*, vol. 4, p. 100206, Dec. 2023, doi: 10.1016/j.health.2023.100206.
- [38] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers, "ChestX-ray8: hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jul. 2017, pp. 3462–3471, doi: 10.1109/CVPR.2017.369.
- [39] P. Mooney, "Chest X-ray images (pneumonia)," *Kaggle*. 2017, Accessed: Dec. 10, 2024. [Online]. Available: <https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia/version/2>.
- [40] J. Irvin *et al.*, "CheXpert: a large chest radiograph dataset with uncertainty labels and expert comparison," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. 01, pp. 590–597, Jul. 2019, doi: 10.1609/aaai.v33i01.3301590.
- [41] S. Wakrim, "Diagnostic performance of chest CT findings of COVID-19 with RT-PCR negative," *African Health Sciences*, vol. 22, no. 4, pp. 502–504, Dec. 2022, doi: 10.4314/ahs.v22i4.56.

BIOGRAPHIES OF AUTHORS



Khushboo Trivedi     holds a master's degree in computer science and engineering from Parul University and currently pursuing Ph.D. in the domain of deep learning and computer vision from Parul University, Vadodara, Gujarat. She has 12+ years of experience in academia. Her research interests are in machine learning, deep learning, artificial intelligence, and computer vision. Currently, she serves as an assistant professor in the computer science and engineering Department at Parul Institute of Technology, Parul University. She can be contacted at email: khushboo.trivedi21305@paruluniversity.ac.in.



Chintan B. Thacker     received a Ph.D. degree in the domain of artificial intelligence and computer vision from Gujarat Technical University in the year 2022. He had served as Head of the Department of Computer Science and Engineering Department at HJD Institute of Technical Education and Research, Kera, India. He has 1+ years of experience in industry and 12+ years of experience in academia. Currently, he serves as an assistant professor in the Computer Science and Engineering Department at Parul Institute of Engineering Technology, Parul University, Vadodara, Gujarat. In addition, he has also guided several doctorate students and has been active in conducting several workshops in the domain of computer vision. His research interests are in machine learning, artificial intelligence, deep learning and computer vision. He can be contacted at email: chintan.thacker19435@paruluniversity.ac.in.