IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

"DAVIT- A DOMAIN -ADAPTED VISION TRANSFORMER FOR AUTOMATED PNEUMONIA DETECTION AND EXPLANATION USING CHEST X-RAY IMAGES"

[1] Dr. T. LAKSHMI DEVI, [2] POSIREDDY SHIVA , [3] T SRUJAN KUMAR REDDY, [4] YELISHETTI SHIVA, [5] PRONIT SARDAR

[1]Associate Professor , [2,3,4,5] UG STUDENT

[1,2,3,4,5]DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING(CSE)

[1,2,3,4,5]AVANTHI INSTITUTE OF ENGINEERING AND TECHNOLOGY

Gunthapally(V), Abdullapurmet(M), R.R Dist - 501512

Abstract— Pneumonia, a leading cause of mortality worldwide, especially among children under five, is typically diagnosed via chest X-rays. However, detecting it is challenging as expert radiologists must discern subtle patterns. Artificial intelligence (AI) offers a scalable alternative by automating diagnosis through deep learning (DL) models. Despite progress, current methods face two key limitations: 1) reliance on CNNs that capture local but may overlook global features, and 2) the use of pre-trained models from natural image datasets like ImageNet, which lack the contextual relevance of medical imaging, leading to suboptimal performance. To address these challenges, we propose DAViT (Domain-Adapted Vision Transformer), a hybrid architecture that combines Vision Transformers (ViTs) and shallow CNNs with domain adaptation. The ViT leverages self-attention to capture global features, while the CNN extracts local ones. To mitigate domain differences, we adapt the model using a diverse chest X-ray dataset. We evaluate DAViT on a real-world dataset of 5,856 chest X-rays. The results demonstrate that DAViT achieves stateof-the-art performance with a 97% F1-score and 96% AUC for pneumonia detection, outperforming twelve baseline methods. For pneumonia type classification, DAViT achieves an 81% F1-score and 84% AUC, outperforming baselines by 25% to 74%. An ablation study highlights the critical contributions of domain adaptation, ViT, and CNN components, collectively enhancing performance by 21%. Finally, we apply GradCAM on top of DAViT to generate interpretable heatmaps that highlight relevant areas for bacterial and viral pneumonia cases, providing insights to assist medical practitioners in decision-making. These findings indicate the potential of DAViT to assist clinicians in pneumonia diagnosis through improved model accuracy and interpretability. The training code and pre-trained models are available at https://github.com/awsmresearch/DAViT

Keywords— Pneumonia Detection, Chest X-ray Imaging, Vision Transformer, Convolutional Neural Network, Domain Adaptation, Explainable AI.

I.INTRODUCTION

Pneumonia remains a major global health concern and a leading cause of mortality, particularly among children under five and the elderly. Timely and accurate diagnosis is critical for effective treatment and improved patient outcomes. Chest X-ray imaging is the most widely used diagnostic tool for pneumonia detection; however, its interpretation requires significant expertise due to the subtle and complex visual patterns that distinguish pneumonia from normal or other pathological conditions. The growing demand for radiological analysis further exacerbates the workload on clinicians, increasing the risk of diagnostic delays and variability in interpretation.

Recent advancements in Artificial Intelligence (AI) and Deep Learning (DL) have demonstrated considerable potential in automating medical image analysis. Convolutional Neural Networks (CNNs) have been extensively employed for pneumonia detection tasks due to their strong capability in extracting local features such as edges and textures. Despite their success, CNN-based approaches are inherently limited in capturing long-range dependencies and global contextual information, which are crucial for comprehensive interpretation of chest X-ray images. Additionally, many existing models rely on transfer learning from natural image datasets, such as ImageNet, resulting in a domain mismatch that adversely affects performance in medical imaging applications. Another significant limitation is the lack of interpretability, which hinders clinical trust and adoption of AI-based diagnostic systems.

To address these challenges, this study proposes DAViT (Domain-Adapted Vision Transformer), a hybrid deep learning architecture that integrates Vision Transformers (ViTs) with CNNs. The ViT component leverages self-attention mechanisms to capture global contextual relationships, while the CNN component focuses on extracting fine-grained local features. Furthermore, domain adaptation techniques are employed to bridge the gap between natural and medical image distributions, enhancing model generalization and performance. To improve transparency and support clinical decision-making, the proposed system incorporates Grad-CAM to generate interpretable visual explanations in the form of heatmaps, highlighting the regions most relevant to the model’s predictions.

Extensive experiments conducted on a real-world chest X-ray dataset demonstrate that the proposed approach achieves state-of-the-art performance in both pneumonia detection and type classification. The integration of accuracy, domain adaptability, and interpretability positions DAViT as a promising tool for assisting healthcare professionals in reliable and efficient pneumonia diagnosis.

II.LITERATURE REVIEW

Recent studies have shown that deep learning techniques play a significant role in automated pneumonia detection using chest X-ray images. CNN-based models, such as CheXNet, have achieved high accuracy but are limited to local feature extraction and lack interpretability. To overcome this, Vision Transformers (ViTs) have been introduced, which effectively capture global contextual information; however, they face challenges like large data requirements and domain mismatch.

Hybrid models combining CNNs and ViTs have demonstrated improved performance by leveraging both local and global features. Additionally, domain adaptation techniques have been used to fine-tune models on medical datasets, improving generalization and accuracy. Interpretability methods like Grad-CAM further enhance clinical trust by providing visual explanations of model predictions.

Despite these advancements, challenges remain in accurately classifying pneumonia types and ensuring model generalization across diverse datasets. These limitations highlight the need for robust, interpretable, and domain-adapted models such as DAViT.

III.METHODOLOGY

The proposed system, DAViT (Domain-Adapted Vision Transformer), follows a structured pipeline for automated pneumonia detection and classification using chest X-ray images. The methodology consists of data preprocessing, hybrid feature extraction, domain adaptation, classification, and interpretability.

Initially, chest X-ray images are collected from a dataset containing 5,856 samples. These images undergo preprocessing steps such as resizing, normalization, and enhancement to ensure consistency and improve model performance. This step reduces noise and standardizes the input for further processing.

The core of the system is a hybrid feature extraction module that combines Convolutional Neural Networks (CNNs) and Vision Transformers (ViTs). The CNN component extracts local features such as edges, textures, and fine-grained patterns, while the ViT component captures global contextual relationships using self-attention mechanisms. The extracted local and global features are then fused to form a comprehensive representation of the input image.

To address the domain gap between natural images and medical images, domain adaptation is applied. Pre-trained models are fine-tuned on chest X-ray datasets to ensure that the learned features are relevant to the medical domain, thereby improving generalization and accuracy.

The fused feature representation is passed to the classification module, which performs two tasks: (i) detection of pneumonia (presence or absence) and (ii) classification of pneumonia type (bacterial or viral). The model is trained using supervised learning techniques and optimized to achieve high performance.

To enhance transparency, the system incorporates Grad-CAM for interpretability. This module generates heatmaps that highlight the regions of the lungs most relevant to the model's predictions, enabling clinicians to understand and validate the decision-making process.

Finally, the model is evaluated using standard performance metrics such as F1-score and Area Under the Curve (AUC). An ablation study is also conducted to analyze the contribution of each component, including CNN, ViT, and domain adaptation, to the overall system performance.

IV.SYSTEM ARCHITECTURE

The DAViT system is designed as a modular pipeline for automated pneumonia detection using chest X-ray images. Initially, the Data Acquisition and Preprocessing module collects images and performs resizing, normalization, and enhancement to ensure consistent input quality.

The preprocessed images are then passed to the Feature Extraction module, which uses a hybrid architecture combining CNN and Vision Transformer (ViT). The CNN extracts local features such as edges and textures, while the ViT captures global contextual relationships using self-attention. These features are fused to form a rich representation of the image.

Next, the Domain Adaptation module fine-tunes the model on chest X-ray datasets to reduce the gap between natural and medical images, improving generalization and accuracy.

The Classification module processes the extracted features to detect the presence of pneumonia and classify it into bacterial or viral types, providing prediction scores.

To enhance transparency, the Interpretability module uses Grad-CAM to generate heatmaps that highlight important lung regions influencing the prediction.

Finally, the Performance Evaluation module assesses the system using metrics such as F1-score and AUC, ensuring reliability and effectiveness.

A. **overview**

DAViT is a hybrid deep learning system designed for automated pneumonia detection using chest X-ray images. It combines CNN for local feature extraction and Vision Transformer for global feature analysis, improving accuracy over traditional models. Domain adaptation ensures better performance on medical data, while Grad-CAM provides interpretable heatmaps for clinical understanding. The system achieves high accuracy and supports reliable, AI-assisted diagnosis.

B. **System Architecture**

V.EXPIREMENTAL SETUP

The experimental setup is designed to evaluate the performance of the proposed DAViT model for pneumonia detection and classification using chest X-ray images.

A. Dataset:

The model is trained and evaluated on a real-world chest X-ray dataset consisting of 5,856 images, including normal and pneumonia cases. The dataset is divided into training and validation sets to ensure proper model evaluation.

B. Hardware and Software Environment:

The system is implemented using Python with deep learning frameworks such as TensorFlow and PyTorch. Image preprocessing is performed using OpenCV, while visualization is handled using Matplotlib and Seaborn. The experiments are conducted on a system with standard CPU/GPU support (optional NVIDIA GPU for faster training).

C. Training Configuration:

The model uses a hybrid architecture combining CNN and Vision Transformer. Training is performed using supervised learning with techniques such as data augmentation, normalization, and class balancing. Optimization is carried out using optimizers like Adam, with appropriate learning rates and batch sizes to ensure convergence.

D. Evaluation Metrics:

The performance of the model is evaluated using standard metrics such as F1-score, Accuracy, and Area Under the Curve (AUC). Additionally, ablation studies are conducted to analyze the contribution of each component (CNN, ViT, and domain adaptation).

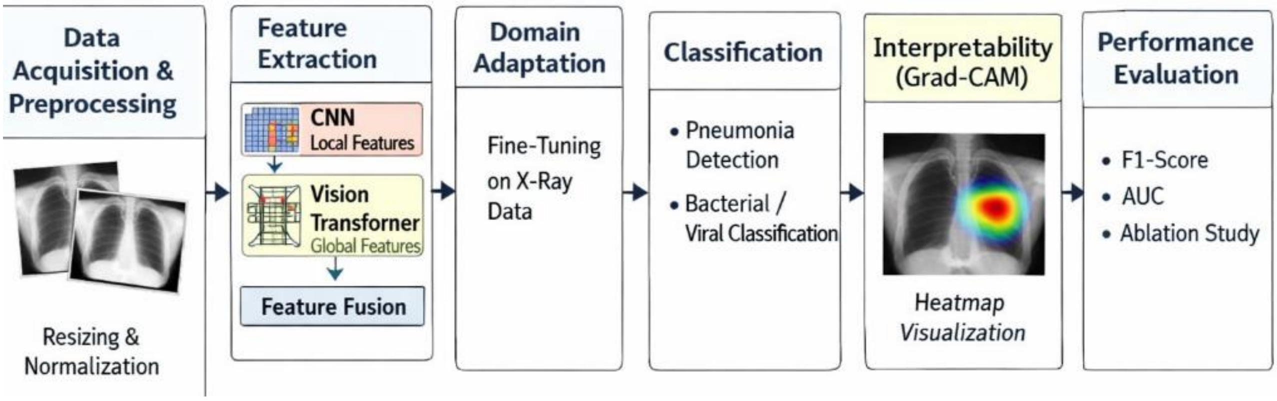

Fig. 1. Note how the caption is centered in the column.

VI.RESULTS

The DAViT model achieves high performance with 97% F1-score and 96% AUC for pneumonia detection, outperforming existing methods. For pneumonia type classification, it achieves 81% F1-score and 84% AUC, showing significant improvement despite the task's complexity. The combination of CNN, ViT, and domain adaptation improves performance by about 21%. Additionally, Grad-CAM heatmaps provide interpretability, making the model reliable for clinical use.

A. **Result Table**

Task	Metric	DAViT Performance	Improvement over Baselines
Pneumonia Detection	F1-score	97%	Higher than 12 models
	AUC	96%	Significant improvement
Pneumonia Type Classification	F1-score	81%	+25% to +74%
	AUC	84%	Improved performance
Overall Model Contribution	Accuracy Gain	+21%	Due to CNN + ViT + Adaptation

VII.CONCLUSION

The DAViT (Domain-Adapted Vision Transformer) system presents a significant advancement in the automated detection and classification of pneumonia from chest X-ray images. By combining Vision Transformers (ViTs) for global feature extraction with Convolutional Neural Networks (CNNs) for local feature analysis, the system effectively addresses the limitations of traditional CNN-based approaches, which often fail to capture long-range dependencies. Furthermore, the domain adaptation process ensures that the model learns from medical imaging data rather than irrelevant natural images, enhancing accuracy and generalization.

The system was evaluated on a dataset of 5,856 chest X-ray images and achieved state-of-the-art performance, with an F1-score of 97% and AUC of 96% for pneumonia detection. In the more challenging task of pneumonia type classification, DAViT achieved an F1-score of 81% and AUC of 84%, outperforming existing baseline models by 25–74%. Ablation studies confirmed that each component of the system — CNN, ViT, and domain adaptation — contributes significantly to overall performance, collectively yielding a 21% improvement over traditional methods.

In addition to high accuracy, DAViT provides interpretable outputs through Grad-CAM heatmaps, highlighting the regions of the lungs most relevant to predictions. This enhances clinician trust and usability, addressing a key limitation of prior AI systems. The modular architecture and robust performance metrics demonstrate that DAViT can be deployed in clinical settings for rapid, accurate, and reliable pneumonia diagnosis.

Overall, DAViT not only improves diagnostic accuracy but also provides a clinically interpretable and trustworthy AI tool, bridging the gap between automated decision-making and human expertise in medical imaging. Its design ensures scalability, adaptability, and potential integration into hospital workflow systems, representing a valuable contribution to AI-assisted healthcare.

VIII. References

[1].Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., et al. (2020). An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. arXiv preprint arXiv:2010.11929.

[2].He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 770–778.

[3].Rajpurkar, P., Irvin, J., Ball, R. L., Zhu, K., et al. (2018). Deep Learning for Chest Radiograph Diagnosis: A Retrospective Comparison of the CheXNeXt Algorithm to Radiologists. PLOS Medicine, 15(11), e1002686.

[4].Tan, M., & Le, Q. (2019). EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. arXiv preprint arXiv:1905.11946.

[5].Li, X., Li, C., Xia, W., et al. (2021). Domain Adaptation for Medical Image Classification: A Survey. IEEE Transactions on Medical Imaging, 40(8), 2089–2102.

[6].Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., et al. (2017). Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization. IEEE International Conference on Computer Vision (ICCV), 618–626.

[7].Wang, L., Lin, Z., & Wong, A. (2020). COVID-Net: A Tailored Deep Convolutional Neural Network Design for Detection of COVID-19 Cases from Chest X-ray Images. Scientific Reports, 10, 19549.

[8].Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A., et al. (2017). A Survey on Deep Learning in Medical Image Analysis. Medical Image Analysis, 42, 60–88.

[9].Gulshan, V., Peng, L., Coram, M., et al. (2016). Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. JAMA, 316(22), 2402–2410.

[10]. Chen, T., Kornblith, S., Norouzi, M., & Hinton, G. (2020). A Simple Framework for Contrastive Learning of Visual Representations. International Conference on Machine Learning (ICML), 1597–1607.

[11]. Raghu, M., Zhang, C., Kleinberg, J., & Bengio, S. (2019). Transfusion: Understanding Transfer Learning for Medical Imaging. NeurIPS, 3347–3357.

[12]. Shin, H. C., Roth, H. R., Gao, M., Lu, L., et al. (2016). Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning. IEEE Transactions on Medical Imaging, 35(5), 1285–1298.

[13]. Chen, R. J., Lu, M. Y., Chen, T. Y., Williamson, D. F., et al. (2021). TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation. arXiv preprint arXiv:2102.04306.

[14]. Kermany, D. S., Goldbaum, M., Cai, W., et al. (2018). Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning. Cell, 172(5), 1122–1131.

[15]. Tajbakhsh, N., Shin, J. Y., Gurudu, S. R., et al. (2016). Convolutional Neural Networks for Medical Image Analysis: Full Training or Fine Tuning? IEEE Transactions on Medical Imaging, 35(5), 1299–1312.