JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Pneumonia Detection Using Hybrid Deep Learning: ResNet34 and MaxViT with Attention Mechanism

Ayush
Under The Guidance of: Mr. Ashish Verma

Abstract

Pneumonia detection through chest X-ray (CXR) imaging is a crucial diagnostic process that requires expert evaluation. However, manual interpretation is time-consuming, subjective, and prone to errors. This research presents a novel **hybrid deep learning model** that integrates **ResNet34 and MaxViT** architectures along with an **attention-based feature fusion mechanism** to improve pneumonia classification accuracy. The **ResNet34 model** extracts **local features**, while **MaxViT captures global dependencies**, providing a robust representation of lung abnormalities. The attention mechanism optimally fuses these features, ensuring the model emphasizes the most critical pneumonia-related patterns. Performance evaluation on a large CXR dataset shows that our approach significantly outperforms previous models, achieving **99.04% accuracy, 0.9750 Kappa score, 99.48% sensitivity, and 97.78% specificity**. The results suggest that this **hybrid model can assist radiologists** in automating pneumonia diagnosis with high reliability, particularly in resource-constrained settings.

Keywords— Pneumonia Detection, Deep Learning, ResNet34, MaxViT, Attention Mechanism, Medical Imaging, Chest X-ray.

1. Introduction

A significant number of deaths due to pneumonia occur in locations with limited resources and an inadequate number of radiologists. Manually analysing chest X-rays is a time-consuming and subjective process, yet it is the principal diagnostic tool for pneumonia. As a result of advancements in artificial intelligence and deep learning, automated computed tomography (CXR) processing has emerged as a potential new tool for the accurate and early detection of pneumonia.

Using classic CNNs, like as ResNet34, has substantially improved the classification of medical pictures. Since CNNs are mostly effective at catching local characteristics, it is possible that they may miss global structural problems in pneumonia patients.**Vision Transformers (ViTs)**, particularly **MaxViT**, can model long-range dependencies across an image but often require large datasets and high computational resources. This research **combines both architectures** to **leverage their complementary strengths**, ensuring comprehensive feature extraction.

Additionally, **attention-based feature fusion** is introduced to intelligently combine CNN and transformer features, ensuring that only the most **clinically relevant patterns** contribute to decision-making. By incorporating **advanced optimization techniques**, **test-time ensembling**, and **class imbalance handling**, this research **improves pneumonia classification accuracy and robustness**, making it suitable for real-world deployment.

2. Methodology

Datasets

Chest X-ray images of paediatric children ranging in age from one to five were included in the samples retrieved from the retrospective cohort dataset from Guangzhou Women and Children's Medical Centre, as shown in Figure 1. It was from this dataset that the Kaggle pneumonia dataset was built. Normal and pneumonia X-rays taken as part of routine clinical practice are both included in this collection. The training dataset consists of 13,411 normal samples and 3,875, pneumonia samples. Of the 390 pneumonia samples that were analysed, 234 were found to be within normal limits. Given its small size of just 16 images, the Kaggle validation dataset was not used for any studies. Based on the classifications, we divided the dataset in two using an 80:20 split: one for training and one for validation.

2.1 Data Preprocessing and Augmentation

The dataset consists of labeled CXR images categorized as **Normal** or **Pneumonia**. To enhance generalization, we apply:

- **Resizing (224×224 pixels)**
- **Normalization (ImageNet mean and std)**
- **Augmentations**: Random Horizontal Flip, Random Rotation (30°), Color Jitter, and Random Erasing
- **Class Balance Strategies**: Synthetic data augmentation and weighted sampling

2.2 Model Architecture

Hybrid Architecture = ResNet34 (CNN) + MaxViT (Transformer) + Attention Fusion

- **ResNet34**: Extracts fine-grained **local features** using residual connections, enhancing depth without degradation.
- **MaxViT**: Captures **global dependencies** using self-attention over image patches, modeling relationships across lung regions.
- **Attention Mechanism**: Fuses the **CNN and Transformer** outputs by dynamically weighting feature contributions based on relevance.

2.3 Training and Optimization

- **Loss Function**: CrossEntropy with **Label Smoothing (0.1)** to prevent overconfidence.
- **Optimizer**: **Adam** with **CosineAnnealingLR** for dynamic learning rate adaptation.
- **Batch Size**: 8, ensuring computational efficiency.
- **Test-Time Ensembling (TTA)**: Multiple augmented versions of the test image are evaluated, and results are aggregated for improved robustness.

3. Results and Performance Analysis

Our model achieves **significant improvements** over previous CNN-based approaches:

Metric	**This Research**
Accuracy	**99.04%**
Kappa Score	**0.9750**
Sensitivity	**0.9948**
Specificity	**97.78%**
PPV (Precision)	**99.23%**

- **Sensitivity remains near 1.0**, meaning pneumonia cases are detected reliably.
- **Specificity improved**, reducing false positives, ensuring accurate classification.
- **Higher Kappa Score** reflects strong agreement with expert radiologists.

Pneumonia Detected

4. Discussion

4.1 Comparison with Existing Methods

Traditional CNN-based models, such as **ResNet34 alone**, lack global context understanding, making them prone to misclassification. On the other hand, pure **transformer models (MaxViT alone)** require significantly more data and computational resources. Our **hybrid approach** effectively **balances local and global feature extraction**, leading to improved generalization.

4.2 Clinical Relevance

This model can be integrated into **telemedicine platforms** and **hospital information systems (HIS)**, assisting radiologists in diagnosing pneumonia efficiently. The **attention-based fusion mechanism** ensures that model decisions align with **clinically relevant features**, enhancing trust in AI-driven diagnostics.

5. Conclusion and Future Scope

This research presents an **optimized hybrid deep learning model** for pneumonia detection, integrating **CNN-based local feature extraction with transformer-based global feature analysis**. The inclusion of an **attention-**

based fusion mechanism significantly improves model performance. Compared to previous methods, our approach demonstrates **higher accuracy (99.04%) and specificity (97.78%)**, reducing false positives and improving clinical applicability.

Future Work

1. **Multi-Disease Classification**: Expanding the model to detect other lung diseases such as **COVID-19, Tuberculosis, and Lung Cancer**.
2. **Real-Time Deployment**: Optimizing the model for **edge AI devices** like mobile phones and IoT-enabled hospital equipment.
3. **Explainability & XAI Integration**: Implementing **Grad-CAM and SHAP** for model interpretability.
4. **Federated Learning for Privacy**: Training models across hospitals without sharing patient data to comply with **HIPAA regulations**.

References

[1] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition*. [2] Dosovitskiy, A., et al. (2020). An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. *arXiv preprint arXiv:2010.11929*. [3] Tan, M., & Le, Q. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. *International Conference on Machine Learning (ICML)*

[2] M. Badrish, K. G. N. Prabhanjali, and A. Raghuvira Pratap, “Comparative Analysis of CNN Models with Vision Transformer on Lung Infection Classification,” *SpringerLink*, 2020. https://doi.org/10.1007/.

[3] H. Sharma, J. S. Jain, P. Bansal, and S. Gupta, “Feature Extraction and Classification of Chest X- ray Images Using CNN to Detect Pneumonia,” *Confluence 2020*, 2020. https://doi.org/10.1109/Confluence47617.2020.9057809.

[4] D.-P. Fan, W. Zuo, F. Zhou, and L. Wang, “Inf-Net: Automatic COVID-19 Lung Infection Segmentation From CT Images,” *IEEE Trans Med Imaging*, 2020.

[5] E. F. Ohata, G. M. Bezerra, J. V. S. Chagas, A. V. L. Neto, A. B. Albuquerque, V. H. C. de Albuquerque, and P. P. Reboucas Filho, “Automatic Detection of COVID-19 Infection Using Chest X- Ray Images Through Transfer Learning,” *IEEE/CAA J Autom Sinica*, 2020.

[6] J. G. Lee, et al., "Deep Learning in Medical Imaging: General Overview," *Korean J Radiol*, 2017. https://doi.org/10.3348/kjr.2017.18.4.570.

[7] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” *CVPR*, 2016. https://doi.org/10.1109/CVPR.2016.90.

[8] S. V. Militante, N. V. Dionisio, and B. G. Sibbaluca, “Pneumonia and COVID-19 Detection Using Convolutional Neural Networks,” *IEEE ICVEE 2020*, 2020

[9] D. Srivastav, A. Bajpai, and P. Srivastava, “Improved Classification for Pneumonia Detection Using Transfer Learning with GAN-Based Synthetic Image Augmentation,” *Confluence 2021*, 2021. https://doi.org/10.1109/Confluence51648.2021.9377062.

[10] Z. Liu et al., “Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows,” *arXiv*, 2021.