As per UGC guidelines an electronic bar code is provided to secure your paper

International Journal for Modern Trends in Science and Technology
Volume 11, Issue 07, pages 59-64
ISSN: 2455-3778 online
Available online at: http://www.ijmtst.com/vol11issue07.html
DOI: https://doi.org/10.5281/zenodo.15770273

Efficient Pneumonia Detection Using Vision Transformers on Chest X-Rays

MD.Baig Mohammad, Ch.Jagan, A.Durga Prasad, Sk. Raheem, K.Bhanu Teja

Department of Electronics and Communication Engineering, Andhra Loyola Institute of Engineering & Technology, Vijayawada, India.

To Cite this Article
MD.Baig Mohammad, Ch.Jagan, A.Durga Prasad, Sk. Raheem & K.Bhanu Teja (2025). Efficient Pneumonia Detection Using Vision Transformers on Chest X-Rays. International Journal for Modern Trends in Science and Technology, 11(07), 59-64. https://doi.org/10.5281/zenodo.15770273

Article Info
Received: 06 June 2025; Accepted: 26 June 2025.; Published: 29 June 2025.

KEYWORDS

Computer-aided diagnosis, Machine learning, Deep learning, Vision transformer Efficient neural networks, Chest X-ray

ABSTRACT

Pneumonia remains a significant health concern worldwide, with timely and accurate diagnosis being crucial for effective treatment and patient recovery. Traditional methods of pneumonia diagnosis using chest X-rays often rely on manual interpretation by radiologists, which is time- consuming and prone to human error. In this project, we propose an Efficient Pneumonia Detection System leveraging Vision Transformers (ViTs) to analyze chest X-ray images and enhance diagnostic accuracy. ViTs, unlike convolutional neural networks (CNNs), utilize a self-attention mechanism that captures both global and local dependencies, enabling superior feature extraction and classification. The proposed model is trained on a large dataset of labeled chest X-ray images, ensuring robustness and generalizability. Extensive experimentation is conducted to optimize hyperparameters and improve model efficiency, resulting in a highly accurate system capable of distinguishing between normal and pneumonia- affected cases. The results demonstrate that the Vision Transformer-based approach outperforms traditional CNN models in terms of precision, recall, and overall classification accuracy. This system has the potential to assist healthcare professionals by providing an automated, reliable, and scalable solution for early pneumonia detection, ultimately reducing diagnostic delays and improving patient care.

1. INTRODUCTION

Pneumonia is a lung infection caused by various pathogens, including viruses, bacteria, and fungi [1,2]. It affects either one or both lungs and results in the inflammation of lung parenchyma, i.e., the portion of the lung tissue that is responsible for gas exchange, including pulmonary alveoli. The inflammation causes the lung's alveoli to fill up with pus or fluid, therefore resulting in symptoms like shortness of breath, cough, chest pain, fever, etc. Common pathogens responsible for viral pneumonia include influenza viruses, respiratory syncytial virus, and the SARS-CoV-2 virus, while the most common pathogen of bacterial pneumonia is Streptococcus pneumoniae [3].

Pneumonia is a significant global health threat, leading to high mortality rates worldwide. It is a major cause of death of children (under five years) in developing countries as well as of elderly people (above 65 years) in developed countries [4,5,6]. Recently, the outbreak of SARS-CoV-2 virus has also caused much havoc across the world. Globally, up till the time of writing, there have been more than 702 million confirmed cases of the virus, resulting in around 6.9 million deaths [7]. In many cases, the SARS-CoV-2 virus leads to pneumonia in the infected person [8] and therefore requires the hospitalization of the patient. Since SARS-CoV-2 has a relatively high transmission rate with an R0 of around 2.5 for the original variant [9], the medical and healthcare systems of even developed countries have sometimes been overwhelmed by the high number of COVID-19 patients [10]. Due to the aforementioned problems, timely and accurate pneumonia diagnoses are needed for prompt curative treatment, which in turn helps mitigate the pneumonia-associated crisis.

Some common tools to diagnose pneumonia include chest X-rays (CXRs), chest computed tomography (CT) scans, magnetic resonance imaging (MRI) of the chest, chest ultrasound scans, etc. [11]. Despite CXRs' lower sensitivity for detecting pneumonia compared to that of some other diagnostic tools, like chest CT and chest ultrasound scans [12,13], CXRs are still considered the gold standard for diagnosing pneumonia according to most clinical guidelines globally [14]. Moreover, CXRs are very economical [15], easily accessible, and have very low radiation doses compared to CT scans [16]. For example, a CXR scan delivers only 0.1 millisievert (mSv) of ionized radiation to the patient compared to 7 mSv delivered by a chest CT scan; i.e., one's exposure to radiation is seventy times higher in the case of the CT scan. Compared to MRIs, CXRs also have some benefits. They have lower costs, are much quicker to perform, and are much more commonly and easily available even in resource-constrained parts of the world. For these reasons, CXRs are the most commonly performed radiological scans in the world [17].

In the context of the COVID-19 pandemic, CXRs enable the rapid triaging of patients during a COVID-19 wave [18]. This is because the gold- standard diagnostic test for SARS-CoV-2 is reverse transcription-polymerase chain reaction (RT-PCR) [19] but this is time-consuming, involves a laborious manual process [20], and also suffers from a low level of sensitivity [20]. The use of CXR-based diagnosis in parallel with RT-PCR (which takes longer) can help prioritize patients and improve survival rates. In addition, portable CXRs ensure patient isolation and help in preventing the spread of the virus [18,21]. All this is possible only because COVID-19 pneumonia has some unique manifestations on CXRs [21], which are different from other forms of pneumonia.

Unfortunately, despite all the aforementioned benefits of using CXRs as a tool to diagnose pneumonia, there is a dearth of expert radiologists [6,15,22], especially in developing countries, that can accurately interpret CXRs; i.e., there is often a serious imbalance between the number of patients and the number of available radiologists. Moreover, since the resolution of CXRs is lower than that of CT and MRI scans, there is always a chance that even an expert radiologist or clinician may miss out some important pattern or manifestation present in a CXR [6]. Computer-aided diagnostic (CAD) tools can help medical staff like radiologists and clinicians in pneumonia diagnosis. Deep-learning (DL)-based methods have been extensively evaluated as an underlying CAD technology for pneumonia diagnosis [1,2,6,8,11,15,18,23,24,25,26].

Over the past decade, DL methods have dramatically improved the state of the art in visual recognition tasks [27,28]. This includes diagnostic accuracy in medical imaging [29], such as pneumonia diagnosis in CXR. Unlike the traditional machine learning (ML) approach, DL methods do not require domain expertise or hand-crafted features. This is because in DL, features are learned from the data directly using a general-purpose learning procedure called backpropagation [27]. This enables automatic end-to-end feature extraction and image classification. A convolutional neural network (CNN) is a type of deep neural network that specializes in visual recognition tasks such as pneumonia diagnosis in CXRs [30]. An important challenge in exploiting CNNs for diagnostic tasks is the scarcity of high-quality correctly labeled large- size medical image datasets. Training on small datasets often leads to problems such as overfitting, poor generalization, etc. Moreover, training on poorly labeled datasets will also likely lead to misleading diagnoses. The public availability of some high-quality CXR datasets [1,31,32,33,34,35] has certainly helped promote research in the area of automated chest disease diagnoses. In some research studies [36,37], generative adversarial network (GAN)-based synthetic CXR generation techniques have been successfully exploited to overcome the problems of overfitting and poor generalization.

2. LITERATURE REVIEW

In recent years, the focus of research on diagnosing and categorizing lung diseases, including pneumonia, through medical imaging has intensified, driven by advances in machine learning and deep learning technologies. Precisely segmenting lung areas in chest X-ray (CXR) images is essential for reliable disease identification and thorough analysis. This section explores deep learning techniques for segmenting and diagnosing lung diseases in chest X-ray (CXR) images. It focuses on the U-Net architecture and its variations, including attention mechanisms and transformer blocks, which have revolutionized lung disease segmentation. The approaches are categorized into basic Deep Learning models, transfer learning, fine- tuning, and custom models, highlighting the progression and impact of these advanced techniques on improving diagnostic outcomes.

2.1 Segmentation:

2.1.1U-Net for CXR Segmentation

The U-Net architecture, with its encoder-decoder structure and skip connections, has occurred as a leading method for CXR segmentation. This setup, which captures high-level semantic information and low-level details, is crucial for accurately outlining lung boundaries. Studies have consistently shown U-Net's effectiveness in segmenting lung regions with high accuracy, a factor that significantly comforts the potential of this technology in improvingdiagnosticoutcomes fateh2023persian; fa teh2022providing.U-Net,introducedby Ronneberger et al., has become a fundamental tool in medical image segmentation ronneberger2015u. Islam et al. islam2018towards showcased U-Net's ability in accurately tracing lung boundaries, which has enhanced diagnostic precision. Additionally, Liu et al. liu2022automatic employed a pre-trained EfficientNet-B4 and developed an enhanced version of U-Net for identifying and segmenting lung regions.

2.1.2U-Net Enhancements with Transformers: Recent research has significantly advanced lung segmentation by enhancing the U-Net architecture with attention mechanisms. Oktay et al. oktay2018attention introduced mechanisms that enable the model to concentrate on the most crucial areas within chest X-rays using Attention Gates (AGs). This innovation enhances segmentation accuracy and sensitivity to disease characteristics. Additionally, research by Wu et al. wu2023transformer, Gu et al. gu2018recent, and Liu et al. liu2024sta has shown that incorporating attention mechanisms into the U-Net framework significantly improves lung segmentation performance, underscoring the effectiveness of this method in precisely outlining lung boundaries and enhancing diagnostic outcomes.

Khaniki et al. khaniki2024novel enhanced U-Net by incorporating a Convolutional Block Attention Module (CBAM), which integrates channel, spatial, and pixel attention to boost segmentation accuracy. Azad et al. and Chen et al. extended the U-Net framework with transformers, demonstrating significant improvements in capturing intricate details and achieving top-tier results in lung segmentation tasks azad2019bi; chen2021transunet.

3. EXISTING MODEL OBJECTIVE

Existing Model: CNN-Based Pneumonia Detection Architecture:

1. Input: Chest X-ray images resized to a fixed dimension.

2. Feature Extraction: Convolutional Neural Networks (CNNs) extract local features through convolution and pooling layers.

3. Flattening and Dense Layers: Extracted features are passed through fully connected layers for classification.

4. Output: Binary classification – Pneumonia or Normal.

Advantages:

1. High accuracy for localized patterns.

2. Efficient feature extraction for smaller datasets.

3. Well-established models such as VGG, ResNet, and Inception have been successfully used.

Limitations:

1. Focuses primarily on local features, missing long-range dependencies.

2. Performance degrades with large-scale and high-resolution datasets.

3. Limited capacity to capture global context, resulting in possible misclassification.

4.PROPOSED METHODOLOGY:

Vision Transformer (ViT)-Based Pneumonia Detection

Architecture:

1. Input: Chest X-ray images are divided into fixed- size patches (e.g., 16x16 pixels).

2. Patch Embedding: Each patch is flattened and passed through a linear projection to create embedding's.

3. Positional Encoding: Positional information is added to the patch embedding's to retain spatial relationships.

4. Transformer Encoder: Multiple self-attention and feed-forward layers analyse patch embedding's to capture both local and global dependencies.

5. Classification Token: A special classification token is appended, and its final state is used for classification.

6. Output: Binary classification – Pneumonia or Normal.

Advantages:

1. Superior performance by capturing global context using self-attention mechanisms.

2. Effective at handling high-resolution medical images.

3. Reduces reliance on large labeled datasets by utilizing pre-trained models.

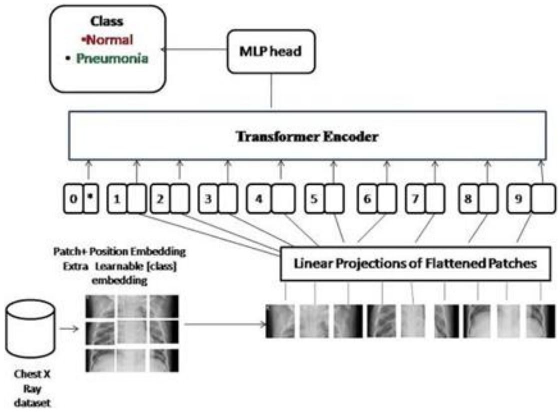

Fig 1. The proposed system design architecture.

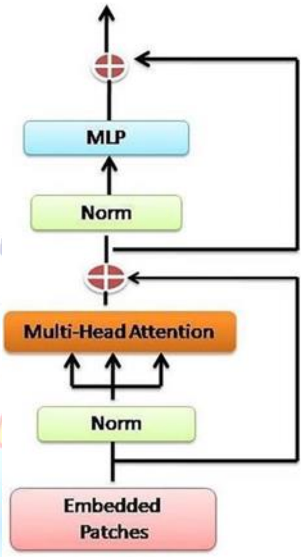

Fig 2. Internal design of a transformer encoder

5. EXPERIMENTAL RESULT

Fig3:pneumonia detection web page interface

Fig4:uploading normal chest x ray

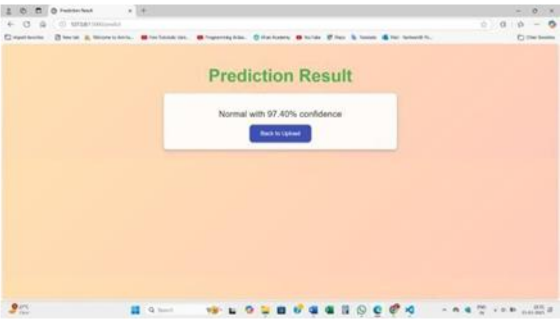

Fig5:Results normal chest x ray

Fig6:uploading pneumonia chest x ray photo

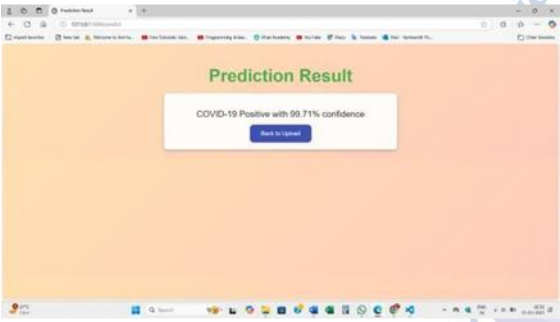

Fig7:Results of pneumonia chest x-ray

6. CONCLUSION :

The article conducts a thorough analysis of a Vision Transformer (ViT) framework for pneumonia detection in chest X-rays. ViTs' ability to analyze complex image relationships is showcased, demonstrating superior performance over traditional deep learnigs and other advanced techniques. ViTs excel in capturing global context, spatial relations, and handling variable image resolutions, leading to accurate pneumonia detection. The study aims to assess this method's effectiveness by comparing it to state-of-the-art models on a diverse CXR dataset. The results reveal ViT's superiority with an accuracy of 97.61%, sensitivity of 95%, and specificity of 98%. In conclusion, the ViT-based approach holds promise for early pneumonia detection in CXRs, offering substantial development potential in this field. However, limitations include data scarcity and the need for real-world validation. Future directions encompass enhancing interpretability, addressing model robustness, and conducting clinical trials for practical deployment.

Conflict of interest statement

Authors declare that they do not have any conflict of interest.

REFERENCES

[1] Pneumonia in children. WHO (2019). https://www.who.int/news-room/fact-sheets/detail/pneumonia

[2] Khan, S. H. et al. COVID-19 detection and analysis from lung CT images using novel channel boosted CNNs. Expert Syst. Appl. 229, 120477 (2022).

[3] Khan, S. H. et al. COVID-19 detection in chest X-ray images using deep boosted hybrid learning. Comput. Biol. Med. 137, 104816 (2021).

[4] Khan, S. H., Sohail, A., Zafar, M. M. & Khan, A. Coronavirus disease analysis using chest X-ray images and a novel deep convolutional neural network. Photodiagnosis Photodyn. Ter. 35, 102473 (2021).

[5] Singh, S., Tripathi, B. K. & Rawat, S. S. Deep quaternion convolutional neural networks for breast Cancer classifcation. Multimed. Tools Appl. 82, 31285–31308 (2023).

[6] Liang, G. & Zheng, L. A transfer learning method with deep residual network for pediatric pneumonia diagnosis. Comput. Methods Programs Biomed. 187, 104964 (2020).

[7] Nishio, M., Noguchi, S., Matsuo, H. & Murakami, T. Automatic classifcation between COVID-19 pneumonia, non-COVID-19 pneumonia, and the healthy on chest X-ray image: Combination of data augmentation methods. Sci. Rep. 10, 1–6 (2020).

[8] Asif, S., Zhao, M., Tang, F. & Zhu, Y. A deep learning-based framework for detecting COVID-19 patients using chest X-rays. Multimed. Syst. https://doi.org/10.1007/s00530-022-00917-7 (2022).

[9] Suryaa, V. S., Annie, A. X. & Aiswarya, M. S. Efcient DNN ensemble for pneumonia detection in chest X-ray images. Int. J. Adv. Comput. Sci. Appl. 12, 759–767 (2021).

[10] Singh, S., Kumar, M., Kumar, A., Verma, B. K. & Shitharth, S. Pneumonia detection with QCSA network on chest X-ray. Sci. Rep. 13, 9025 (2023).

[11] Duong, L. T., Nguyen, P. T., Iovino, L. & Flammini, M. Automatic detection of COVID-19 from chest X-ray and lung computed tomography images using deep neural networks and transfer learning. Appl. Sof Comput. 132, 109851 (2023).

[12] Duong, L. T., Le, N. H., Tran, T. B., Ngo, V. M. & Nguyen, P. T. Detection of tuberculosis from chest X-ray images: Boosting the performance with Vision Transformer and transfer learning. Expert Syst. Appl. 184, 115519 (2021).

[13] Duong, L. T., Nguyen, P. T., Iovino, L. & Flammini, M. Deep learning for automated recognition of COVID-19 from chest X-ray images. medRxiv. https://doi.org/10.1101/2020.08.13.20173997 (2020).

[14] Kazemzadeh, S. et al. Deep learning detection of active pulmonary tuberculosis at chest radiography matched the clinical performance of radiologists. Radiology 306, 124–137 (2023).

[15] Ramachandran, P. et al. Stand-alone self- attention in vision models. Adv. Neural Inform. Process. Syst. 32 (2019).

[16] Guo, M.-H., Lu, C.-Z., Liu, Z.-N., Cheng, M.-M. & Hu, S.-M. Visual attention. Network. 14, 1–12 (2022).

[17] Xu, K. et al. Show, attend and tell: Neural image caption generation with visual attention. in 32nd Int. Conf. Mach. Learn. ICML 2015 3, 2048–2057(2015).