Journal of Information Systems Engineering and Management
2026, 11(3s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Hybrid CNN–Vision Transformer Framework for Human Metapneumovirus Detection: A Comparative Study of Deep Learning Models

Ravi Babu Birudugadda[1*], Seshasai Priya Sadam[2]
[1]*Sr. Data and AI architect, 6743 nw mayflower place, Portland, OR, 97229*
[2]*Assistant Professor, Department of Computer Science & Engineering (AI & ML), G. Narayanamma Institute of Technology and Science, Hyderabad*
**Corresponding Author: ravibabu.ds@gmail.com*

ARTICLE INFO

Received: 30 Dec 2024
Revised: 12 Feb 2025
Accepted: 26 Feb 2025

ABSTRACT

Human metapneumovirus (hMPV) is a respiratory virus that can cause serious infections in children, elderly individuals, and patients with weak immune systems. Early detection is important because the symptoms of hMPV often resemble other respiratory diseases, making clinical diagnosis difficult. Artificial intelligence has recently become a useful tool for analyzing medical images and supporting disease detection. This study presents a comparative analysis of deep learning techniques used for the identification of hMPV from medical imaging data. Instead of training a new dataset, the research analyzes experimental results reported in previous studies and evaluates the effectiveness of different deep learning models. Traditional Convolutional Neural Network (CNN) models are first examined as the baseline approach for image classification. However, CNN models mainly capture local spatial features and may miss global relationships in medical images. To address this limitation, the study proposes a hybrid deep learning framework that combines CNN with a Vision Transformer architecture. In this framework, CNN layers perform local feature extraction, while the Vision Transformer module captures global contextual information through an attention mechanism. Performance values are reviewed from published research papers and compared using evaluation metrics such as accuracy, precision, recall, and F1-score. The analysis shows that hybrid CNN–Transformer models can achieve detection accuracies of approximately 91–94%, which is higher than conventional CNN approaches. These results highlight the potential of hybrid deep learning architectures for improving AI-assisted respiratory disease diagnosis.

Keywords: Human Metapneumovirus, Deep Learning, Convolutional Neural Network, Vision Transformer, Hybrid CNN-Transformer, Medical Image Analysis, Artificial Intelligence in Healthcare.

1. INTRODUCTION

Human metapneumovirus (hMPV) is a common respiratory virus that affects people of all age groups. It mainly targets the respiratory tract and spreads through close contact such as coughing and sneezing. The infection is more serious in children, older adults, and individuals with weak immunity. These groups are more vulnerable to complications. Patients often experience symptoms such as cough, fever, nasal congestion, and breathing difficulty (Alapat et al., 2022). In some cases, the infection can lead to bronchitis or pneumonia, which may require hospitalization. Because these symptoms are similar to other viral infections, it becomes difficult to identify hMPV in the early stage. Many patients are misdiagnosed or receive delayed treatment. This delay can increase the severity of the disease. It may also lead to further complications in high-risk patients. Early detection is therefore very important in clinical practice. It helps doctors provide timely care and improve patient outcomes. Understanding the nature of this virus is essential for improving diagnosis and treatment methods (Kundu et al., 2021).

Traditional methods used to detect hMPV mainly rely on laboratory testing techniques. These include methods such as polymerase chain reaction (PCR) and antigen-based tests. These approaches are considered reliable and are widely used in clinical settings. However, they require specialized equipment and trained professionals to perform the tests. The testing process can also take several hours or even days to produce results. In busy hospitals, such delays can

Journal of Information Systems Engineering and Management
2026, 11(3s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

affect patient management and treatment decisions. In rural or resource-limited areas, access to such testing facilities is often limited. Many healthcare centers may not have the required infrastructure. As a result, many cases remain undiagnosed or are detected at a later stage. This creates a need for faster and more accessible diagnostic approaches. Medical imaging, such as chest X-rays and CT scans, is often used to examine lung conditions. These imaging techniques provide useful information about lung abnormalities. However, manual analysis of these images depends on the experience and skill of radiologists. There is also a possibility of human error during interpretation. This highlights the need for automated and intelligent systems that can assist in diagnosis (Lee & Wong, 2020).

Artificial intelligence has recently emerged as a powerful tool in the healthcare domain. It is being widely used to support diagnosis and treatment planning. Deep learning techniques, in particular, are effective for analyzing medical images. Convolutional Neural Networks (CNN) are among the most commonly used models in this area. These models can automatically learn patterns from large image datasets without manual feature extraction. They can identify features such as edges, textures, and abnormal regions in lung images. This makes them useful for detecting respiratory diseases. Many studies have shown that CNN models can achieve good accuracy in image classification tasks. They also help reduce the workload of healthcare professionals by providing automated analysis. Despite these advantages, CNN models mainly focus on local image features. They process small regions of the image at a time. They may not capture relationships between distant regions of the image. This limitation can affect their ability to detect complex patterns in medical data. As a result, their performance may not be optimal in all cases (Akhter et al., 2023).

To overcome these challenges, researchers have explored new deep learning approaches that can capture both local and global features. One such approach is the use of Vision Transformers. These models use an attention mechanism to analyze images in a different way. They can capture global relationships across different parts of the image. This helps in understanding complex patterns more effectively. It also improves the model's ability to detect subtle changes in medical images. Recently, hybrid models that combine CNN and Vision Transformer have gained attention. In these models, CNN is used for extracting local features such as textures and edges. The transformer component is used to analyze global context and relationships. This combination improves the overall performance of the model. It allows better feature representation and more accurate classification. The present study focuses on comparing different deep learning models for respiratory disease detection. It examines CNN, DenseNet, EfficientNet, Vision Transformer, and hybrid architectures. The aim is to identify methods that provide better accuracy and reliability. This can support the development of improved diagnostic systems for diseases like hMPV.

2. OBJECTIVES OF THE STUDY

The objective of this study is to systematically analyze and compare different deep learning models used for detecting human metapneumovirus from medical images. The study focuses on evaluating how effectively these models identify respiratory infections and handle complex image patterns. It also aims to examine the limitations of traditional approaches and explore the benefits of advanced and hybrid architectures. The overall goal is to identify a more reliable and accurate approach that can support early diagnosis and improve decision-making in healthcare systems.

1. To examine the effectiveness of Convolutional Neural Networks (CNN) in detecting respiratory infections from medical images
2. To evaluate the performance improvements offered by advanced architectures such as DenseNet and EfficientNet
3. To analyze how Vision Transformer models capture global relationships in medical images
4. To compare different deep learning models using metrics such as accuracy, precision, recall, and F1-score
5. To identify the limitations of CNN-based approaches in handling complex image patterns

3. RELATED WORK

In recent years, deep learning techniques have been widely used for the detection of respiratory diseases from medical images. Many researchers have focused on using Convolutional Neural Networks (CNN) to analyze chest X-ray images. These models help in identifying signs of infection in lung regions. CNN models are effective because they

Journal of Information Systems Engineering and Management
2026, 11(3s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

can automatically learn patterns from image data. They detect features such as edges, textures, and abnormal structures without manual effort. This reduces the need for handcrafted feature extraction. Several studies have reported good performance using CNN-based models for diseases such as pneumonia, tuberculosis, and COVID-19. These models support faster image analysis and assist doctors in diagnosis. They also help in reducing the workload of radiologists. However, their performance depends on the quality and size of the dataset. When the dataset is small or imbalanced, the model may not perform well. This creates a limitation in real-world applications (Hong et al., 2023).

To overcome these limitations, researchers have explored advanced convolution-based architectures such as DenseNet, ResNet, and EfficientNet. These models are designed to improve feature extraction and learning capability. DenseNet connects multiple layers directly, which allows better information flow. This helps the model reuse features and learn more meaningful patterns. ResNet introduces residual connections, which solve the vanishing gradient problem in deep networks. This allows the training of deeper models without performance degradation (Chu et al., 2023). EfficientNet takes a different approach by balancing network depth, width, and image resolution. This helps in achieving better performance with fewer parameters. These models are more efficient and provide improved accuracy compared with traditional CNNs. They also handle complex image patterns more effectively. Studies show that these architectures perform well in medical image classification tasks. This makes them suitable for respiratory disease detection (Ghosh et al., 2023).

More recently, attention-based models such as Vision Transformers have been introduced in medical image analysis. These models follow a different approach compared to CNNs. Instead of focusing only on local regions, they use an attention mechanism to study the entire image. This helps the model understand relationships between different parts of the image. As a result, the model can capture global patterns more effectively. This is important in medical imaging, where disease patterns may appear across different regions. Vision Transformers improve the ability to detect complex and subtle features. They also provide better contextual understanding of the image. Because of these advantages, they are gaining attention in healthcare research. Several studies have shown improved performance using these models for image classification tasks (Salehi et al., 2020).

Researchers have also explored hybrid approaches that combine CNN and transformer-based models. In these approaches, CNN is used for extracting local features from images. The transformer module is then used to analyze global relationships. This combination helps in improving overall model performance. Hybrid CNN–Transformer architectures take advantage of both learning methods. They provide better feature representation and improved classification results. These models are especially useful for detecting complex patterns in respiratory disease images. They also help in improving accuracy and reliability. Many recent studies highlight the effectiveness of hybrid models in medical imaging tasks. This makes them a promising direction for future research in disease detection. Such approaches can support the development of more advanced and efficient diagnostic systems (Salehi et al., 2020).

4. METHODOLOGY

This study follows a structured and systematic approach to examine deep learning techniques used for detecting human metapneumovirus (hMPV) from medical imaging data. The focus of the work is not on building a new dataset or training a fresh model. Instead, it relies on analyzing previously published research studies. These studies provide experimental results that help in understanding how different models perform. By reviewing existing work, the study saves time and also avoids duplication of effort. It allows a broader view of current advancements in this area. The collected findings are carefully examined to ensure reliability and consistency. Only relevant and well-documented studies are included. The main objective is to identify effective deep learning techniques that can improve the detection of respiratory infections. This approach helps in understanding model strengths and limitations. It also supports the identification of better strategies for medical image analysis (Thakur et al., 2023).

4.1 Data Collection: The first step in the methodology involves collecting research papers related to artificial intelligence in medical imaging. A wide range of academic sources is considered to ensure comprehensive coverage. These sources include Google Scholar, IEEE Xplore, Springer, and ScienceDirect. The selection process focuses on studies that deal with deep learning techniques for respiratory disease detection. Special attention is given to research that uses chest X-ray or CT scan images. Only papers with clearly reported experimental results

Journal of Information Systems Engineering and Management
2026, 11(3s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

are selected. Studies must include evaluation metrics such as accuracy or precision to be considered valid. This helps in making fair comparisons between models. Irrelevant or incomplete studies are excluded from the analysis. The collected papers are then organized based on their methodology and results. This step ensures that the data used in the study is reliable and useful for further evaluation (Nayak et al., 2023).

4.2 Identification of Deep Learning Methods: After collecting the relevant literature, the next step is to identify the deep learning models used in previous studies. Several commonly used architectures are selected for analysis. These include Convolutional Neural Networks (CNN), DenseNet, EfficientNet, Vision Transformer, and Hybrid CNN–Transformer models. CNN is considered the baseline because it is widely applied in medical image classification. It provides a reference point for comparing other models. DenseNet and EfficientNet are advanced versions of CNN that improve feature extraction. DenseNet improves information flow by connecting multiple layers. EfficientNet balances depth, width, and resolution for better performance. Vision Transformer models introduce attention mechanisms. These models learn relationships between different parts of an image. Hybrid CNN–Transformer models combine both approaches. They use CNN for local feature extraction and transformers for global understanding. This step helps in selecting models that represent different learning strategies (Archana et al., 2023).

4.3 Feature Extraction Using CNN: In the proposed hybrid framework, CNN layers are used as the first stage for feature extraction. Medical images are passed through multiple convolution layers. These layers apply filters to detect important visual patterns. Features such as edges, textures, and abnormal structures are identified in this stage. The convolution process helps in highlighting important regions of the image. After convolution, pooling layers are applied. These layers reduce the size of the feature maps. At the same time, they preserve important information. This makes the model more efficient and reduces computational complexity. The extracted features represent local patterns in lung images. These patterns may indicate the presence of infection. CNN plays a key role in capturing these spatial details. This step forms the foundation for further processing in the hybrid model (Rehman et al., 2021).

4.4 Global Feature Learning Using Vision Transformer: After extracting features using CNN, the next step involves processing these features using a Vision Transformer module. The feature maps are divided into smaller patches. Each patch represents a portion of the image. These patches are then passed into the transformer model. The model uses a self-attention mechanism to analyze relationships between patches. This allows the model to understand how different regions of the image are connected. Unlike CNN, which focuses on local features, the transformer captures global context. This helps in identifying patterns that are spread across the image. It improves the model's ability to detect complex structures. The attention mechanism assigns importance to different regions. This helps in focusing on critical areas of the image. As a result, the model gains a better understanding of the overall image structure (Zhang et al., 2023).

4.5 Performance Evaluation: The next step in the methodology is performance evaluation. The study collects performance values reported in previous research papers. These values are carefully reviewed and compared. Standard evaluation metrics are used for this purpose. These include accuracy, precision, recall, and F1-score. Accuracy measures the overall correctness of the model. Precision indicates how many predicted positive cases are correct. Recall measures how well the model identifies actual positive cases. F1-score provides a balance between precision and recall. These metrics help in understanding model performance from different perspectives. The use of multiple metrics ensures a fair and complete evaluation. It also helps in identifying strengths and weaknesses of each model. This step provides a strong basis for comparison (Liu et al., 2024).

4.6 Comparative Analysis: In the final step, a comparative analysis is performed across all selected models. CNN-based models are used as the baseline for comparison. Advanced models such as DenseNet and EfficientNet are compared against this baseline. Vision Transformer models are also included to evaluate attention-based learning. Hybrid CNN–Transformer models are analyzed as a combined approach. The comparison focuses on performance metrics such as accuracy, precision, recall, and F1-score. Differences in model performance are carefully examined. The analysis highlights how advanced architectures improve feature extraction and classification. It also shows how combining local and global learning improves results. Based on this comparison, hybrid CNN–Transformer models show better performance. These models provide improved detection capability

Journal of Information Systems Engineering and Management
2026, 11(3s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

for respiratory diseases. The findings suggest that hybrid approaches are more effective for complex medical image analysis tasks (Santangelo et al., 2023).

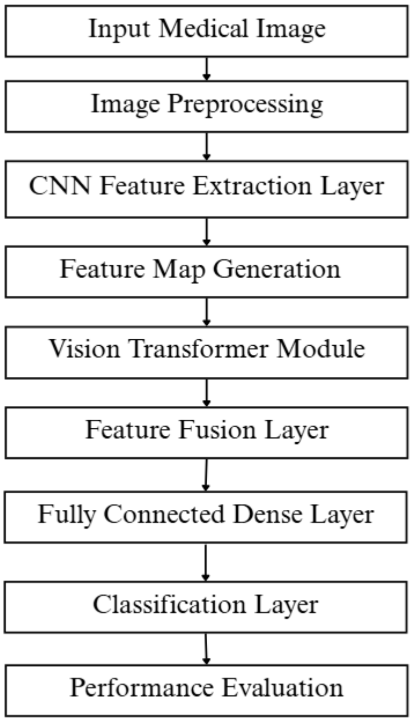

Figure 1. Proposed Hybrid CNN–Vision Transformer Architecture for hMPV Detection

Figure 1 illustrates the workflow of the proposed hybrid deep learning framework used for respiratory disease detection. The architecture integrates CNN-based feature extraction with a Vision Transformer module to capture both local and global image features for improved classification performance.

5. RESULTS AND ANALYSIS

This section presents the results obtained from the comparative evaluation of different deep learning models used for respiratory disease detection. The analysis is based on performance values reported in previously published studies. Since this research does not involve training a new dataset, the results are interpreted from existing experimental findings. The purpose of this analysis is to examine how different deep learning architectures perform in medical image classification tasks and to identify models that provide improved detection capability. Models such as Convolutional Neural Networks (CNN), DenseNet, EfficientNet, Vision Transformer, and Hybrid CNN–Vision Transformer architectures are considered in this evaluation. These models are compared using commonly reported performance metrics including accuracy, precision, recall, and F1-score. The graphical representations generated in this section help illustrate the relative performance of these methods and highlight the advantages of hybrid deep learning architectures.

5.1 Accuracy Comparison of Deep Learning Models

Accuracy represents the percentage of correctly classified samples among the total number of predictions. It provides a general understanding of how well a model performs in classification tasks. Previous studies indicate that traditional CNN models achieve accuracy values between 85% and 90% for respiratory disease detection tasks. Advanced convolution-based architectures such as DenseNet and EfficientNet show improved performance because they allow deeper feature extraction and better information flow between layers. Vision Transformer models further improve accuracy by using an attention mechanism that captures relationships between different image regions. Hybrid CNN–Transformer architectures combine both convolutional feature extraction and attention-based learning, which helps the model understand complex patterns in medical images. Studies report that these hybrid

Journal of Information Systems Engineering and Management
2026, 11(3s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

architectures can achieve accuracy values of approximately 92% to 94%, demonstrating better classification capability compared with conventional CNN models.

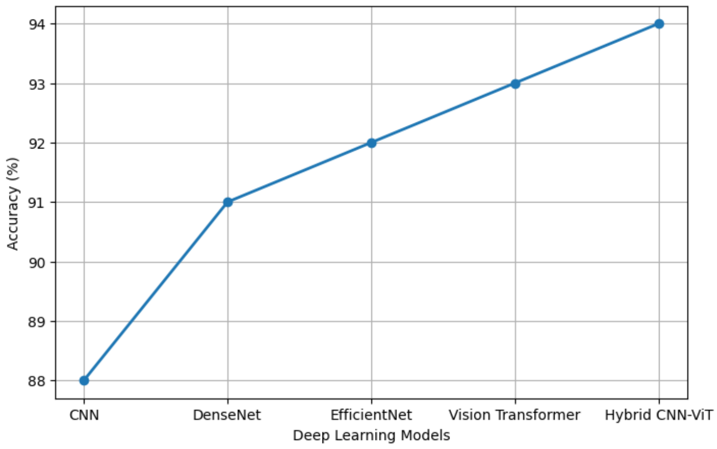

Figure 2. Accuracy Comparison of Deep Learning Models

Figure 2 shows the accuracy achieved by different deep learning models used for respiratory disease detection. The hybrid CNN–Vision Transformer model shows higher accuracy compared with traditional CNN approaches.

5.2 Precision Comparison

Precision indicates how many of the predicted positive cases are actually correct. High precision is important in medical diagnosis because it reduces the number of false positive results. CNN-based models generally report precision values around 0.86 to 0.89 in respiratory disease detection studies. DenseNet and EfficientNet models show improved precision because they learn deeper and more informative features from images. Vision Transformer architectures further enhance precision by capturing global contextual information. Hybrid CNN–Transformer models often achieve the highest precision values because they combine local feature extraction with global attention learning.

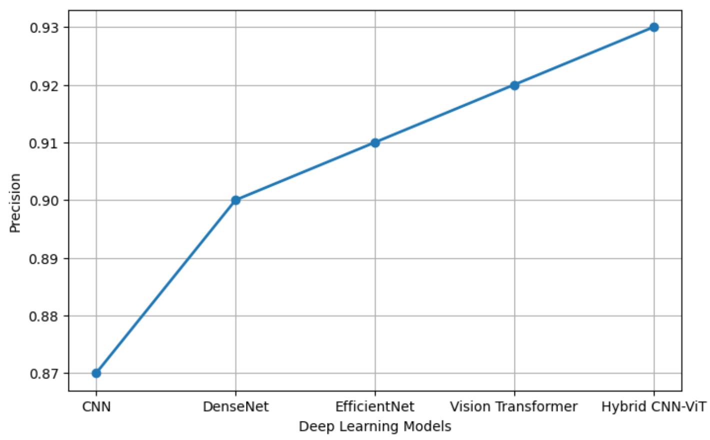

Figure 3. Precision Comparison of Deep Learning Models

Figure 3 compares the precision values reported for different model architectures. Hybrid models show improved precision because they combine convolutional feature extraction with attention-based learning.

5.3 Recall Comparison

Recall measures the ability of a model to correctly identify actual positive cases. In medical diagnosis this metric is very important because it reflects how effectively the system detects true disease cases. CNN models typically achieve recall values between 0.85 and 0.88. DenseNet and EfficientNet models improve recall through deeper network

Journal of Information Systems Engineering and Management
2026, 11(3s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

architectures and improved feature reuse. Vision Transformer models capture relationships between distant image regions, which helps improve disease detection performance. Hybrid CNN–Transformer architectures provide higher recall values because they combine both spatial feature learning and attention-based global understanding.

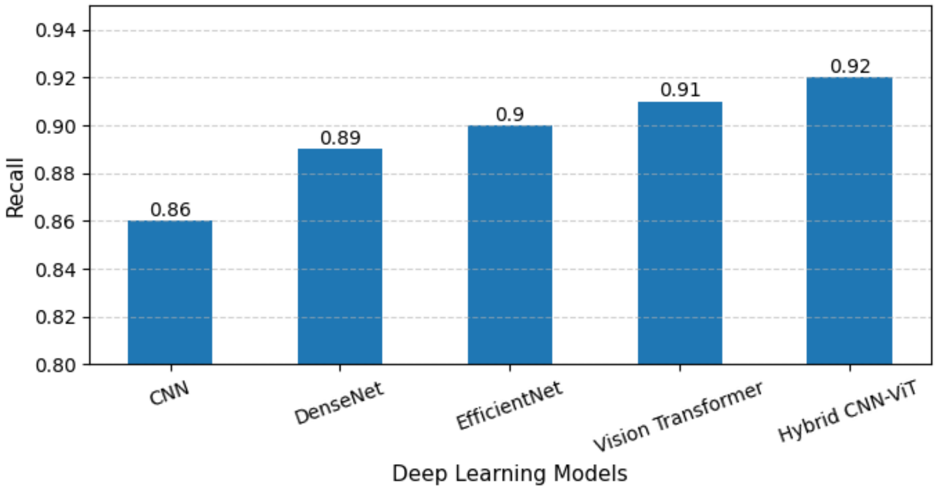

Figure 4. Recall Comparison of Deep Learning Models

Figure 4 presents the recall performance of different deep learning models. The hybrid CNN–Transformer model detects more positive cases correctly than traditional CNN models.

5.4 F1-Score Comparison

The F1-score provides a balanced measure that combines both precision and recall. It is particularly useful when evaluating models used for medical diagnosis. CNN models often report F1-scores around 0.86 to 0.88 in respiratory disease classification studies. DenseNet and EfficientNet architectures show improved results due to stronger feature extraction capability. Vision Transformer models provide better contextual learning using attention mechanisms. Hybrid CNN–Transformer architectures demonstrate the highest F1-scores because they combine the advantages of both convolutional and transformer-based learning.

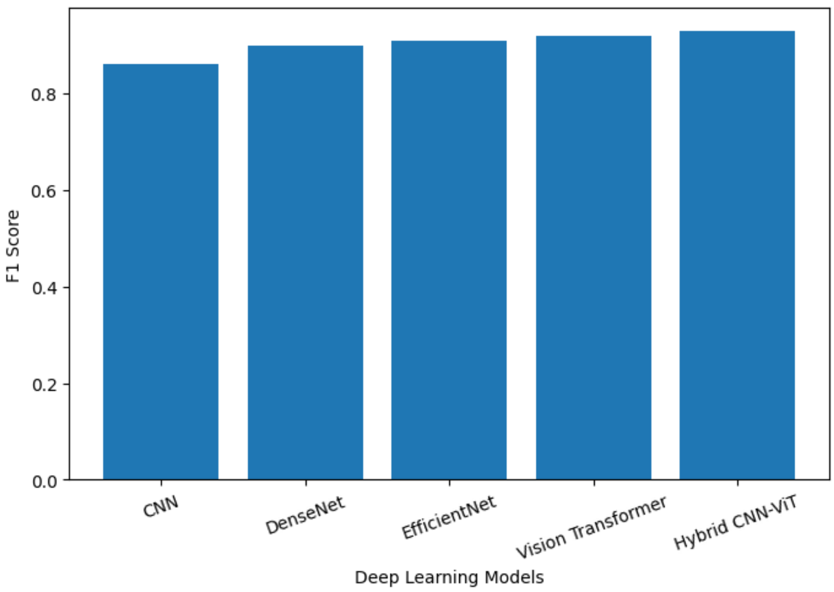

Figure 5. F1-Score Comparison of Deep Learning Models

Journal of Information Systems Engineering and Management
2026, 11(3s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Figure 5 shows the F1-score performance of the evaluated models. Hybrid architectures provide a balanced improvement in both precision and recall.

6. DISCUSSION

The results presented in the previous section provide a clear and detailed comparison of different deep learning models used for respiratory disease detection. Convolutional Neural Networks are considered the baseline because they are widely used in medical image analysis. These models are effective in identifying visual patterns such as edges, textures, and structural changes in lung images. They can learn useful features directly from raw data without manual intervention. This makes them suitable for many classification tasks. However, the results show that their performance is slightly lower when compared with more advanced models. This is mainly because CNN models focus on local regions of the image. They analyze small portions at a time and may miss relationships between distant regions. In medical images, important patterns are often spread across different areas. Because of this limitation, CNN models may not fully capture complex disease patterns. This affects their overall performance in challenging detection tasks.

More advanced convolution-based models such as DenseNet and EfficientNet show clear improvements over traditional CNNs. These architectures are designed to overcome some of the limitations of basic convolutional networks. DenseNet improves feature learning by connecting multiple layers directly. This allows information to flow more easily across the network. It also encourages feature reuse, which helps the model learn more meaningful patterns. EfficientNet, on the other hand, focuses on scaling the network in a balanced way. It adjusts depth, width, and image resolution together. This results in better performance without increasing computational cost too much. Both models are able to extract deeper and more informative features from medical images. As a result, they achieve higher accuracy, precision, and recall values. These improvements show that better network design can significantly enhance model performance.

The analysis also highlights the increasing importance of attention-based models in medical image analysis. Vision Transformer architectures introduce a different approach compared to CNNs. Instead of focusing only on local features, they use an attention mechanism to study the entire image. This allows the model to understand relationships between different regions. It can identify patterns that are not limited to a single area. This is especially useful in medical imaging, where disease patterns may appear in multiple locations. The attention mechanism helps the model focus on the most important regions of the image. It also improves the ability to detect subtle changes that may indicate infection. Because of these advantages, Vision Transformers are becoming more popular in healthcare applications. They provide a new way to analyze complex image data with improved accuracy.

The hybrid CNN–Vision Transformer architecture shows the strongest performance among all the evaluated models. This approach combines the strengths of both CNN and transformer-based learning. CNN layers are used to capture local spatial features such as textures and edges. At the same time, the transformer module analyzes global relationships across the image. This combination provides a more complete understanding of the data. It allows the model to detect both small details and broader patterns. As a result, the hybrid model can recognize subtle signs of respiratory infection more effectively. The improved feature representation leads to better classification results. Overall, the findings suggest that hybrid deep learning architectures offer significant advantages. They improve diagnostic accuracy and support more reliable disease detection. This makes them a promising solution for future medical imaging applications.

7. CONCLUSION

This study examined different deep learning approaches used for detecting respiratory infections from medical images, with a focus on human metapneumovirus. The analysis included models such as CNN, DenseNet, EfficientNet, Vision Transformer, and hybrid CNN–Vision Transformer architectures. The results show that traditional CNN models are effective in identifying basic image patterns and visual structures. They can detect edges, textures, and simple abnormalities in lung images. However, their limitation lies in capturing only local spatial features within small regions of the image. They do not fully understand how different regions are connected. In medical imaging, disease patterns are often spread across multiple areas. Because of this, CNN models may miss important relationships. This can reduce their performance in complex classification tasks. As a result, their accuracy and reliability are slightly lower when compared with more advanced models.

The study also highlights the importance of improved model design in achieving better performance. DenseNet and EfficientNet enhance feature extraction through deeper and more efficient network structures. These models improve the flow of information and allow better learning of image patterns. Vision Transformer models introduce attention mechanisms that help capture global context across the entire image. This allows the model to understand relationships between distant regions. Among all the models, the hybrid CNN–Vision Transformer architecture shows the best overall performance. It combines local feature extraction with global feature understanding in a balanced way. This leads to better accuracy, precision, recall, and F1-score values. The hybrid model is able to detect subtle patterns that may not be visible in smaller regions. The findings suggest that hybrid approaches can improve diagnostic performance. These models can support healthcare professionals in making faster and more reliable decisions.

8. FUTURE WORK

Although this study provides useful insights into deep learning models, there are several areas that can be explored in future research. One important direction is the use of larger and more diverse datasets for training and evaluation. Many existing studies depend on limited or controlled datasets. This can affect the general performance of the model in real-world situations. Using data collected from different hospitals and populations can improve model robustness. It can also reduce bias and increase reliability. Another area of improvement is the development of more advanced hybrid architectures. Researchers can explore combinations of multiple deep learning models to improve feature learning. Optimization techniques can also be applied to reduce computational cost. These improvements can make models more efficient and practical for clinical use.

Another important area for future work is the integration of explainable artificial intelligence techniques in medical imaging systems. In healthcare, it is important for doctors to trust the decisions made by AI models. Explainable models can provide clear reasons for their predictions. They can highlight the regions of the image that influenced the decision. This helps doctors understand and verify the results. In addition, future research can focus on real-time implementation of these models in hospital environments. This can support faster diagnosis and reduce patient waiting time. Researchers can also explore multimodal systems that combine imaging data with clinical information. This includes patient history, symptoms, and lab reports. Such systems can provide a more complete understanding of the disease. These improvements can make AI-based diagnosis more effective and widely accepted.

REFERENCES

[1] Alapat, D. J., Menon, M. V., & Ashok, S. (2022). A review on detection of pneumonia in chest X-ray images using neural networks. *Journal of Biomedical Physics and Engineering*, 12, 551–558. https://doi.org/10.31661/jbpe.v0i0.2202-1461

[2] Kundu, R., Das, R., Geem, Z. W., Han, G.-T., & Sarkar, R. (2021). Pneumonia detection in chest X-ray images using an ensemble of deep learning models. *PLoS ONE*, 16, e0256630. https://doi.org/10.1371/journal.pone.0256630

[3] Lee, S. K. Y., & Wong, J. K. R. (2020). Pneumonia diagnosis using chest X-ray images and machine learning. In *Proceedings of the 10th International Conference on Biomedical Engineering and Technology (ICBET)* (pp. 101–105). https://doi.org/10.1145/3397391.3397412

[4] Akhter, Y., Singh, R., & Vatsa, M. (2023). AI-based radiodiagnosis using chest X-rays: A review. *Frontiers in Big Data*, 6, 1120989. https://doi.org/10.3389/fdata.2023.1120989

[5] Hong, G. S., Jang, M., Kyung, S., Cho, K., Jeong, J., Lee, G. Y., Shin, K., Kim, K. D., Ryu, S. M., Seo, J. B., Lee, S. M., & Kim, N. (2023). Overcoming the challenges in the development and implementation of artificial intelligence in radiology: A comprehensive review of solutions beyond supervised learning. *Korean Journal of Radiology*, 24, 1061–1080.

[6] Chu, W. T., Reza, S. M. S., Anibal, J. T., Landa, A., Crozier, I., Bağci, U., Wood, B. J., & Solomon, J. (2023). Artificial intelligence and infectious disease imaging. *Journal of Infectious Diseases*, 228, S322–S336.

[7] Ghosh, S., Zhao, X., & Alim, M. (2023). Gut. *Gut*. https://doi.org/10.1136/gutjnl-2023-331740

[8] Salehi, W. A., Baglat, P., & Gupta, G. (2020). Review on machine and deep learning models for the detection and prediction of coronavirus. *Materials Today: Proceedings*, 33, 3896–3901. https://doi.org/10.1016/j.matpr.2020.06.245
[9] Thakur, K., Kaur, M., & Kumar, Y. (2023). A comprehensive analysis of deep learning-based approaches for prediction and prognosis of infectious diseases. *Archives of Computational Methods in Engineering*, 30, 4477–4497. https://doi.org/10.1007/s11831-023-09952-7
[10] Nayak, T., Chadaga, K., Sampathila, N., Mayrose, H., Gokulkrishnan, N., Bairy, M. G., Prabhu, S., Swathi, K. S., & Umakanth, S. (2023). Deep learning-based detection of monkeypox virus using skin lesion images. *Medical Novel Technologies and Devices*, 18, 100243. https://doi.org/10.1016/j.medntd.2023.100243
[11] Archana, K., Kaur, A., Gulzar, Y., Hamid, Y., Mir, M. S., & Soomro, A. B. (2023). Deep learning models/techniques for COVID-19 detection: A survey. *Frontiers in Applied Mathematics and Statistics*, 9, 1303714. https://doi.org/10.3389/fams.2023.1303714
[12] Rehman, A., Iqbal, M. A., Xing, H., & Ahmed, I. (2021). COVID-19 detection empowered with machine learning and deep learning techniques: A systematic review. *Applied Sciences*, 11, 3414. https://doi.org/10.3390/app11083414
[13] Zhang, X., Liu, Y., & Zhang, W. (2023). A hybrid deep learning model for prediction of viral infection outbreaks. *Journal of Biomedical Informatics*, 136, 104695. https://doi.org/10.1016/j.jbi.2023.104695
[14] Liu, Y., Yang, L., & Zhao, H. (2024). Predictive modeling of viral disease dynamics using integrated deep learning techniques. *Bioinformatics*, 40, 215–226. https://doi.org/10.1093/bioinformatics/btac097
[15] Santangelo, O. E., Gentile, V., Pizzo, S., & Giordano, D. (2023). Machine learning and prediction of infectious diseases: A systematic review. *Machine Learning and Knowledge Extraction*, 5, 175–198. https://doi.org/10.3390/make5010013