International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 11 | Nov - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Pneumonia Detection using Machine Learning

Anand Patwa, Raheel Hassan
Department of Computer Science and Applications, Sharda school of Computing Science and Engineering Greater Noida, UP.

Abstract

Over the last few years, strides made in artificial intelligence have greatly improved the identification of chest X-ray pneumonia. This paper examines and contrasts multiple approaches where contemporary neural networks and data-centric methods have been employed to enhance diagnostic precision. These approaches feature a combination of EfficientNetB7 and Squeeze-and-Excitation modules, a PD-Net variant based on ResNet50, and a YOLOv8 system reinforced by the generation of synthetic images using FILM and PCA methods. Every strategy has its own merits: EfficientNetB7's feature extraction skill is unparalleled, the balance YOLOv8 maintains between the quality of detection and the speed of processing is commendable, and ResNet50 is a solid transfer learning backbone tailored for hospital settings. This review additionally stresses the requirement for balanced datasets, the application of explainable artificial intelligence (AI) techniques like Grad-CAM for model debugging, augmented data, and explainable AI frameworks to provide more transparency to the end-users, the clinicians. The integration of diverse and interpretable data with computational efficiency is highlighted by this study as the main factor for reliable AI systems for the detection of pneumonia in vivo clinical settings.

Keywords— Prenumonia detection/ Deep learning/ Comparative review/ CNN/ Explainable AI/ XAI/ Radiology/ Chest X-ray/ CXR/ Medical imaging

INTRODUCTION

Pneumonia still causes a significant amount of global illness. Pneumonia causes a significant amount of global illness and death and remains a large burden on healthcare systems worldwide, regardless of development. [1] Healthcare systems in both developing and developed countries feel the burden of poorly managed Pneumonia in the population. Diagnosis of Pneumonia has classically relied on chest imaging, with chest x rays and chest CT being the most commonly used. While the imaging techniques as mentioned will always continue to be the gold standard in diagnosis, the diagnosis remains the dependant on the subjective interpretation of the radiologist. This subjectivity, especially with the more subtle cases and the atypical cases, can cause delays and mistakes to the point of the presentation being life threatening. Pneumonia's poorly managed cases highlight the greater necessity of accurate and efficient diagnostic methods that have a degree of objectivity to aid the clinician and enhance the outcome for the patient, both on a cost and resource usage basis. [1][3]

The adoption of Artificial Intelligence (AI), especially Deep Learning (DL), has been extraordinarily rapid in medicine, propelled mainly by high-performance computing and large-scale medical datasets [1]. Among various techniques, Convolutional Neural Networks (CNNs) have been very effective in the automatic identification of lung diseases [1], [4], [5]. Models of this kind are currently even capable of matching or surpassing human diagnostic accuracy, thus, they are able to make diagnoses quicker, more stable, and of a high degree of trust [1]. The previous works in this field have been based on conventional CNN architectures [4]–[6]. However, continuous breakthroughs have led to more advanced and specialized deep learning architectures.The paper states that primary to these phenomenal results is the ensembling of models like PneuX-Net which achieve classification accuracies of up to 99.91% [7]. Similarly, object detection frameworks like YOLOv8 that are refined by artificial image generation to compensate for dataset imbalance have attained very good results—around 97% accuracy in differentiating pneumonia and COVID-19 cases [8]. Recently, researchers have also looked at alternative concepts such as Kolmogorov–Arnold Networks (KANs) which, apart from high accuracy (98.84%), also offer better interpretability of the model than a standard CNN classifier [9].

I. The significant advancements in supercomputing power over the past few years combined with the availability of very large medical datasets has considerably facilitated the adoption of Artificial Intelligence, particularly Deep Learning, in the field of medical image interpretation [1]. Among these techniques, Convolutional Neural Networks have demonstrated brilliant ability in the automated detection of lung abnormalities [1], [4], [5]. Presently, such systems have the potential of equaling and even outperforming the diagnostic accuracy of human experts, as they provide faster, more standardized, and highly reliable image evaluations [1]. The transition of this field of study was based on the use of conventional CNN architectures [4]–[6]. Nevertheless, continuous research and development have been leading to more and more sophisticated and finely tuned deep learning architectures. According to recent publications, such superlative results are reported as a consequence of the use of ensembling PneuX-Net models reaching up to 99.91% classification accuracy [7]. In the same vein, the rather impressive performance of around 97% accuracy in the identification of pneumonia and COVID-19 samples is being attributed to the use of the object detection framework like YOLOv8 that was improved through generating synthetic images to tackle the problem of dataset imbalance [8]. Most recently, the possibility of using novel structures such as Kolmogorov–Arnold Networks (KANs) has been considered showing that in addition to very high accuracy (98.84%) KANs also provide an easily understandable model when compared to the conventional CNN classifiers [9].

Besides binary classification, some machine learning models have been employed in prediction of the disease progression and outcomes, such as the prediction of mortality risk and ICU admission which can be regarded as the first step in the creation of clinical decision support systems [10]. Despite this, the clinical utilization of these types of technology is still at a very low level. One of the main reasons for this is the "black box" issue of deep learning that may cause the clinicians' trust in the system to diminish [1]. The invention of Explainable AI (XAI) Gradient-weighted Class Activation Mapping (Grad-CAM), which tries to help the user understand the deep learning predictions in a visual manner, can be considered a step towards not only lessening the problem but also toward AI decision-making transparency [8], [11]. Furthermore, the thorough surveys in this domain have recognized that besides these problems, the major issues of the reviewed papers are very high bias, lack of dataset diversity, and limited external validation [10]. When promotion of high accuracy performance is made, a certain degree of skepticism may be present based on the assumption that the issues are there, and hence, it may be difficult to recognize the actual impact on practice.

This paper examines the developments of deep learning techniques in detecting pneumonia from medical images. It traces the changes of the main neural network architectures—starting from simple convolutional neural networks, through ensembles, and up to the latest neural networks—and the role of explainable artificial intelligence in gaining the trust of the clinical environment [1], [8], [11]. The article also embraces the continuing struggle against problems arising from data quality, model evaluation, and clinical implementation [1], [10]. Understanding of these two communities of interest is what the paper strives for by indicating the development of AI technology that is not only accurate but also explainable, trustworthy, and clinically usable.

LITERATURE REVIEW

Over the years, research in pneumonia detection through Deep Learning (DL) continues to be dominated by Convolutional Neural Networks (CNNs). This is largely because CNNs excel in recognizing and learning the tiered visual structures embedded in medical images. The research focus has shifted from using standard CNNs, to advanced ensemble systems, and then to more sophisticated neural network architectures tailored for enhanced performance and dependability. The most used approach in the literature is transfer learning, putting to work models that have been trained on large-scale image datasets, like ImageNet. CNNs, including VGG16, DenseNet121, MobileNet, and the ResNet family, are commonly used as feature extractors and baseline models for more complex architectures. Several studies have, for instance, customized DenseNet121 or used modified ResNet50V2 for benchmarking. Although there has been work on system optimization to enhance overall architecture performance, these systems have traditionally served as the baseline on which more specialized architecture designs are developed.

To go beyond the performance of baseline models, the researchers have progressively opted for hybrid and ensemble-based approaches. Ensemble learning combines the power of several architectures, thus enhancing stability and predictive performance. To this effect, combinations of diverse CNN structures may be used as an example, VGG16 and ResNet50V2 [6] being integrated with a customized CNN, or feature-level ensembling may be applied for more in-depth representation learning.

One of the significant works in this area is PneuX-Net [7] that not only extracts features with an Xception backbone but also blends them with the outputs of several machine learning classifiers, among them Random Forest (RF), Gaussian Naïve Bayes (GNB), and K-Nearest Centroid (KNC). A classification accuracy as high as 99.91% is what this model got. In the same way, object detection-based architectures like YOLOv8 have been talked about; Hasib et al. [8] proved its strong performance, thus achieving 97% accuracy for the differentiation between pneumonia and COVID-19 cases, which is a step forward as compared to the previous CNN models like InceptionV3.

Researchers call for hybrid and ensemble-based approaches to surpass the capabilities of baseline models. Ensemble learning merges the strengths of multiple architectures, thereby improving stability and predictive performance. These models could, for instance, combine different CNN architectures, such as a tailored CNN, VGG16, and ResNet50V2 [6], or use feature-level ensembling for deeper representation learning. A significant stepping stone in this field is PneuX-Net [7] that uses an Xception backbone to get features and combines the outputs with a bunch of machine learning classifiers like Random Forest (RF), Gaussian Naive Bayes (GNB), and K-Nearest Centroid (KNC). The model got an astonishing classification accuracy of 99.91%. Correspondingly, object detection-based architectures like YOLOv8 have drawn the interest of the researchers; Hasib et al. [8] exhibited its solid performance, achieving 97% accuracy in differentiating pneumonia from COVID-19 cases, thus leading to a performance better than that of the earlier CNN models like InceptionV3.

Innovations of the recent past are also willing to bet on new neural paradigms that are different from the current ones. One of the very few examples of such is JAM-net by Romano et al. [9], which uses Kolmogorov–Arnold Networks (KANs) in place of traditional Multi-Layer Perceptrons (MLPs). Besides the very high accuracy of 98.84% it attained, this new structure also shows the way to the model interpretability to become easier as compared to the use of traditional CNN-based models [9].

With overall classification accuracy going higher and higher, research has taken a wider turn to cover more clinically meaningful goals. Diseases severity prediction is one of the emerging goals beyond infection detection. Modelling systems have been challenged to predict patient mortality and the probability of Intensive Care Unit (ICU) admission thus coming a big step closer to real-world clinical decision support [10]. Simultaneously, the so-called "black box" problem is still there - a large number of deep learning models are not interpretable, which can cause the distrust of clinicians [1]. Therefore, Explainable AI (XAI) techniques like Gradient-weighted Class Activation Mapping (Grad-CAM) gain more and more prominence. Such visualization tools give a hand in understanding model logic by creating the areas of interest on the input, which is typical for YOLOv8 and other CNN-based frameworks [8], [11], thus making a step closer to human understanding from algorithmic performance.

Even with the rapid technological progress, the literature has been pointing out the same limitations and challenges that have not been solved yet. The most significant of these is the problem of dataset imbalance, as pneumonia datasets usually have a lot more normal samples than abnormal ones, which results in bias to training outcomes. In order to eliminate this bias completely, scientists are using synthetically created augmentation techniques to make underrepresented classes bigger in terms of the number of samples and to get the performance to be stable as well [8]. Nevertheless, there are still some major problems that remain. A recent systematic review [10] revealed that many of the high-performing studies have problems with limited diversity of the dataset, bias, and lack of external validation. For example, PneuX-Net [7] models that have been held up as accurate are usually tested on one dataset—most of the time a pediatric one—which makes it hard to tell if they can work with different populations and clinical contexts.

To sum up, DL models have been able to make the right diagnosis of pneumonia most of the time, but their use outside the laboratory is still restricted by data problems, interpretability problems, and lack of testing across different hospitals. Without overcoming these problems, it will not be possible to use the high-accuracy algorithms as clinically reliable AI systems that can be deployed in the real world.

Title / Study	Author(s) / Year	Techniques / Methods Used	Algorithms or Models	Primary Objective	Limitations / Observed Drawbacks
JAM-net: A KAN-based Deep Neural Network for Pneumonia Detection in Chest X-Rays	*Irene Amerini (2025)*	Transfer learning, advanced neural network design	DenseNet169, Kolmogorov–Arnold Networks (KAN), Convolutional KANs (CKAN)	To evaluate whether newer architectures such as KANs and CKANs can outperform traditional MLP models in pneumonia detection, while also enhancing interpretability [2908, 2932].	High computational complexity and memory requirements; significantly longer training times than MLP-based systems [3071, 3072].
YOLOv8 Framework for COVID-19 and Pneumonia Detection Using Synthetic Image Augmentation	*Lip Yee Por & Chin Soon Ku (2025)*	Synthetic image augmentation, transfer learning, Explainable AI (XAI)	YOLOv8, InceptionV3, DenseNet, ResNet, Grad-CAM	To create a robust model using synthetic data and XAI tools to improve pneumonia and COVID-19 classification accuracy and interpretability [3789].	Synthetic data may not reflect full real-world variability; occasional misclassification of normal X-rays as pneumonia persists [4312, 4317].
PneuX-Net: Enhanced Feature Extraction and Transformation Approach for Pneumonia Detection in X-ray Images	*Atiq Ur Rehman & Muhammad Usama Tanveer (2025)*	Ensemble-based feature extraction	Xception backbone with Random Forest (RF), Gaussian Naïve Bayes (GNB), and K-Nearest Centroid (KNC) classifiers	To propose an ensemble framework that strengthens feature extraction, reduces overfitting, and achieves higher diagnostic accuracy [1191, 1195].	Validation limited to a single pediatric dataset; generalizability across ages and clinical settings remains uncertain [1761].
Application of Artificial Intelligence in Radiological Image Analysis for Pulmonary Disease Diagnosis: A Review	*Karolina Zalewa (2025)*	Literature review and analysis of AI methods in radiology	Surveys CNN, ML, and DL-based architectures [545, 551]	To review AI applications in pulmonary disease diagnosis and assess current integration challenges in clinical workflows [545, 546].	Identifies field-wide barriers including limited dataset diversity, model opacity (“black box” issue), and workflow integration difficulties [553, 722, 725].
The Use of Machine Learning-based Models to Predict the Severity of Community-Acquired Pneumonia: A Systematic Review	*Brian W. Johnston (2025)*	Systematic literature review	Neural Networks, Naïve Bayes, Random Forest, Support Vector Machine (SVM)	To evaluate ML models used to predict CAP severity, focusing on ICU admission and mortality prediction [1909].	High variability in methodology and datasets, substantial bias, and lack of reproducibility across studies limit cross-comparison [1913].

In spite of the rapid technical progress, literature keeps pointing out several limitations and challenges that have not been solved yet. The most important one is the issue of dataset imbalance, as pneumonia datasets usually have a lot more normal samples than abnormal ones, which means that training outcomes are biased. To solve it, researchers have started using synthetic augmentation techniques to artificially create more samples in the underrepresented classes and thus make the models' performance more stable [8]. However, there are still bigger problems beyond these. A recent systematic review [10] has pointed out that many of the high-performing studies have problems with limited dataset diversity, bias, and lack of external validation. For example, PneuX-Net [7] type of models that show impressive accuracy and have been validated mostly on a single dataset—usually a pediatric one—thus it is difficult to support the claims of generalizability to different populations and clinical settings.

Summing up, DL models have set the bar high in terms of diagnostic accuracy of pneumonia and have been able to detect the disease with great precision, yet their practical implementation is hampered by limitations in data, interpretability, and lack of cross-institutional testing. Overcoming these obstacles is the only way of making the high-accuracy algorithms trusted AI systems that can be used in clinical practice.

6. Evaluation Metrics and Results

Model	Accuracy	Precision	Recall	F1-Score	AUC	Remarks
Behera (EfficientNetB7 + SE)	98.31%	≈0.99	≈0.99	≈0.99	≈1.00	Highest reported accuracy; pediatric cohort
Hasib (YOLOv8 + FILM/PCA)	97%	0.97	0.97	0.97	—	Balanced performance; improved recall post-augmentation
PD-Net (ResNet50)	95%	0.94	0.96	0.95	0.97	Strong recall; simple architecture

Interpretation:

In principle, all models are capable of diagnosing with high performance, but the content of the dataset and the strategies for augmenting the data have a significant impact on the comparison of the models. The figures of Behera are, to all numerical extents, the highest, however, they are from a children's dataset, so it is necessary to be cautious when making generalizations.

7. Strengths, Limitations, and Practical Insights

Study	Strengths	Limitations
Behera (EfficientNetB7 + SE)	Exceptional accuracy; strong feature recalibration through SE blocks; effective use of Grad-CAM.	Computationally intensive; potential overfitting risk; large model footprint.
Hasib (YOLOv8 + FILM/PCA)	Directly addresses class imbalance; integrates localization and XAI; reproducible setup.	Synthetic data may introduce artifacts; requires validation against real samples.
PD-Net (ResNet50)	Lightweight and clinically deployable; high recall for screening.	Slightly lower accuracy; lacks advanced augmentation or attention layers.
Review / Implementation Perspective	Suggests efficient attention-based transformer architectures for constrained devices.	Highlights need for larger, diverse datasets for generalization.

8. Practical Recommendations

Interpretation:

All models have been able to deliver excellent diagnostic performance. However, the composition of the datasets and augmentation strategies have a significant impact on how the results can be compared. Behera's outcomes are the highest in numbers, but since they are from a pediatric dataset, it is necessary to take caution when generalizing.

To achieve the highest classification accuracy: EfficientNetB7 + SE (Behera) should be replicated, external data should be used for validation, and Grad-CAM should be employed for interpretability.

In the case of imbalanced datasets or real-time detection: Hasib's method of using YOLOv8 + FILM/PCA can be chosen, however, synthetic images should be clinically validated.
If there are limited computational resources: PD-Net (ResNet50) with transfer learning can be utilized, which is a simpler method but still robust.
Per-class metrics, confusion matrices, and Grad-CAM heatmaps should always be reported, and ablation studies comparing real and synthetic data should be conducted.

9. Reproducibility Overview

Study	**Reproducibility Features Provided**
Hasib et al.	**Full FILM-PCA augmentation procedure, dataset counts, YOLOv8 hyperparameters.**
Behera et al.	**Model design (EfficientNetB7 + SE), dataset splits, Grad-CAM outputs, training settings.**
PD-Net	**Architecture outline (ResNet50 + dense head), key metrics, and basic training parameters.**

Conclusion and Summary

The variety of contemporary deep learning techniques used for the detection of pneumonia from chest X-ray images is demonstrated by this comparative study. Although the authors' architectural designs, data handling strategies, interpretability mechanisms, and practical feasibility for real-world healthcare applications vary greatly, all of the models evaluated in the paper demonstrate strong diagnostic capabilities.

With a reported accuracy of 98.31% and nearly flawless AUC values, Behera et al.'s EfficientNetB7 enhanced with Squeeze-and-Excitation (SE) blocks is the most performance-oriented configuration to convey the longest line of their work. The integration of Grad-CAM visualization in the study elevates clinical transparency, which is why radiologists' validation is crucial. However, without further training or trimming, the EfficientNetB7 is not the best option for general clinical settings due to its weight and exclusive dependence on a pediatric dataset.

However, the YOLOv8 system by Hasib et al. is a good illustration of how operational practicality and diagnostic performance can be balanced. The authors were able to overcome the issue of dataset imbalance and attain very comparable precision and recall values (~97%) for all classes by creating synthetic images using FILM and PCA. The authors' innovative approach to integrating explainable AI into diagnostic pipelines that involve humans is the use of LLM to aid interpretability. However, clinical staff must exercise caution to ensure that images are realistic and to prevent the introduction of bias due to the reliance on artificially generated data.

PD-Net is a practical approach that prioritizes simplicity, generalizability, and deployment feasibility. It is derived from ResNet50 with a custom lightweight head. Its high performance and low computational cost make it appropriate for edge devices and low-resource healthcare settings. Its accuracy is approximately 95%, and its recall is 0.96. However, the model's simple design allows for enhancement by incorporating sophisticated augmentation techniques or attention mechanisms.

Grad-CAM and associated explainability tools are consistently used in all studies, indicating a clear shift toward transparent AI in medical imaging. Establishing ethical AI deployment in diagnostic workflows and fostering clinician trust depend on this trend. However, more significant issues still exist: even the most accurate models' ability to be broadly applied is still constrained by a lack of standardized benchmarking, inconsistent external validation, and a small diversity of datasets.
Practically speaking, the results point to three complementary avenues for further study and application:
To verify their stability outside of pediatric cohorts, high-accuracy systems like EfficientNetB7+SE must undergo extensive testing on sizable, multi-institution datasets.
YOLOv8+FILM/PCA and other augmentation-driven frameworks should validate synthetic data with domain experts and explore hybrid augmentation technQiques that combine generated and real samples.
Because lightweight models like PD-Net are specifically designed for environments with constrained computational resources, they can be expanded to further clinical deployment.
Actually, the next generation of clinically viable AI models for pneumonia detection will probably be built around the

combination of these three approaches: interpretability, data diversity, and architectural efficiency. Future research should concentrate on transparency in reporting, reproducibility, and cross-dataset evaluation. Deep learning solutions should transform from successful experiments into dependable, trustworthy instruments for real-world healthcare diagnostics.

References

[1] K. Zalewa, J. Olszak, W. Kapłan, D. Orłowska, L. Bartoszek, M. Kaus, and N. Klepacz, "Application of artificial intelligence in radiological image analysis for pulmonary disease diagnosis: A review of current methods and challenges," *Journal of Education, Health and Sport*, vol. 77, p. 56893, 2025.

[2] M. K. Dalei and S. Mahapatra, "Pneumonia detection and classification from chest X-ray images using modified DenseNet121 deep learning approach," in *Proc. Int. Conf. on Ambient Intelligence in Health Care (ICAIHC)*, 2025, pp. 1–5.

[3] S. Sharmila, N. Navya, R. S. Bhavana, K. Naveena, and A. B. Priya, "Automated pneumonia detection from chest X-rays using enhanced deep learning techniques," in *Proc. Int. Conf. on Multi-Agent Systems for Collaborative Intelligence (ICMSCI)*, 2025, pp. 905–912.

[4] M. R. Kumar and B. G. Vani, "Implementation of deep learning mechanisms for detection and classification of pneumonia on chest X-ray dataset," in *Proc. Int. Conf. on Computational Intelligence for Green and Sustainable Technologies (ICCIGST)*, 2024, pp. 1–6.

[5] G. Saikrishna and C. Lakshmi, "Pneumonia detection using convolutional neural network," in *Proc. 6th Int. Conf. on Inventive Research in Computing Applications (ICIRCA)*, 2025, pp. 1426–1431.

[6] A. Yaqin, L. P. O. T., B. Satya, K. Kraugusteeliana, A. C. Frobenius, and P. Ferdiansyah, "Pneumonia disease analysis using deep learning on X-ray images," in *Proc. Int. Conf. on Informatics, Multimedia, Cyber and Information System (ICIMCIS)*, 2024, pp. 804–808.

[7] K. Munir, M. U. Tanveer, H. J. Alyamani, A. Bermak, and A. U. Rehman, "PneuX-Net: An enhanced feature extraction and transformation approach for pneumonia detection in X-ray images," *IEEE Access*, vol. 13, pp. 84024–84037, 2025.

[8] U. A. Hasib, R. M. Abu, J. Yang, U. A. Bhatti, C. S. Ku, and L. Y. Por, "YOLOv8 framework for COVID-19 and pneumonia detection using synthetic image augmentation," *Digital Health*, vol. 11, pp. 1–18, 2025.

[9] M. Romano, J. Tedeschi, and I. Amerini, "JAM-net: A KAN-based deep neural network for pneumonia detection in chest X-rays," in *Proc. IEEE Symp. on Computational Intelligence in Health and Medicine Companion (CIHM Companion)*, 2025, pp. 1–5.

[10] C. Lythgoe, D. O. Hamilton, B. W. Johnston, S. Ortega-Martorell, I. Olier, and I. Welters, "The use of machine learning-based models to predict the severity of community-acquired pneumonia in hospitalised patients: A systematic review," *Journal of the Intensive Care Society*, vol. 26, no. 2, pp. 237–248, 2025.

[11] S. K. Behera, K. M. Gopal, and S. B. Punuri, "A deep learning-based pneumonia detection system with explainable AI for medical decision support," in *Proc. 11th Int. Conf. on Communication and Signal Processing (ICCSP)*, 2025, pp. 694–699.