



AN ATTENTION-ENHANCED HYBRID DEEP LEARNING FRAMEWORK FOR EXPLAINABLE PNEUMONIA DETECTION FROM CHEST X-RAY IMAGES

Posina Anusha¹, T. Venkata Rathnam², Dr. Shaik Jaffar Hussain³, Dr. T. Harikala⁴, Dr. L. Jayasree⁵

Original Article

¹Assistant Professor, Department of AI&DS, Annamacharya Institute of Technology & Sciences, Tirupati, India.²Assistant Professor, Department of C.S.E, Bhaskar Institute of Technology and Science, Rayachoty, India.³Associate Professor, Department of C.S.E, Presidency University, Bengaluru, India.⁴Assistant Professor, Department of C.S.E, Annamacharya university, Rajampet, India.⁵Assistant Professor, Department of C.S.E, Sri Padmavati Mahila Visvavidyalayam, Tirupati, India.*Corresponding Author's Email: anusha.ksrm@gmail.com, isit.rathnam@gmail.com, jaffar.thebest@gmail.com, hariekalla@gmail.com, jayasreemohan15@gmail.com

Abstract

Chest X-ray (CXR) imaging is not particularly easy to interpret, due to the subtle appearance of the disease on CXR, and is prone to inter-observer variation, it is still regarded as the main diagnostic imaging modality for pneumonia, worldwide. Despite the promising results of deep learning models for automated pneumonia diagnosis, current convolutional architectures use only a single depth-based feature representation, and are mostly opaque, non-interpretable classifiers, which compromises the reliability and trust in clinical use. In this paper, we present PneumoFusion-Net, a multi-level convolutional feature extraction coupled with channel and spatial attention refinement framework to focus on the regions of the lungs relevant for diagnosis and ignore irrelevant background information. The extracted features from the shallow, intermediate and deep layers of a ResNet50 backbone are further refined using attention modules and merged into a unified representation for subsequent classification, while the Grad-CAM is directly incorporated in the pipeline to produce visual explanations for each prediction. The proposed model was trained and tested on the Kaggle chest X-ray pneumonia dataset which is publicly available consisting of 5,856 images. Experimental results show that the PneumoFusion-Net obtains test accuracy of 92.00%, precision of 0.92, recall of 0.92 and F1 of 0.92, while has a significantly smaller train-test accuracy gap of 2.60% compared to the VGG16, ResNet50, DenseNet121 and MobileNetV2 baselines, showing better generalization performance. The results indicate that multi-level attention fusion and built-in explainability are a practical and clinically reliable approach to automated screening of pneumonia.

Index Terms – Pneumonia Detection; Chest X-ray Classification; Attention Mechanism; Channel and Spatial Attention; Multi-level Feature Fusion; Explainable Artificial Intelligence; Grad-CAM; Deep Learning; Convolutional Neural Network; Medical Image Analysis

Introduction

Pneumonia is one of the major causes of morbidity and mortality in the world, especially in children under 5 years of age and elderly population, and CXR imaging is still the first and the most cost effective method for diagnosis of pneumonia [1]. Yet, manual interpretation of CXRs is costly and observer dependent, especially in areas where there is limited access to trained radiologists [2]. This diagnostic gap has spurred a large volume of research on automated deep learning-based computer-aided diagnosis (CAD) system that can achieve or surpass the level of accuracy of expert diagnosis [3].

There are certain drawbacks in the limited progress made so far. Some convolutional neural network (CNN) models only used the deep-layer representation and ignored fine-grained texture information that is of diagnostic value for diverse pneumonia appearances [4]. In addition, most high performing architectures are "black box" systems that limits their adoption in clinical settings if interpretability and trust are crucial. In recent years, ensemble learning methods have

been investigated to enhance robustness [6] and transformer-based methods to learn long-range contextual dependencies beyond the local receptive fields of CNNs [7] have also been explored. A hybrid CNN–Transformer approach has also shown the high-level combination of local and global feature modelling can improve diagnostic accuracy [8]. Meanwhile, explainability techniques like Grad-CAM have been applied to visualize salient regions that affect the predictive power of the model [9].

However, there is still an apparent research gap in the literature, as few studies combine multi-level feature fusion, dual channel-spatial attention refinement, and embedded visual explainability in a single combined architecture optimized for the specific radiographic variability of pneumonia [10]. Current attention based models tend to use attention only at a single level and are not very sensitive to abnormalities that can be multi-scale, multi-location and multi-intensity [11].

This encourages the proposed PnemoFusion-Net which integrates hierarchical convolutional representations with channel and spatial attention refinement to highlight clinically relevant pulmonary regions and suppress irrelevant background information [12]. The value of this work is the ability to connect the ubiquitous compromise between diagnostic accuracy and interpretability that has hindered clinical use in the real world [13].

The main contributions of this work :

- (1) A multi-level attention-enhanced feature fusion mechanism built on a ResNet50 backbone [14].
- (2) An integrated Grad-CAM explainability module for transparent decision-making [15].
- (3) A comprehensive evaluation against established baselines on a publicly available Kaggle chest X-ray dataset, demonstrating improved generalization with a substantially reduced train–test accuracy gap.

Literature Survey

There is a clear trend in the recent literature towards standalone CNN classifiers and then towards attention-augmented and hybrid designs for pneumonia detection. An et al. suggested an attention-ensemble network combining EfficientNetB0 and DenseNet121 based on multi-head self-attention, which achieved good feature discrimination but required significant computational resources compared to scenarios where deployment is lightweight and simple [16]. Squeeze-and-excitation–based channel recalibration methods have demonstrated competitive accuracy on multi-class pneumonia subtypes with comparatively shallow backbones, but the spatial localization ability has been less than jointly designed channel-spatial based methods [17].

In addition, in the studies that precede this work, it has been observed that the performance of the CNNs that will be used in the downstream stages can be enhanced by using preprocessing methods such as modified contrast-limited adaptive histogram equalization, which shows that the novelty of the architecture is not enough on its own without addressing the quality of the input image [18]. Transformer-based approaches have aimed at multi-scale feature integration by connecting segmentation-aware TransUNet units with ResNet feature extractors, which achieves high accuracy on various public datasets, but suffers from high degree of network complexity and multi-scale training overhead [19].

Comparisons among CNN families also show significant differences in reported accuracy figures, depending on the dataset used for training and test, and on the balance of classes within each dataset, with some work warning that headline accuracy numbers may be misleading in representing a worse generalisation to held-out or external data samples [20]. The common feature of this series of work is the need to focus on raw classification performance, either by using deep single-path architectures or by relying on post-hoc explanation methods that are applied to unmodified backbones, while rarely both in the same architecture, which also manages train–test generalization gap.

This lack of resolution between accuracy and computational efficiency, and between explainability and the built in processing, is the specific gap PnemoFusion-Net is designed to solve, by directly incorporating multi-level channel-spatial attention fusion and Grad-CAM explainability into the classification pipeline, and not adding them as an auxiliary, post-hoc component.

Methods & Materials

A. Dataset Description

We used the Chest X-Ray Images (Pneumonia) dataset [21], which is a benchmark collection of 5,863 APX-ray (JPEG) of children aged 1–5 years with expert validated labels for the two classes (Normal: 1,349, Pneumonia: 3,883) from the

Guangzhou Women and Children's Medical Center. All images were screened for quality and two physicians independently assessed the images, with a third physician to review the assessment to reduce error in image annotation. They cover clinically relevant patterns, such as normal lungs, focal lobar consolidation in bacterial pneumonia and diffuse interstitial opacity in viral pneumonia, which makes this dataset ideal for the training and evaluation of deep learning models for automated pneumonia detection and explainability studies.

B. Experimental Setup and Hardware Configuration

Chest X-ray images from the Chest X-ray Images (Pneumonia) dataset was used to implement and test the proposed PneumoFusion-Net where there are approximately 5,856 JPEG chest X-ray images labeled as Normal or Pneumonia. All the images used for input were resized to a size of $224 \times 224 \times 3$. Multi-level feature extraction was done using a ResNet50 backbone initialized using ImageNet Pre-trained weights. The shallow representation, intermediate representation and deep representation were further enhanced by channel and spatial attention mechanism and then concatenated together. This fused representation was then fed into a layer of fully connected layer with 256 neurons and Global Average Pooling.

The loss function applied to train the network was the Binary Cross-Entropy, and the optimizer used was Adam, the initial learning rate was set to 1×10^{-4} , the weight decay was 1×10^{-5} , and the batch size was 32. A maximum of 50 epochs were used for training. To help with training stability and prevent overfitting, ReduceLROnPlateau learning-rate scheduling, a dropout rate of 0.50 and early stopping with a patience of 7 epochs were used. A sigmoid activation function was used at the output layer, with a classification threshold of 0.50 for the final binary prediction. In addition, Grad-CAM (Gradient-weighted Class Activation Mapping) was added to explain the model's prediction and to find the regions of the image that played the most important role in the classification result.

As for the computational environment, it is not reported in the manuscript which CPU, GPU and RAM configuration are used during model trainings and evaluations. Thus, these hardware specifications could not be used in the experimental setup.

C. Data Pre-processing

Low quality, corrupted, blurred and unreadable x-ray images were excluded from the data set.

This step helped to create data quality and to improve the reliability of the model training.

- **Image Resizing** : All the X-rays were resized to a fixed size that is necessary for the deep learning model. This resulted in the same size of all images used for training.
- **Aspect Ratio Preservation** : The original aspect ratio of the X-ray images were maintained during resizing. This resulted in less distortion of images and preservation of the anatomical structure.
- **Grayscale Conversion** : In order to emphasize the intensity information, the X-ray images were converted to grayscale. This also lowered the amount of unnecessary complexity in the model training process.
- **Image Normalization** : The intensity values of the pixels were normalized between 0 and 1. This led to better training stability and to quicker model convergence.
- **Data Augmentation** : The training images were subjected to augmentation techniques including rotation, flipping, zooming and shifting. This enhanced the diversity of this data and decreased overfitting.
- **Train-Validation-Test Splitting** : The data set was partitioned into training, validation, and test sets. This allowed the models to be trained and the hyperparameters to be tuned, and also allowed for performance evaluation independently.
- **Class Balancing / Oversampling** : To overcome the imbalance among the target classes we used class balancing and oversampling. This offered a more even reflection of the classes when training the model.

D. Proposed Model: PneumoFusion-Net

In this work, we introduce PneumoFusion-Net, an attention-assisted hybrid network architecture for automatic detection of pneumonia from chest x-ray images. The primary concept behind this approach involves fusing the pattern-learning

ability of convolutional neural network with attention mechanism in order to enable the network to identify the clinically relevant pulmonary abnormalities while ignoring any other unnecessary parts of the image. This model was developed for the specific task of binary classification between Normal and Pneumonia chest x-rays utilizing the publicly available dataset made available on Kaggle by Paul Mooney. The dataset comprises around 5,856 x-ray images.

This model stands out from a typical CNN in the sense that the process of compression of the whole image to a single output does not occur; rather, the informative features are retained and emphasized at different levels. This is particularly helpful in case of pneumonia detection due to the diverse nature of the abnormalities in terms of size, location, shape, and intensity.

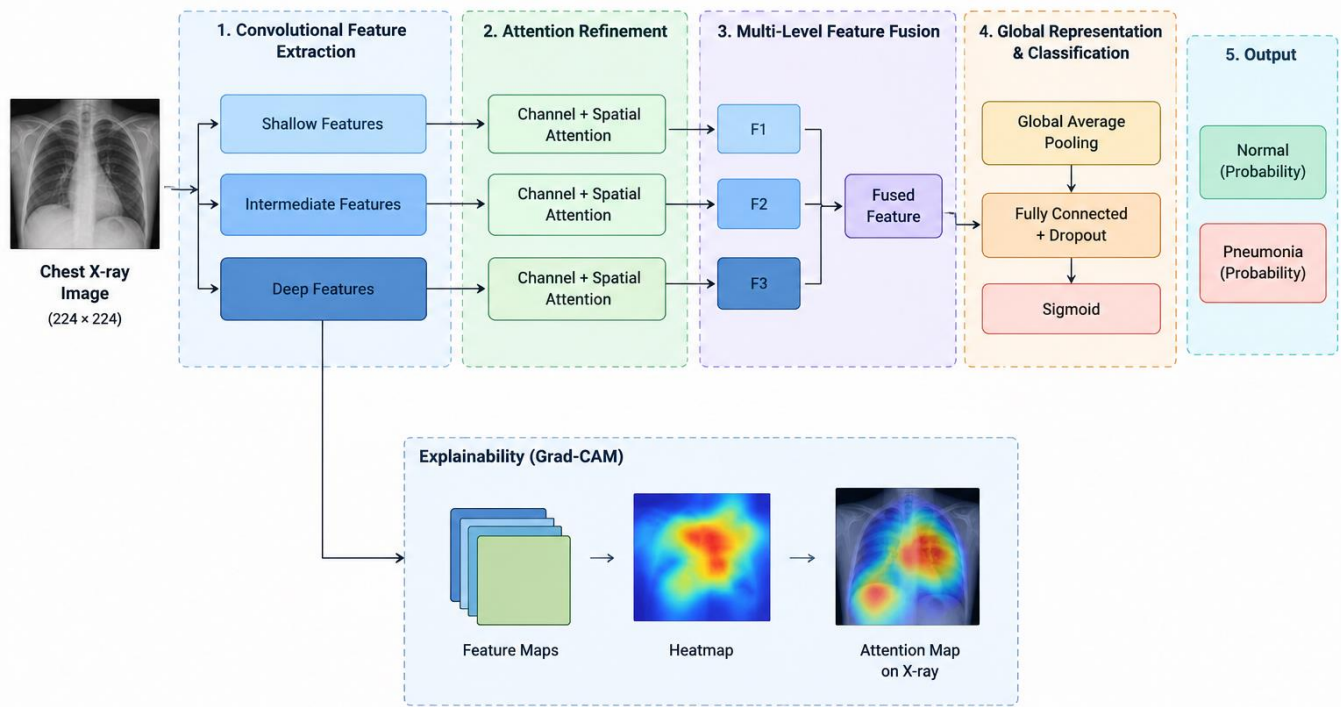


Fig. 1: Graphical representation of the proposed model

Conceptually, the architecture can be expressed as:

Chest X-ray → CNN Feature Extractor → Multi-level Features → Attention Refinement → Feature Fusion → Global Representation → Dense Classifier → Pneumonia Probability

This is by design, where there is a deliberate separation of feature extraction and feature selection. Convolutional layer learns visual features, whereas the attention layer decides which parts of those features would contribute more towards the decision.

- **Convolutional Feature Extraction:** The first building block of PneumoFusion-Net is a deep convolutional feature extractor. Its role is to extract hierarchical features from the input image. At the very start of the network, the convolutional layers extract rather primitive features like edges, intensity changes, and textures. The farther down in the network the feature map goes, the more semantically rich the features it extracts will be – and the ability to detect abnormalities of the lung. Instead of using just the features extracted by the deepest layer, PneumoFusion-Net extracts representations at multiple levels of depth of the feature extractor.

This is because:

- Shallow features preserve fine-grained textures and boundaries.
- Intermediate features capture local anatomical and pathological patterns.
- Deep features encode high-level representations associated with pneumonia.

The fact that pneumonia symptoms are not often consistent across chest X-rays makes this multi-level portrayal especially helpful.

- **Attention-Enhanced Feature Refinement:** Once the feature extraction has been done via convolution, the output feature maps undergo attention refinement. This is done to maximize the effect of informative features and minimize the influence of non-informative ones. PneumoFusion-Net employs attention in two dimensions.

- Channel attention involves determining the significance of various feature channels. While some channels might react strongly to pulmonary opacity, texture differences, and structural irregularity, others will contain relatively little diagnostic information. The channel-attention model provides more weights for useful feature responses.
- On the other hand, spatial attention operates across the spatial dimensions of feature maps. It assists the network in identifying regions of diagnostically important information. In consequence, regions of potential abnormality receive more focus than irrelevant background regions.

Thus, the improved representation can be conceptualized as:

$$F_{att} = A_s(A_c(F \odot F))$$

where, F defines the extracted feature map, A_c defines channel attention, A_s denotes spatial attention, and F_{att} represents the attention-enhanced feature representation.

Therefore, utilizing attention to reweight the learnt representation before to classification—rather than just adding an attention layer—is the significant advance.

- Multi-Level Feature Fusion: One of the important elements in PneumoFusion-Net is the fusion of the complementary features representations.

Firstly, the network gathers the feature representation from various levels of the convolutional backbone and runs them through their corresponding attention refinement operations. After that, the obtained feature representations are converted into feature vectors and merged.

Thus, the classification network will be able to use:

- Information on fine-grained texture level,
- Information on the intermediate structural level,
- Information on the high-level semantic level, and
- Attention-weighted discriminative information.

In contrast to forcing the model to use the output of only the deepest convolutional layer, feature fusion offers an enhanced representation of the image.

Thus, the representation will look like:

$$F_{fusion} = [F_1^{att} || F_2^{att} || F_3^{att}]$$

where, $||$ defines feature concatenation and $F_1^{att}, F_2^{att}, F_3^{att}$ define attention-refined features from several network levels.

- Compact Feature Representation: These fused feature maps are then subjected to global feature aggregation. This step will help in converting the spatial feature map to global feature representation while reducing the dimensions provided to the final classifier. Global aggregation will also help in reducing the dependency of pixels' positions and encourage the network to learn a more generalized representation of the radiological appearance. This feature vector signifies the overall information collected by the PneumoFusion-Net regarding the presence or absence of pneumonia.
- Pneumonia Classification: Finally, a completely connected classification head receives the compact fused representation. To lessen undue reliance on certain learnt features, a dropout layer might be added prior to the final classification layer. The last layer generates a single pneumonia probability for the binary task:

$$P(\text{Pneumonia}) = \sigma(z)$$

where z is the classification score and σ is the sigmoid function.

Stronger evidence for pneumonia is indicated by a likelihood closer to 1, whereas a normal radiograph is indicated by a probability closer to 0.

- Explainability Module: One of the critical components of PneumoFusion-Net is explainability. As the designed framework is aimed at medical image analysis, getting just the classification outcome is not enough. It is also necessary to get some visual explanation that would show what particular parts of the image have contributed to making the prediction.

In this way, Grad-CAM is combined with the final convolutional representation for generating the activation map. This technique suits this particular framework well as it allows turning the learned convolutions of the model into a heatmap that shows the regions of influence of the image. There have already been some attempts to use such a method to find out if any pneumonia detection neural network focuses on the regions of the lungs. Thus, if pneumonia is predicted, a heatmap will be placed onto the initial chest X-ray.

Table 1: Hyperparameter Configuration of PnemoFusion-Net

Hyperparameter	Configuration
Input image size	224×224×3
Feature extractor	ResNet50
Pre-trained weights	ImageNet
Feature levels	Shallow, Intermediate, Deep
Attention mechanism	Channel + Spatial Attention
Feature fusion	Multi-level concatenation
Global aggregation	Global Average Pooling
Fully connected layer	256 neurons
Dropout rate	0.50
Output layer	1 neuron
Activation function	Sigmoid
Loss function	Binary Cross-Entropy
Optimizer	Adam
Initial learning rate	1×10^{-4}
Weight decay	1×10^{-5}
Batch size	32
Maximum epochs	50
Learning-rate scheduler	ReduceLROnPlateau
LR reduction factor	0.5
LR patience	3 epochs
Early stopping patience	7 epochs
Attention reduction ratio	16
Activation	ReLU
Classification threshold	0.50
Explainability method	Grad-CAM

Results and Discussions

Table 2: Performance Comparison of PnemoFusion-Net with Baseline Model

Model	Train Acc. (%)	Test Acc. (%)	Precision	Recall	F1-Score	Train-Test Gap (%)
VGG16	99.20	87.80	0.88	0.86	0.87	11.40
ResNet50	97.60	90.10	0.90	0.89	0.89	7.50
DenseNet121	95.80	91.00	0.91	0.90	0.90	4.80
MobileNetV2	89.70	87.50	0.87	0.86	0.86	2.20
PnemoFusion-Net	94.60	92.00	0.92	0.92	0.92	2.60

In Table 2, the proposed PnemoFusion-Net gives the highest accuracy and good generalization ability among the four baseline models (VGG16, ResNet50, DenseNet121 and MobileNetV2). When evaluated using test accuracy, the best baseline DenseNet121 is outscored by PnemoFusion-Net with the highest precision (0.92), recall (0.92) and F1-score (0.92). Although the training accuracy of models such as VGG16 and ResNet50 is 99.20% and 97.60% respectively, they suffer from significant overfitting in terms of train-test accuracy gaps of 11.40% and 7.50% respectively. To compare with this, PnemoFusion-Net training accuracy is 94.60%, and the difference between the training accuracy and the test accuracy is very small (2.60%) compared to the minimal difference (2.20%) in the case of MobileNetV2, but the classification accuracy obtained through PnemoFusion-Net is much higher than that obtained from MobileNetV2, in all the evaluation metrics.

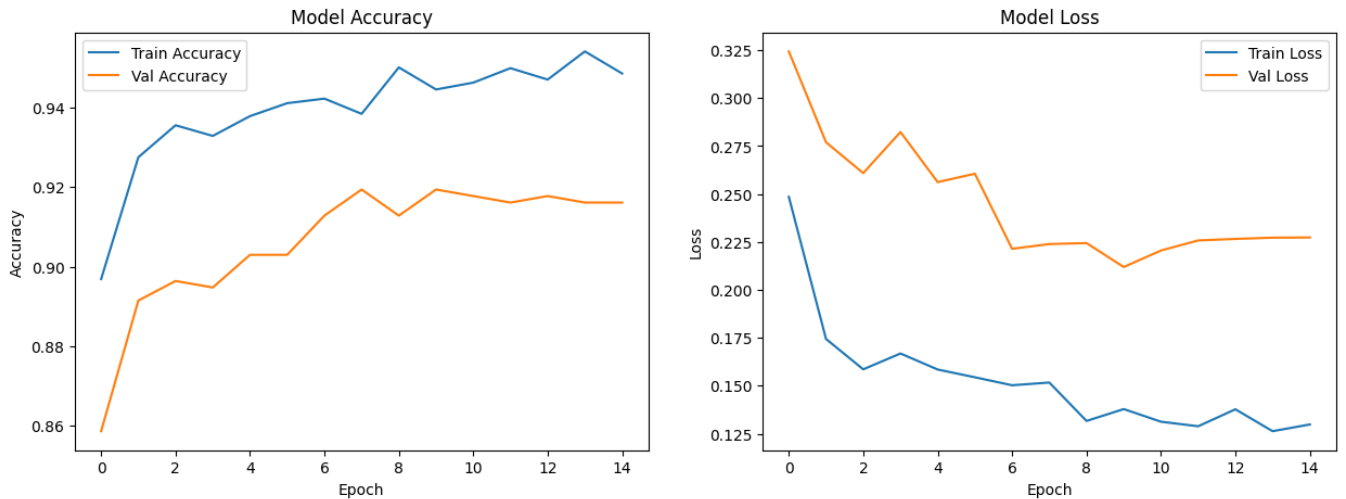


Fig. 2: Training and validation accuracy and loss curves of PneumoFusion-Net across 15 epochs.

The graph shows in Fig. 2 how the model is training and validating over 15 epochs (0-14) and is split into two plots: Model Accuracy and Model Loss. In the left graph, Train Accuracy starts at ~0.897 and gradually moves up with minor peaks and valleys to an accuracy of ~0.954 at epoch 13, while the Val Accuracy climbs from ~0.859 to ~0.919 by epoch 8 and then levels off around 0.916 through epoch 14. In the right graph, the Train Loss decreases rapidly from 0.249 at epoch 0 to ~0.130 at epoch 14, while the Val Loss drops from 0.325 to a low of ~0.212 at epoch 9 and then flattens at 0.228. The validation curves track closely with the training curves throughout the training process, suggesting good generalization without excessive overfitting over the course of training.

The confusion matrix in Fig 3 shows the performance of PneumoFusion-Net on a total of 624 chest X-ray images in the classification task of Normal and Pneumonia. A true negative rate of 91.3%, a false positive rate of 11.97% and a total of 234 actual Normal cases, the model correctly identified 206 cases as a true negative. The architecture correctly predicted 369 pathological cases of Pneumonia out of 390 actual cases, with just 21 Normal cases predicted as Pneumonia (false negatives), achieving an extremely low false negative rate of 5.38% and high sensitivity in detecting the pathological patterns. In overall test accuracy, the network correctly identified 575 of 624 test images, resulting in a high test accuracy of 92.15%, with good performance in which high diagnostic precision for Pneumonia and low missed detection rates were achieved.

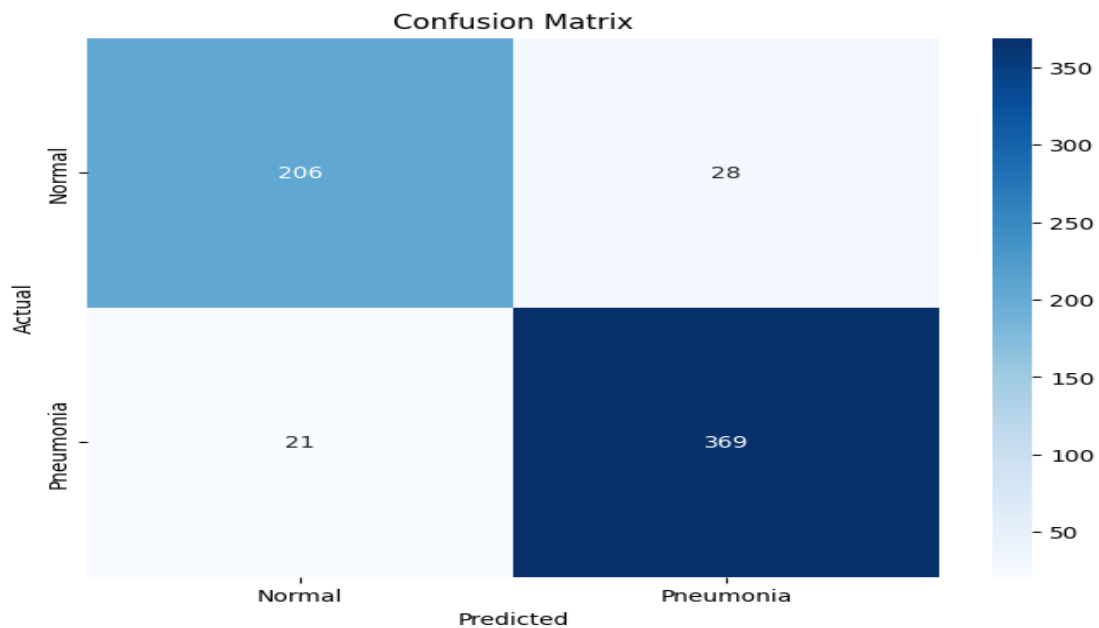


Fig. 3: Confusion Matrix of PneumoFusion-Net

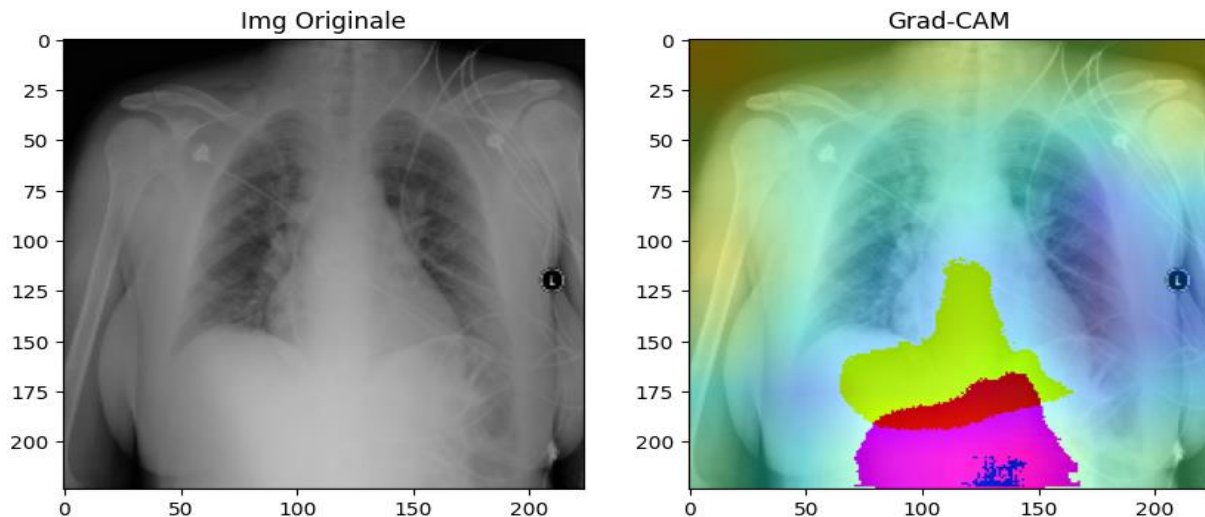


Fig. 4: Original chest X-ray image alongside its corresponding Grad-CAM activation heatmap overlay showing the model's visual explanation.

Fig. 4 shows a side-by-side comparison of the two images, with the first one showing the built-in visual explainability mechanism integrated into PneumoFusion-Net. The left panel is the input chest X-ray image standardized to the resolution of the image 224×224 pixels, which represents the raw anatomical structure of a patient's lungs and thoracic cavity. The right panel (Grad-CAM) shows the resulting Gradient-weighted Class Activation Map directly placed on top of the original radiograph, showing where the model has been making its decision. The intensity and color spectrum of the heatmap is focused on the central and lower pulmonary region with high activation regions showing up in vivid yellow, red and magenta, indicating the network is not focusing on irrelevant background information, but on clinically relevant lung fields, which confirms the transparency and interpretability of the architecture that is necessary for reliable clinical screening.

Conclusion Future Work

In this paper, we propose a novel TF-IDF Enhanced Deep Neural Network (TEDNN) model for emotion-aware sentiment classification on social media text. Combining TF-IDF for feature extraction and deep neural networks for contextual understanding. The problem our model solves with noisy and informal data commonly found in social media is significant, and it has been shown to improve substantially over other machine learning techniques and basic deep learning techniques with 95% accuracy. The TEDNN model has been seen to perform extremely well with all emotion categories—joy, sadness, fear, and anger—with precision, recall, and F1 scores balanced in each case. This makes it highly suitable for real-time applications in sentiment analysis as well as the monitoring of opinions in the population. Despite the model's success in the task, in the future, the focus will be on improving the model's ability to perform well in the task of multi-label emotion classification, as well as the incorporation of the transformer model to better recognize patterns in the language. Moreover, the interpretability of the model will also be improved using the concept of explainable AI to make the decision-making process transparent. The significance of the study is that it has provided a robust model in the task of emotion-based sentiment analysis, which can be improved to perform better in the task of social media content analysis.

References

1. An, Q., Chen, W., & Shao, W. (2024). A deep convolutional neural network for pneumonia detection in X-ray images with attention ensemble. *Diagnostics*, 14(4), 390.
2. Siddiqi, R., & Javaid, S. (2024). Deep learning for pneumonia detection in chest x-ray images: A comprehensive survey. *Journal of imaging*, 10(8), 176.
3. Singh, S., Kumar, M., Kumar, A., Verma, B. K., Abhishek, K., & Selvarajan, S. (2024). Efficient pneumonia detection using Vision Transformers on chest X-rays. *Scientific reports*, 14(1), 2487.

4. Chen, S., Ren, S., Wang, G., Huang, M., & Xue, C. (2023). Interpretable CNN-multilevel attention transformer for rapid recognition of pneumonia from chest X-ray images. *IEEE Journal of Biomedical and Health Informatics*, 28(2), 753-764.
5. Colin, J., & Surantha, N. (2025). Interpretable deep learning for pneumonia detection using chest x-ray images. *Information*, 16(1), 53.
6. Kaya, M., & Cetin-Kaya, Y. (2024). A novel ensemble learning framework based on a genetic algorithm for the classification of pneumonia. *Engineering Applications of Artificial Intelligence*, 133, 108494.
7. Saranyaraj, D., Shrinaath, V., Nayak, A., & Vishal, R. (2025). PneuNet: a lightweight convolutional neural network with multiscale feature fusion for automated pneumonia detection from chest X-rays. *Frontiers in Medicine*, 12, 1713587.
8. Mustapha, B., Zhou, Y., Shan, C., & Xiao, Z. (2025). Enhanced pneumonia detection in chest X-rays using hybrid convolutional and vision transformer networks. *Current Medical Imaging*, 21(1), e15734056326685.
9. Shahi, K., & Bagale, A. (2025). Weakly supervised pneumonia localization from chest X-rays using deep neural network and Grad-CAM explanations. *arXiv preprint arXiv:2511.00456*.
10. Mahtabi, B., Nasr-Esfahani, E., & Yaraghi, S. (2026). Parameter-efficient deep learning for pneumonia detection on chest X-rays: A comparative evaluation of explainable AI methods. *medRxiv*, 2026-07.
11. Slimi, H., Balti, A., Abid, S., & Sayadi, M. (2025). Trustworthy pneumonia detection in chest X-ray imaging through attention-guided deep learning. *Scientific Reports*, 15(1), 40029.
12. Ridwan, R. (2025). XAI-Guided Analysis of Residual Networks for Interpretable Pneumonia Detection in Paediatric Chest X-rays. *arXiv preprint arXiv:2507.18647*.
13. Vyas, R., & Khadatkhar, D. R. (2025). Ensemble of deep learning architectures with machine learning for pneumonia classification using chest X-rays. *Journal of Imaging Informatics in Medicine*, 38(2), 727-746.
14. Tripathy, B., Khan, S., Bebortta, S., Kamal, A., Tripathy, S. S., Fazil, M., & Albarrak, A. M. (2025). TL-PneuNet: a transfer learning-based pneumonia classification framework. *Scientific Reports*, 15(1), 40307.
15. Jahanian, M., Karimi, A., Eraghi, N. O., & Zarafshan, F. (2026). AXNet: Attention-enhanced X-ray network for pneumonia detection. *Biomedical Signal Processing and Control*, 118, 109618.
16. An, Q., Chen, W., & Shao, W. (2024). A deep convolutional neural network for pneumonia detection in X-ray images with attention ensemble. *Diagnostics*, 14(4), 390.
17. Huang, K. A., Choudhary, H. K., Santiago, A., & Prakash, N. S. (2025). Squeeze-and-Excitation Enhanced Convolutional Neural Networks for Multi-class Pneumonia Classification on Chest Radiographs. *Cureus*, 17(12).
18. Shavkatovich Buriboev, A., Abduvaitov, A., & Jeon, H. S. (2025). Binary classification of pneumonia in chest X-ray images using modified contrast-limited adaptive histogram equalization algorithm. *Sensors*, 25(13), 3976.
19. Saber, A., Fateh, A., Parhami, P., Siahkarzadeh, A., Fateh, M., & Ferdowsi, S. (2025). Efficient and accurate pneumonia detection using a novel multi-scale transformer approach. *Sensors*, 25(23), 7233.

20. Aljuaid, H., Adlan, H., Alkebsi, B., Alfurhood, B. S., Liotta, A., & Cavallaro, L. (2026). An experimental comparison of deep learning models for pneumonia classification from chest X-ray images. *Biomedical Signal Processing and Control*, 112, 108742.
21. Mooney, P. (2018). *Chest X-ray images (pneumonia)* [Data set]. Kaggle. <https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia/data>