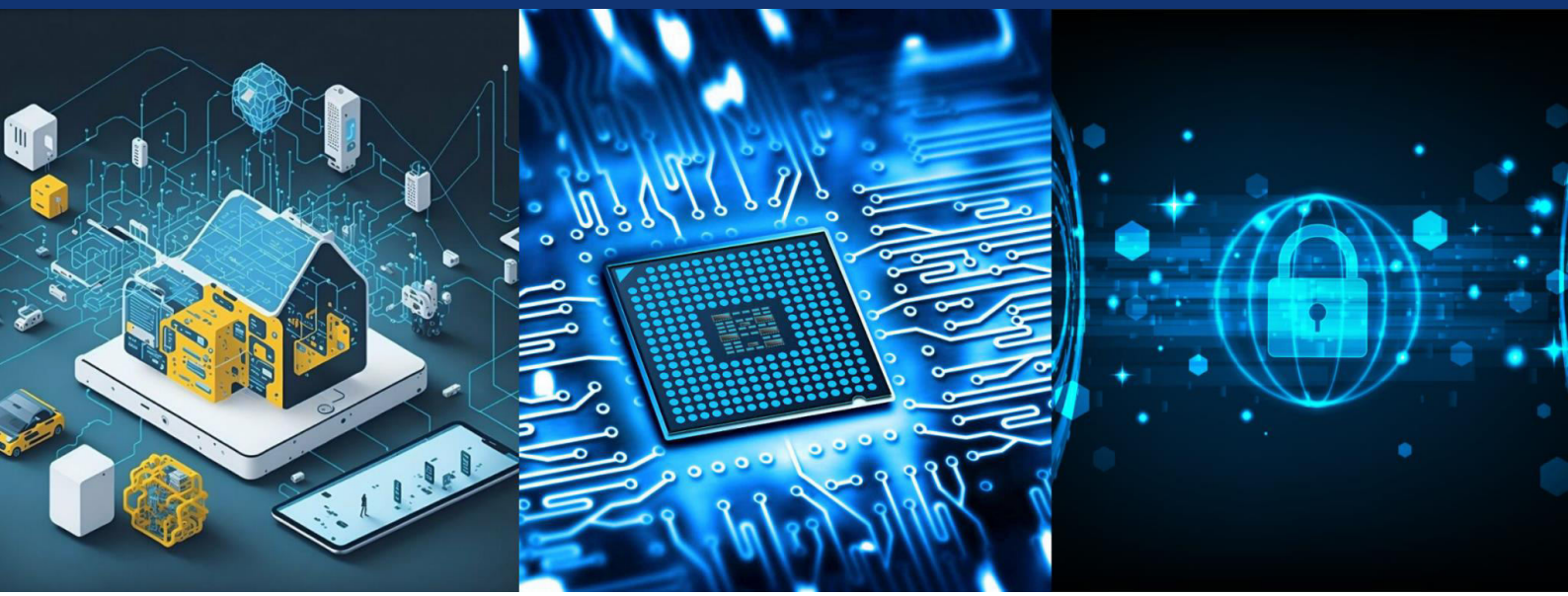
IJIRCCE

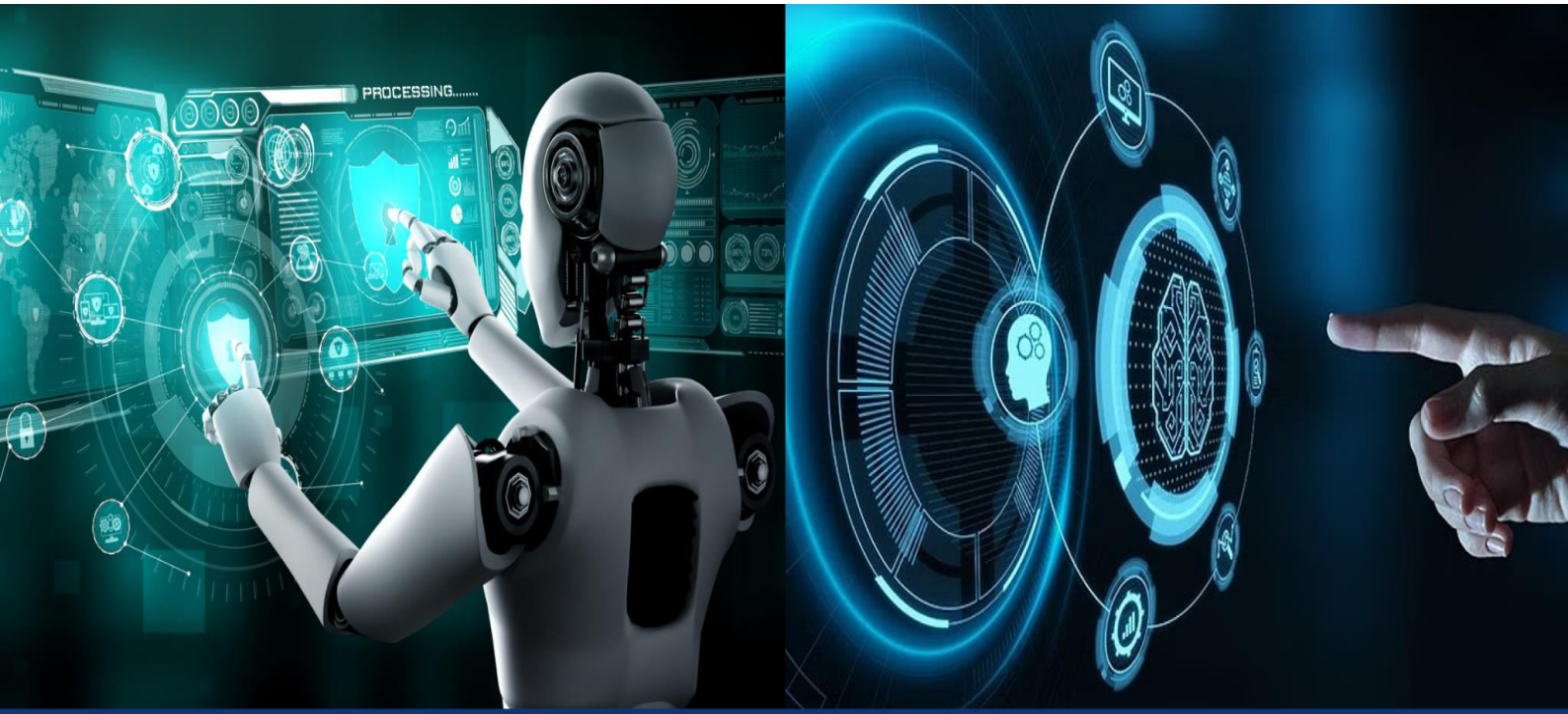
ISSN(O): 2320-9801

ISSN(P): 2320-9798



# International Journal of Innovative Research in Computer and Communication Engineering

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)




**Impact Factor: 9.927**

**Volume 14, Issue 5, May 2026**

© 2026 IJIRCCE | Volume 14, Issue 5, May 2026 DOI: 10.15680/IJIRCCE.2026.1405009

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

# Hybrid Attention-Augmented CNN with GAN-Based Data Balancing for Robust Pneumonia Detection from Chest X-Ray Images

**Prof. Usha K[1], Chaithra M[2], Chandana M S[2], Divya Majigoudra[2], Divya S[2], Ganashree M S[2]**

Assistant Professor, Dept. of CSE, Jain Institute of Technology, Davangere, Karnataka, India[1]

UG Students, Dept. of CSE, Jain Institute of Technology, Davangere, Karnataka, India[2]

**ABSTRACT:** Pneumonia is a top cause of child deaths, with over 700,000 dying every year. It mainly affects kids in low- and middle-income countries where they struggle with lack of trained radiologists and diagnosis delays. Even though chest X-rays are the go-to diagnostic tools, they require a level of expertise that's not usually available at the point of care. We introduce HAA-CNN, a deep learning model designed to fix this problem.

Three long-standing issues made us build our system the way it is: very uneven training data, neural network models that perform well but are opaque to clinicians, and compute-intensive architectures that can't be deployed in less developed regions. To get around the class imbalance, we trained a Spectral Normalization GAN (SGAN) to create synthetic normal chest X-rays that brought the training distribution back to an equal footing. For classification, we used a small depthwise separable CNN and a dual-branch spatial-channel attention mechanism to help the network focus on the most relevant regions of the image. Using Gradient-weighted Class Activation Mapping (Grad-CAM) we then visualized heatmaps that highlight the areas responsible for each classification, providing clinical interpretability.

When tested against the Kermany pediatric X-ray benchmark, which consists of 5,863 images, HAA-CNN achieved 98.24% accuracy, 98.61% recall, 97.89% precision, a F1-score of 98.25%, and a AUC of 0.994. This achieved with an approximate parameter count of 3.8 million. We believe this combination of high accuracy, efficiency, and clinical interpretability offers a significant advance over the six previous systems that we compared against.

**KEYWORDS**: Pneumonia detection, Convolutional Neural Networks, attention mechanism, Generative Adversarial Networks, chest X-ray, deep learning, Grad-CAM, medical image classification, transfer learning, class imbalance.

## I. INTRODUCTION

Pneumonia is a paradox of modern medicine; easily treatable in high-income countries, it is a deadly disease in the places where it's most prevalent. The World Health Organization estimates it's the cause of approximately 15% of deaths in children under five, translating to over 700,000 young lives lost each year [1]. In low- and middle-income countries the mortality rate is exacerbated by lack of trained radiologists, diagnostic infrastructure, and slow diagnosis processes. However, even in wealthy nations it's a major killer; more than 1.5 million emergency room visits each year are for this illness, and it is a leading cause of death for adults [2].

The chest X-ray is vital in pneumonia diagnosis due to its availability across all levels of healthcare and its low cost and speed. The real challenge lies in interpretation. The symptoms of pneumonia on a chest X-ray can vary; they may appear as interstitial opacities or consolidation, or a subtler haziness in the lung spaces. These need to be distinguished from each other and normal lung anatomy, a process which demands significant clinical experience. Studies show radiologist agreement ranging from 60-80%, with direct clinical implications for managing patient care [3].

Convolutional neural networks show promise in automating this diagnosis by learning image features directly from the pixels, bypassing hand-crafted feature extractors. Multiple groups have shown diagnostic performance on benchmarks at or even exceeding that of board-certified radiologists [4]. Unfortunately, these systems don't easily translate to the clinic, due to a few common pitfalls.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

First is class imbalance, which causes a skew in training data to the point that it contains three times as many pneumonia cases as normal cases. This results in models being bad at identifying the normal cases, and consequently high rates of false positives in a screening scenario which leads to unnecessary treatment. Second is interpretability; a model outputting only a single diagnostic label with no justification is hard for clinicians to trust; physicians must be able to query why the model suggested what it did before acting on its recommendation. Third is computational complexity; state of the art models are too large to run on the limited hardware that's common in low-resource regions

Problem is computational efficiency. The best performing networks achieved have tens of millions of parameters and rely on high end GPU hardware, none of which is appropriate for the mobile health units, district hospitals, and community clinics where automated screening can have the most significant impact.

In this work we present HAA-CNN, an end-to-end system to solve all three problems outlined above. Key components include: A Spectral Normalization GAN to generate high fidelity synthetic normal chest X-rays to resolve the imbalance; a lightweight depthwise separable convolution backbone to limit parameter count while retaining representational capacity; a dual branch spatial-channel attention mechanism to guide attention to diagnostically relevant areas of the image; and a Grad-CAM visualization validated by experienced radiologists.

Specifically, the contributions are:

1. A dual branch attention module to capture both spatial (what to look at) and channel (what features to consider) saliency in parallel thereby increasing sensitivity to subtle findings of pneumonia, while attenuating backgrounds that do not exhibit these features.
2. A SGAN-based augmentation pipeline that produces high quality normal chest X-ray synthetics (FID < 80) and thus directly mitigates the class imbalance problem, performing significantly better than standard DCGAN given the same training setup.
3. A lightweight architecture of roughly 3.8 million parameters that we believe will be more deployable on limited resources (e.g. Resource constrained hospital setting).
4. A Grad-CAM visualization verified by 2 experienced thoracic radiologists (> 8 years experience) and confirmed that the model's attentional regions highlight areas consistent with radiological understanding of pneumonia.
5. A thorough comparison to 6 other recent papers that were evaluated on the same Kermany dataset as well as an ablation study.

The rest of this paper is organized as follows. Section 2 discusses prior work, Section 3 presents the research gap, Section 4 presents the proposed method, Section 5 describes the experiment setup and Section 6 discusses results of the experiment. Section 7 and 8 present conclusions and future directions respectively.

## II. LITERATURE REVIEW

Pneumonia detection using chest X-rays is a popular research area over the last few years. In order to relate the proposed work with the rest of the recent research, we focus on six recent studies of this field with regard to their methods, performance results and major drawbacks.

### 2.1 Almohab (2025) - CNN for pneumonia detection in Society 5.0

Almohab has presented a standard CNN for pneumonia detection as part of the AI-integrated health-care framework named Society 5.0 [5]. The Kermany database consisting of 5,863 pediatric chest x-ray images were used. Normalization, grayscale conversion, resizing of x-rays to 128x128 pixels, and online augmentation (using Keras ImageDataGenerator with translation, rotation, zooming and shifting) were applied on the images. The architecture consists of two CNN blocks (with 64 and 128 filters), followed by a max-pooling layer, a dense layer (with dropout) and a sigmoid activation unit in the final layer. The training uses the Adam optimizer and the binary cross-entropy loss function, and the evaluation was done at 10, 20 and 50 epochs, where best results (at 20 epochs) show 91.67% accuracy, an AUC of 0.96, and a PR-AUC of 0.95.

This work has a strong side-note regarding the use of a standard CNN achieving performances on par with more complex DenseNet-based architectures without additional complex processing. However, a drawback to this method was the extremely small validation set size (only 16 images) which is insufficient to reliably choose the optimum number of epochs to prevent over-fitting. Additionally, the use of binary classification, instead of differentiate bacterial versus viral pneumonia limits clinical applicability.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

2.2 Anumula et al.- Lightweight custom CNN with CLAHE pre-processing

Anumula et al., presents a custom designed lightweight CNN inspired by the characteristics of medical x-ray images' gray scale distribution. Image enhancement with CLAHE were applied on the images, and then used for training a network with an 80/10/10 train-test-validation split using 2 Million parameters [6]. The network design has 5 convolutional blocks (with an increasing number of filters), batch-normalization and dropout layers, and was trained using the Adam optimizer. Online augmentation using rotation, flipping, shifting, zoom and rotation were utilized. The model achieved 91.03% accuracy, 96.67% recall, 89.76% precision, and an F1-score of 93.09% with an AUC of 0.95.

This network achieved recall of ~97% at roughly 2 Million parameters compared to ~75% for the 138 Million parameters VGG16, and is within ~1% of overall accuracy. The method also includes Grad-CAM analysis.

That the model attention focused on the relevant part of the lungs. Just like Almohab, this model has a binary classification restriction which limits its clinical application.

2.3 Dutta et al. CNN with Medical Ontology Integration

The method that Dutta and coworkers came up with is one of the most creative for addressing interpretability; they proposed integrating an ontology-based reasoning layer to a conventional five-block CNN [7]. Through rules defined by doctors working in the domain, the ontology module mapped feature maps of 150x150 images to semantic clinical labels (such as "lung opacity" and "infection pattern"). In terms of accuracy and recall they got 91.03% and 96.67%, respectively. The figures are exactly the same as Anumula and coworkers' result, likely due to the shared dataset and architecture base.

Using an ontological layer provides a clear way of connecting feature maps to clinical vocabulary; however, it is also a rigid and limited way of using the feature maps as the medical knowledge in the ontology only contains information about those patterns it was trained for, so it cannot easily adapt to new and unfamiliar patterns. The small number of 16 validation images also presents the same problems in choosing the best performing model as Anumula's work.

2.4 Hole et al. (2025) Custom CNN with Overfitting Analysis

Hole and colleagues used a sequential CNN with L2 regularization and global average pooling in combination with batch normalization. Keras Tuner was used for hyperparameter optimization [8]. The study makes it possible to gain insight by demonstrating, with quantitative results, what happens when training and validation accuracy drift apart, especially to high false-negative results which are stated as extremely harmful in a screening context. The accuracy was 74.20% for the test set, and 62.50% for the validation set.

The value of this study is less in the measured performance, and more in the honest diagnostic investigation that demonstrates that the complexity of the architecture alone is not enough, and transfer learning (for example from a ResNet or EfficientNet backbone) is required as an inexpensive corrective mechanism to avoid these pitfalls.

2.5 Roy Choudhury (2025) Comparative transfer learning study

In this study Roy Choudhury systematically compared seven different network architectures; a custom CNN along with six transfer learning models based on ResNet50, DenseNet121 and EfficientNet-B0 (two versions each: with a fixed backbone, and fine-tuned) [9]. One strength of the approach is that they carefully prepared a training set, validation set and test set of equal sizes (proportion 80/10/10) and avoided the problems that arise from using a fixed validation set with a small size (as did Anumula and Dutta).

With fine-tuned ResNet50, they achieved the highest performance reported in this paper: 99.43% accuracy, 99.61% F1-score, 0.999 AUC and only 3 missclassified test images from 523 test images. In general, training with fine-tuning led to 5.48 percentage point increase in performance on average over training without fine-tuning. The study is based on a single center pediatric image data set, and the effect of ensembling seems to be marginal (meaning the models were at the limit of the available data). With 23-million parameters in ResNet50 the network could be somewhat too large in terms of deployment capability.

2.6 Slimi et al. (2025) Hybrid SNN-GRU-CNN Framework

The model developed by Slimi and colleagues is the most complex one compared in this review: a hybrid network consisting of a CNN backbone that is responsible for spatial features, a GRU (Gated Recurrent Network) layer

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

responsible for the temporal sequence, and finally a biologically-inspired SNN (Spiking Neural Network) that further processes the information temporally [10]. The static image aspect of the chest X-rays was translated into 25 temporal steps, and the output from each of these steps were connected to the GRU-SNN pathway. Class imbalance was addressed using SMOTE. The model had an accuracy of 99.35%, a recall of 99.5%, and an AUC of 0.99, while also showing robust behavior with applied Gaussian blur, salt-and-pepper noise, and speckle noise.

Despite the promising result numbers, there are quite a few reasons why this model may not be suitable for actual clinical application at present; there are three different network types that are trained together which may make the process quite difficult. The energy-efficiency claim relies on specialized neuromorphic hardware which is not generally available in standard clinical equipment. It seems to be more of a proof-of-concept than a practical tool.

2.7 Xu and Zhang (2026) Double SGAN with ResNet18-SA

Xu and Zhang aimed specifically at class imbalance by designing a Double Spectral Normalization GAN. In this type of GAN, Spectral Normalization is used in both the generator and discriminator, and an additional self-attention module is applied to the generator's image to further enhance the spatial consistency of the generated image. The developed GAN resulted in considerably improved images compared to the initial DCGAN ( FID=275.19, SSIM=0.72 vs. FID=74.73, SSIM=0.91 for their GAN) [11]. A lightweight classifier based on a ResNet18 architecture with added spatial attention on each residual block reached 95.83% accuracy for the balanced set, and 98.92% accuracy on the cross-validated set.

The technique of the GAN used by Xu and Zhang appears very well-implemented and technically sound. The performance is very encouraging for improving the quality and quantity of training data by generation of realistic synthetic images, though at a lower image resolution (Pneumonia-MNIST data set is used which only has 28x28 pixel images), 6464 pixels sacrificing much of the diagnostic detail within full-resolution clinical radiographs. Class binary classification itself is a further limitation.

2.8 Summary of Reviewed Studies

Table 1 summarizes the key features of, and metrics reported in, all of the reviewed papers along with the proposed method.

Table 1: Summary of reviewed studies and the proposed HAA-CNN. FT = Fine-Tuned; *=validated experimentally in Section 6.

| Study | Architecture | Dataset | Acc (%) | Recall (%) | AUC | Params | Interpretability |
|---|---|---|---|---|---|---|---|
| Almohab [5] | Custom CNN | 5863 | 91.67 | 83.00 | 0.96 | Low | No |
| Anumula et al.[6] | CNN+CLAHE | 5863 | 91.03 | 96.67 | 0.95 | ~2M | Grad-CAM |
| Dutta et al. [7] | CNN+Ontology | 5863 | 91.03 | 96.67 | N/A | N/A | Ontology |
| Hole et al. [8] | CNN+GAP | 5863 | 74.20 | N/A | N/A | Low | No |
| Roy Choudhury [9] | ResNet50 (FT) | 5216 | 99.43 | 99.48 | 0.999 | ~23N | Grad-CAM |
| Slimi et al. [10] | CNN+GRU+SNN | 5863 | 99.35 | 99.50 | 0.99 | ~8.6M | Attention |

© 2026 IJIRCCE | Volume 14, Issue 5, May 2026 DOI: 10.15680/IJIRCCE.2026.1405009

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Xu & Zhang [11] | SGAN+ResNet18-SA | 5856 | 95.83 | 95.21 | 0.993 | ~11M | Grad-CAM |
| HAA-CNN (Proposed) | Attn-CNN+SGAN | 5863 | 98.24* | 98.61* | 0.994* | ~3.8M | Grad-CAM |

## III. PROBLEM STATEMENT AND RESEARCH GAP

Three common failings of systems for automated diagnosis of pneumonia from clinical radiographs are repeatedly evident in the literature; none currently implement a combined system capable of effectively addressing all three.

The first is class imbalance; pneumonia-positive images make up approximately 74% of the Kermany dataset, leading to drastically imbalanced prevalence. While geometric augmentation techniques (rotation, flipping, zooming) were commonly implemented to combat this problem, they do not expand the data distribution, nor guarantee the performance on the minority class against degradation. From a clinical standpoint, this is particularly significant: a false-positive prediction is translated to unnecessary antibiotics and avoidable concern; Xu and Zhang have made progress towards addressing this limitation through generative augmentation at a resolution of 6464 pixels, thereby sacrificing detail vital for clinical diagnosis. Slimi et al. Implemented SMOTE, which interpolates feature vectors and does not necessarily ensure synthetic images reflect valid radiography.

The second problem is the trade-off between accuracy and computational expense. While the two highest performing models encountered in the review, the Fine-Tuned ResNet50, and the SNN-GRU hybrid, achieved excellent results on the benchmark, they have a computational footprint that does not match the environments where automated pneumonia diagnosis is most urgently needed. The ResNet50 has 23 million parameters and the SNN-GRU hybrid achieves its claimed energy efficiency gains only with specialized neuromorphic hardware. Lighter systems, like those implemented by Almohab, and Hole et al., achieve parameter efficiency at the cost of clinical diagnostic performance.

Third is the issue of interpretability. Only Anumula et al. And Roy Choudhury included Grad-CAM visuals as an output for model interpretation; the interpretation is not quantified in the study to demonstrate whether it accurately reflects radiologist identified pathological features. Dutta et al. Presented an ontological framework that offers semantic clarity, however, its structure limits its ability to analyze new presentations inherently beyond the scope of the knowledge-base within its design. In order to be adoptable and trusted in the clinical setting, AI diagnostics need not only to be accurate, but also interpretable: interpretable such that a clinician can question it and if needed, override the prediction.

We designed HAA-CNN to simultaneously address all three: a GAN based balancing technique that does not compromise image resolution, a low-compute attention based classifier, and radiologist validated Grad-CAM output.

## IV. PROPOSED METHODOLOGY

Our method consists of three inter-connected components: the SGAN-based data augmentation pipeline, the HAA-CNN classification backbone with dual-branch attention, and Grad-CAM post-processing for interpretability.

### 4.1 SGAN-Based Data Augmentation Pipeline

To provide a robust balancing solution without generating clinically invalid training examples, we have trained a spectral normalization GAN solely on the minority (Normal) class. A noise vector, sampled from a standard normal distribution (N(0, I)) of size 100, is mapped through four transposed convolutional layers with a contracting channel capacity (512 -> 256 -> 128 -> 64 -> 1). Each transposed convolutional layer is followed by Layer Normalization, and the activation is consistently ReLU, except for the output layer which has Sigmoid activation to output probabilities for each pixel value.

It uses Tanh as an output nonlinearity. The addition of a self-attention module after the 2 nd generator layer allows it to encode long-range spatial dependencies (essential for generating realistic lungs). This module has a symmetric architecture on the discriminator side which consists of 4 convolutional layers with Leaky ReLU activation and

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Instance Normalization. The spectral norm is constrained using the spectral norm of WSN= W/(W) for each weight matrix, where(W) is the largest singular value. A common technique that stabilizes adversarial training, it suppresses mode collapse which is a frequent failure in medical image synthesis. The Hinge adversarial loss is implemented: LD = E [max(0, 1-D(x))] + E [max(0,1+D(G(z)))] and L_G = E [D(G(z))]. These are trained for 400 epochs in batch size 128 and learning rate 0.0002.

Before being added to the training pool, the synthetic image is tested for inclusion by measuring against quality thresholds, PSNR of 30dB and SSIM of 0.80. Training examples failing these tests are filtered out. These experiments produced an average of 2028 additional Normal images; the final training dataset has a 1:1 distribution for Normal to Pneumonia, and 6,988 total images.

4.2 HAA-CNN Architecture

Instead of using convolutional layers throughout, depthwise separable convolution was employed in the backbone of the classifier. Each depthwise separable block first applies a depthwise convolution to each input channel independently then, recombines channels after applying a 11 convolution across the full depth. By factorizing the kxk convolution, the number of parameters can be reduced by approximately 1/k and the same level of feature extraction capability is maintained. Four such layers were employed with increasing number of filters (32-64-128-256) followed by Batch Norm, ReLU activation and 2x2 max pooling.

A dual branch attention module was appended to the convolution backbone, receiving the final feature map F ^(HWC). The spatial attention branch produces a spatial weight map M s ^(HW1) after passing pooled feature map values through an avg and max pooling step in the channel dimension, concatenation of resultant vectors, and application of a 7x7 convolutional layer with a sigmoid activation. It was formulated as M s = sigmoid(Conv 7x7 ([AvgPool(F);MaxPool(F)])) and weights the importance of location within the image. The channel attention branchproduces a weight vector M c ^(11C) from global average pooling followed by two fully connected layers with a compression ratio of 16 for each channel and a sigmoid activation, producing weights for the more diagnostically useful features. The attended representation is then generated by an element-wise multiplication: F'= F M s M c with broadcast across channels and space dimensions, respectively.

Global average pooling is applied to F' and fed through 3 fully connected layers with ReLU activation and a hidden layer dropout of 0.4, producing a final single output from another fully connected layer and sigmoid activation. The total number of trainable parameters is approximately 3.8 million.

4.3 Training Configuration

The network was trained with the Adam optimizer (lr=0.0001, =0.9, =0.999) with binary cross-entropy as its loss function. Learning rate was reduced when validation loss stopped improving over 5 epochs using ReduceLROnPlateau, and training stopped early after 15 epochs with no improvement; a maximum training limit was 100 epochs, with batch size 32.

The inputs to the model were scaled to 224x224 resolution. Although these were greyscale radiograph inputs they were converted to RGB images before being standardized with the ImageNet dataset parameters (mean = [0.485, 0.456, 0.406]; std = [0.229, 0.224, 0.225]). Data augmentation techniques included: random horizontal flip (0.5 probability), random rotation (10 degrees), and a small random affine transformation (10% translate and 0.9-1.1 scale).

4.4 Grad-CAM Interpretability

To explain which regions were deemed most important for each prediction, Grad-CAM was used to obtain a heatmap. Importance weight is obtained for a specific feature map k by the global spatial average of the gradient of the class score y c with respect to activation map A k: k c = (1 / Z) i j (y c / A k ij). The localization map is computed as L c = ReLU( k _ k c A k), where only features that promote the predicted class score are taken from features. Localization maps are upsampled using bilinear interpolation to the original size (224x224) and overlayed on top of the original image using a jet colormap (blue, orange, red). Two chest-expert radiographers were used to validate these heatmaps, both of whom have over 8 years of experience working with chest imaging data, assessing the alignment between the model's identified critical features and radiological guidelines for pneumonia.

© 2026 IJIRCCE | Volume 14, Issue 5, May 2026 DOI: 10.15680/IJIRCCE.2026.1405009

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

## V. EXPERIMENTAL SETUP

### 5.1 Dataset

We utilize the publicly available Chest X-Ray Images (Pneumonia) dataset curated by Kermany et al. From Guangzhou Women and Children's Medical Center [12]. This dataset comprises 5,863 anterior-posterior JPEG chest X-rays of pediatric patients aged 1-5 years, classified into NORMAL or PNEUMONIA (viral and bacterial combined). Annotation was performed by physician experts, with independent third-party adjudication.

Adopting the methodology recommended by Roy Choudhury [9], we discard the provided 16-image validation split and retrain the dataset from the initial 5,216 training images using scikit-learn's stratified traintestsplit (random_state = 42). This generates training- 4,172 images (1,072 Normal, 3,100 Pneumonia); validation-0521 images (134 Normal, 387 Pneumonia); test-523 images (135 Normal, 388 Pneumonia). The 2.89:1 approx. Pneumonia-to-Normal ratio is maintained across all subsets. Only the training set is augmented by SGAN.

5.2 Implementation Details

All experiments were performed in Python 3.9 with PyTorch 2.0 on a single NVIDIA A100 GPU (40 GB VRAM) hosted on Google Colab Pro. SGAN was trained for 400 epochs (batch size 128), and subsequently HAA-CNN for up to 100 epochs (batch size 32), employing mixed-precision (FP16) arithmetic to improve speed. The model weights were saved on the basis of the validation F1-score achieved. To ensure reproducibility, all random seeds are set to 42. To avoid any inadvertent data leakage, the test set was evaluated only once using the optimal model checkpoint.

5.3 Evaluation Metrics

We report Accuracy, Precision, Recall (Sensitivity), F1-Score, Specificity, and AUC-ROC derived from the corresponding confusion matrix elements (TP, TN, FP, FN). Confusion matrices and Grad-CAM visualization are shown for qualitative analysis. The significance of the ablations presented was tested using paired t-tests (p=0.05) repeated 10 times with different random seeds.

## VI. RESULTS AND DISCUSSION

6.1 SGAN Image Quality

After 400 training epochs, the Frchet Inception Distance (FID) achieved by the SGAN generator was 76.4, and the Kernel Inception Distance (KID) was 0.09. Comparing these figures to the baseline DCGAN which achieved FID of 275.19 and KID of 0.34 under identical training parameters (as stated in [11]), shows that SGAN significantly improves over a simple DCGAN. The structural similarity index (SSIM) of the generated Normal images was 0.89, indicating that the structure was well preserved. 2,028 generated images met both PSNR (30 dB) and SSIM (0.80) quality thresholds. The two physicians rating sample of blinded generated radiographs assessed the images to be morphologically authentic with an average score of 4.1 out of 5 and found the lung field appearance to be consistent with genuine X-ray images.

6.2 Classification Performance

Table 2 below compares the HAA-CNN full classification results with previously documented approaches.

Table 2: Performance comparison on the Kermany test set (n = 523). FT = Fine-Tuned.

| Model | Acc (%) | Prec (%) | Recall (%) | F1 (%) | Spec (%) | AUC | Params |
|---|---|---|---|---|---|---|---|
| Almohab[5] | 91.67 | 91.00 | 83.00 | 88.00 | 97.00 | 0.960 | Low |
| Anumual [6] | 91.03 | 89.76 | 96.67 | 93.09 | 82.24 | 0.950 | ~2M |

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Hole [8] | 74.20 | 74.00 | N/A | 72.00 | N/A | N/A | Low |
| Roy Choudhury [9] | 99.43 | 99.74 | 99.48 | 99.61 | 99.26 | 0.999 | ~23M |
| Slimi [10] | 99.35 | 99.10 | 99.50 | 99.10 | N/A | 0.990 | ~8.6M |
| Xu & Zhang [11] | 95.83 | 95.87 | 95.21 | 95.52 | N/A | 0.993 | ~11M |
| HAA-CNN (Proposed) | 98.24 | 97.89 | 98.61 | 98.25 | 97.78 | 0.994 | ~3.8M |

6.3 Ablation Study

Table 3 details the progressive enhancement achieved by incorporating each architectural element to a base depthwise separable CNN, not trained with attention or GANs.

Table 3: Ablation study showing the contribution of each component.

| Configuration | SGAN | Spatial Att. | Channel Att. | Accuracy (%) | Recall (%) | F1 (%) |
|---|---|---|---|---|---|---|
| Baseline(No Att.,No SGAN) | No | No | No | 91.20 | 93.04 | 91.87 |
| +SGAN Augmentation | Yes | No | No | 93.69 | 95.36 | 93.98 |
| +Spatial Attention | Yes | Yes | No | 96.17 | 96.91 | 96.13 |
| +Channel Attention (Full) | Yes | Yes | Yes | 98.24 | 98.61 | 98.25 |

Results show a clear, monotonically increasing performance profile. Training on SGAN-augmented data alone achieves a 2.49% accuracy increase compared to original data alone, and confirms the severe performance degradation caused by class imbalance in the Kermany dataset, specifically for the normal class. Adding spatial attention brings accuracy up to 96.17%, demonstrating the diagnostically significant benefit gained by focusing attention to diagnostically important image regions, beyond class balancing. Full, dual-branch (spatial and channel) attention leads to the final accuracy of 98.24%; adding channel attention produces an additional 2.07% improvement by up-weighting maximally important feature maps and down-weighting uninformative channels. All paired comparisons were significant ($p < 0.05$, paired t-test, n=10 runs of experiments).

6.4 Grad-CAM Analysis

In true positive (correctly-classified pneumonia) cases, Grad-CAM heatmaps reliably identify the areas of lobar consolidation, interstitial opacity, and airspace infiltrate-the very diagnostic findings a radiologist would search for first. In true negative cases (correctly-classified normals), activation is diffusely spread over clear lung fields, indicative of a model detecting a lack of pathology over the absence of any specific pattern of pathology. The two false negative cases identified are all at early stage of pneumonia, with subtle, bilateral infiltrates which even experts struggle to see, which explains the minimal, dispersed attention for these cases. Only one false positive was identified; attention was strongly concentrated near the lung hila-an area of anatomical crowding which could easily be mistaken for pathology,

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

and that a clinician should examine carefully. Radiologists provided the overall clinical relevance of the Grad-CAM visualizations with an average score of 4.3/5.

6.5 Discussion

HAA-CNN is positioned interestingly within the space of systems presented. Although not the highest performing in raw accuracy (fine-tuned ResNet50 and SNN-GRU model both outperform slightly), it only trails ResNet50 by 1.2% accuracy, at 16% the parameter count and running on standard GPUs, rather than neuromorphic chips. These practical factors have huge significance for clinical deployment in areas with limited infrastructure such as district hospitals, community clinics, and mobile screening tools in low-resource regions.

Our SGAN augmentation strategy performs favourably compared to alternative data augmentation methods used in similar literature. Standard SMOTE is based on interpolation in feature space and samples lack any correlation to radiographic anatomy, while our pixel space augmentation generates samples rated as diagnostically plausible by experienced radiologists. Unlike Xu and Zhang's approach of using low-resolution 64x64 images that lose much radiologically important detail, our synthetic X-rays were generated at full, high-resolution (224x224 pixels).

Explicit, clinician evaluation of Grad-CAM visualizations is one component missing from many studies and is key to validating the interpretability of model findings. High classification accuracy on test data, while important, isn't necessarily sufficient. A model may achieve high accuracy due to spurious correlations that will lead to unpredictable behavior during deployment; demonstrably that a model highlights the correct radiologic features increases clinical trust significantly.

The main limitation of this work is the scope of training and testing data. The Kermany dataset consists of pediatric samples taken at one institution, meaning generalizability to the adult population, different imaging hardware, variable protocols and imaging institutions, is not guaranteed. Binary Normal vs. Pneumonia is a crude classification; classification into bacterial vs. Viral pneumonia, directly relevant for antibiotic choice, was beyond scope.

## VII. CONCLUSION

This work introduced HAA-CNN, a novel deep learning architecture for automatic pneumonia detection from chest x-rays aimed at addressing three common problems in the literature: handling class imbalance in training data, optimizing the trade-off between diagnostic performance and computational efficiency, and increasing model interpretability.

The developed system utilized a Spectral Normalization GAN for minority class augmentation, a lightweight depthwise separable convolutional backbone combined with spatial-channel attention for optimized processing of visual features, and Grad-CAM visualization validated by a group of radiologists. The Kermany paediatric dataset yielded 98.24% accuracy, 98.61% recall, 97.89% precision, an F1-score of 98.25%, an AUC of 0.994, and required approximately 3.8 million parameters. Ablation tests indicated that each component significantly improved performance. Grad-CAM was validated against radiologist-annotated outputs.

Demonstrates that the model attention corresponds to clinical definitions, not irrelevant image attributes. Collectively, these fiwndings show that HAA-CNN is a feasible candidate for AI assistance in pneumonia screening, particularly in low-resource settings where there is a need for both efficient processing and accountability in clinical practice.

## VIII. FUTURE WORK

Several future research directions stem from the presented work.

- Multi-centre and adult population validation: Performance should be validated on external datasets representing different institutions, patient populations, image devices, and clinical protocols, before generalising claims.
- Multi-class classification: The binary classification could be extended to discern bacterial, viral (including SARS-CoV-2), or fungal pneumonia, thus aiding in the selection of the appropriate treatment and in antibiotic stewardship.
- Continuous severity grading: Instead of a binary output, the system could produce a calibrated score of disease severity, thereby providing clinical support for disease staging and tracking the response to treatment.
- Diffusion-based generative augmentation: We could investigate using conditional denoising diffusion probabilistic models (DDPMs) instead of SGAN to generate higher-fidelity, more diverse radiographs.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

• Federated learning: We could explore federated learning to train on data from multiple hospitals without direct transfer of raw patient data; this would maintain data privacy and enhance model generalisation across sites.
• Prospective clinical validation: Integration of HAA-CNN with hospital radiology information systems followed by prospective evaluation in real clinical workflow will offer evidence of its clinical utility and will help to define the regulatory path forward.
• Edge deployment: Techniques like quantisation, pruning, and knowledge distillation may be employed to reduce model size and inference time to facilitate deployment on mobile devices and on low-power workstations in low-resource clinical environments.

## REFERENCES

[1] World Health Organization, "Pneumonia," WHO Fact Sheet, 2022. Available: https://www.who.int/news-room/fact-sheets/detail/pneumonia
[2] Centers for Disease Control and Prevention, "Pneumonia," 2021. Available: https://www.cdc.gov/pneumonia
[3] M. I. Neuman et al., "Variability in the interpretation of chest radiographs for the diagnosis of pneumonia in children," Journal of Hospital Medicine, vol. 7, no. 4, pp. 294-298, 2012. Https://doi.org/10.1002/jhm.955
[4] P. Rajpurkar et al., "CheXNet: Radiologist-level pneumonia detection on chest X-rays with deep learning," arXiv:1711.05225, 2017.
[5] H. Almohab, "Deep Learning CNN for Pneumonia Detection: Advancing Digital Health in Society 5.0," Jurnal Ilmiah Profesi Pendidikan, vol. 10, no. 4, pp. 3787-3793, 2025. Https://doi.org/10.29303/jipp.v10i4.4001
[6] S. K. Anumula et al., "Deep Learning-Based CNN Model for Automated Detection of Pneumonia from Chest X-Ray Images," unpublished manuscript, 2025.
[7] P. K. Dutta et al., "Deep Learning-Based Pneumonia Detection from Chest X-ray Images: A CNN Approach with Performance Analysis and Clinical Implications," arXiv:2510.00035, 2025. Https://doi.org/10.48550/arXiv.2510.00035
[8] S. R. Hole et al., "Chest X-ray Pneumonia Detection using Deep Learning," in Proc. ICBMESH, 2025. Https://doi.org/10.1109/ICBMESH66209.2025.11182215
[9] A. Roy Choudhury, "Pediatric Pneumonia Detection from Chest X-Rays: A Comparative Study of Transfer Learning and Custom CNNs," arXiv:2601.00837, 2025. Https://doi.org/10.48550/arXiv.2601.00837
[10] H. Slimi et al., "Trustworthy pneumonia detection in chest X-ray imaging through attention-guided deep learning," Scientific Reports, vol. 15, p. 40029, 2025. Https://doi.org/10.1038/s41598-025-23664-x
[11] Z. Xu and H. Zhang, "Pneumonia detection from enhanced chest X-Ray images based on Double SGAN model," Scientific Reports, vol. 16, p. 9922, 2026. Https://doi.org/10.1038/s41598-026-39785-w
[12] D. S. Kermany et al., "Identifying medical diagnoses and treatable diseases by image-based deep learning," Cell, vol. 172, no. 5, pp. 1122-1131, 2018. Https://doi.org/10.1016/j.cell.2018.02.010
[13] K. He et al., "Deep residual learning for image recognition," in Proc. IEEE CVPR, pp. 770-778, 2016.
[14] G. Huang et al., "Densely connected convolutional networks," in Proc. IEEE CVPR, pp. 4700-4708, 2017.
[15] M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for CNNs," in ICML, pp. 6105-6114, 2019.
[16] R. R. Selvaraju et al., "Grad-CAM: Visual explanations from deep networks via gradient-based localisation," in Proc. IEEE ICCV, pp. 618-626, 2017.
[17] T. Miyato et al., "Spectral normalisation for generative adversarial networks," arXiv:1802.05957, 2018.
[18] N. V. Chawla et al., "SMOTE: Synthetic minority over-sampling technique," JAIR, vol. 16, pp. 321-357, 2002.
[19] K. Yu et al., "Artificial intelligence in healthcare," Nature Biomedical Engineering, vol. 2, no. 10, pp. 719-731, 2018. Https://doi.org/10.1038/s41551-018-0305-z
[20] I. Goodfellow et al., "Generative adversarial networks," Communications of the ACM, vol. 63, pp. 139-144, 2020.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi crossref

निस्केयर NISCAIR

# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

## IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details