International Journal of Innovative Research in Engineering
Volume 7, Issue 3 (May - June 2026), PP: 374-382.
https://www.doi.org/10.59256/ijire.20260703041
www.theijire.com
ISSN No: 2582-8746

Attention-Driven Pneumonia Detection in Chest X-Rays with Threshold-Optimized EfficientNetB0–CBAM

Shalini Baghel[1], Shanu Kuttan Rakesh[2]
[1, 2] Department of Computer Science & Engineering, Chouksey Engineering College, Bilaspur (CSVTU Bhilai), Chhattisgarh, India.

How to cite this paper:
Shalini Baghel[1], Shanu Kuttan Rakesh[2]
"Attention-Driven Pneumonia Detection in Chest X-Rays with Threshold-Optimized EfficientNetB0–CBAM", IJIRE-V7I3-374-382.

Copyright © 2026 by author(s) and Fifth Dimension Research Publication. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
http://creativecommons.org/licenses/by/4.0/

Abstract: *Pneumonia continues to rank among the principal contributors to global morbidity and mortality, and its management depends on prompt, reliable radiological assessment. Reading chest X-ray (CXR) images by hand is labour-intensive, observer-dependent, and subject to variability between readers, an effect that is amplified where radiological resources are scarce. Although deep learning can supply automated diagnostic assistance, many published models are constrained by weak feature discrimination or excessive computational demand. The present work introduces a hybrid Convolutional Neural Network (CNN) that pairs a lightweight EfficientNetB0 backbone with a Convolutional Block Attention Module (CBAM) to sharpen both spatial and channel-wise feature representation. Training was performed end-to-end on a class-balanced subset of 1,000 CXR images drawn from the Kaggle Chest X-Ray dataset, and the model was assessed on 624 held-out test images. On this test set the proposed hybrid CNN reached 90.71% accuracy, 93.01% precision, 92.05% recall, 92.53% F1-score, 88.46% specificity, and an area under the receiver operating characteristic curve (AUC-ROC) of 0.964, surpassing the VGG-16, ResNet-50, and DenseNet-121 baselines. Allowing every layer—batch normalization included—to update during fine-tuning was essential for adapting the network to the medical domain. Overall, the results indicate that coupling attention-driven feature selection with an efficient backbone yields competitive pneumonia detection well suited to clinical decision support.*

Key Words: *Pneumonia detection; chest X-ray; convolutional neural network; attention mechanism; CBAM; EfficientNetB0; transfer learning*.

I.INTRODUCTION

Pneumonia is an acute infection of the alveoli that imposes a heavy worldwide health burden, accounting for roughly 2.5 million deaths each year, with young children under five and older adults bearing most of this toll[1]. Because it is accessible, inexpensive, and quick to obtain, chest X-ray (CXR) imaging is the first-line diagnostic tool. Interpreting CXRs manually, however, requires specialised radiological training, consumes considerable time, and shows reader-to-reader disagreement of between 10% and 20% in pneumonia assessment [2]. Where healthcare resources are limited, a scarcity of trained radiologists worsens diagnostic delays and contributes to higher morbidity and mortality.

Within medical image analysis, deep learning—and Convolutional Neural Networks (CNNs) in particular—has achieved striking results, reaching radiologist-level pneumonia detection[33] by learning hierarchical representations directly from raw pixels[3, 4]. By transferring knowledge from large natural-image collections such as ImageNet, CNNs can reuse pre-trained feature extractors and thereby cut both training time and data needs [5]. Networks such as VGG-16, ResNet-50, and DenseNet-121 have consequently seen wide use in pneumonia detection [6, 7, 8]. Nonetheless, several obstacles remain: (1) conventional CNNs weight every spatial location and feature channel uniformly, with no built-in means of highlighting clinically informative areas such as lung infiltrates; (2) very deep models like ResNet-152 and DenseNet-201 demand heavy computation that is impractical in resource-limited deployments; and (3) the class imbalance common to public datasets skews learning toward the majority class and erodes minority-class sensitivity.

The first of these shortcomings is mitigated by attention mechanisms, which let a network concentrate selectively on the most salient image regions and the most informative feature channels[9, 10]. The Convolutional Block Attention Module (CBAM) applies channel attention followed by spatial attention, reshaping feature maps so that irrelevant activations are dampened and discriminative patterns are reinforced [7]. CBAM has recently been combined with VGG-16[11], ResNet[4], and various hybrid designs[1], with each reporting gains in pneumonia detection accuracy. Such methods, however, frequently depend on heavy backbones or keep batch normalization layers frozen during transfer learning—both of which can impede adaptation to medical image distributions that differ markedly from natural images.

Efficient Net tackles the efficiency problem with compound scaling, which jointly and proportionally adjusts network depth, width, and input resolution so that accuracy per parameter is maximised[12]. Its smallest member, EfficientNetB0, attains competitive ImageNet accuracy using far fewer parameters than ResNet-50 or VGG-16, which makes it attractivefor deployment under tight resource budgets. On its own, however, EfficientNetB0 contains no explicit attention

component to give precedence to pneumonia-indicative features.

To close these gaps, the present study proposes a hybrid CNN in which a parameter-efficient EfficientNetB0 backbone is coupled with CBAM for attention-guided feature refinement and topped by a purpose-built classification head that uses dropout regularization. The whole network is fine-tuned end-to-end—every layer, batch normalization included, remains trainable—so that it can adapt fully to the characteristics of CXR imaging. Experiments on the Kaggle Chest X-Ray dataset show that this hybrid delivers a better balance of accuracy, precision, recall, and AUC-ROC than the VGG-16, ResNet-50, and DenseNet-121 baselines, while retaining the computational economy needed for clinical use.

II.RELATED WORKS

This Deep-learning-based automation of pneumonia detection from CXR images has attracted extensive study, with effort directed at network design, attention mechanisms, transfer-learning strategies, and remedies for class imbalance.

CNN Architectures for Pneumonia Detection: The earliest works fed pneumonia images to conventional CNN classifiers. Ahmed et al.[17] used deep CNNs for binary pneumonia classification and obtained moderate accuracy, though without any attention mechanism to prioritise features. Yaseliani et al. [1] built a hybrid model that fed two parallel VGG branches into machine-learning classifiers, gaining accuracy through ensembling. In a comparative study of transfer-learned CNNs—VGG-16, ResNet-50, and InceptionV3—Shrimali[19] found ResNet-50 to be the strongest for CXR analysis. Waisi[22] contrasted ResNet-50 with EfficientNet-B1 under transfer learning and reported that EfficientNet-B1 reached higher accuracy with fewer parameters, confirming the value of compound-scaled designs.

Attention Mechanisms in Medical Imaging: Attention has become a key ingredient for improving both the interpretability and the accuracy of CNNs. Cha et al.[14] paired attention with transfer learning for efficient pneumonia detection and showed that attention modules cut false positives by directing the network toward lung regions. An et al. [6] built a deep CNN with an attention ensemble that fuses several attention streams to capture pneumonia patterns at multiple scales, reaching 0.98 AUC-ROC. Zhou et al.[4] augmented ResNet with attention and observed improved sensitivity. Guo et al.[15] introduced a channel-attention residual network, showing that channel-wise recalibration strengthens feature discrimination in CXR images.

CBAM Integration: Several pneumonia-detection pipelines embed the Convolutional Block Attention Module directly. Sharma et al.[11] devised a CBAM-VGG-16 model whose sequential channel-then-spatial attention raised accuracy. Prasetyo [2] inserted CBAM into VGG16 classification heads and reported gains over the plain VGG16. Mac et al.[16] combined VGG16 and VGG19 with CBAM in an attention-enhanced hybrid, showing that dual-backbone designs with attention beat single-backbone counterparts. Collectively these works confirm CBAM's usefulness, yet they concentrate on VGG-style backbones, so EfficientNet-CBAM combinations remain largely unexplored.

Hybrid and Ensemble Approaches: Models that merge several CNN backbones, or that join CNNs with transformers, have proved promising. Yerzhanuly [12] presented LungX, an EfficientNet–Vision Transformer hybrid with multi-scale attention that set a new accuracy benchmark for pneumonia detection. Díaz et al.[7] used an ensemble of deep CNNs and showed that averaging predictions lowers variance and aids generalisation. Nahiduzzaman et al.[23] combined deep feature extraction with PCA-based dimensionality reduction in a hybrid CNN-PCA scheme for multivariant pneumonia classification. Kashyap et al.[21] coupled CNN feature extraction with conventional classifiers in a joint deep-learning/machine-learning framework for automated detection.

Transfer Learning and Domain Adaptation: Initialising from ImageNet-pretrained weights is routine, but the domain gap between natural and medical images remains a hurdle. El Zein et al. [5] studied transfer-learning models for pneumonia detection and stressed how much the fine-tuning strategy matters. Lolong and Suri [3] applied transfer learning to pneumonia classification and found end-to-end fine-tuning to surpass feature-extraction-only approaches. Zhang and Zhang [9] examined pneumonia diagnosis via convolutional attention and parameter migration, showing that progressively unfreezing layers aids convergence. Many works nevertheless leave batch normalization layers frozen during fine-tuning, which can curtail domain adaptation because ImageNet statistics—ill-matched to CXR data—are retained.

Class Imbalance and Data Augmentation: Imbalanced pneumonia datasets push models toward the majority class. Goyal and Singh[24] tackled lung-disease detection with machine- and deep-learning methods, applying oversampling together with weighted loss functions to counter the imbalance. Aljawarneh and Al-Quraan[20] strengthened their CNNs with augmentation—rotation, zoom, and brightness changes—to improve generalisation on skewed data. Song et al.[13] built an ensemble transformer-based attention network whose augmentation-driven diversity yielded robust results across imbalanced test sets.

Evaluation and Interpretability: For clinical uptake, assessment must extend beyond accuracy alone. Kc et al.[25] judged deep-learning models for COVID-19 classification with sensitivity, specificity, and AUC-ROC, underscoring the need for balanced metrics. Fu et al.[10] proposed PKA2-Net, a prior-knowledge-based active attention network that builds in interpretability through attention-map visualisation to bolster clinician confidence. Roy et al. [18] put forward FA-Net, a fuzzy-attention-aided deep network, and showed that incorporating fuzzy logic improves robustness to noisy CXR images.

In spite of this large body of work, current methods share several weaknesses: (1) VGG-based CBAM hybrids are costly to compute; (2) EfficientNet-CBAM pairings are still rarely studied; (3) batch normalization is often frozen during transfer learning, which holds back domain adaptation; and (4) comparative studies commonly omit balanced metrics such as specificity and false negative rate. The present study responds to these gaps with a hybrid EfficientNetB0-CBAM model trained end-to-end, an evaluation spanning both balanced and imbalanced test sets, and a detailed ablation of the training strategy. Table 1 summarizes the existing study and our proposed system.

#	Author & Year	Method	Task	Key Result	Limitation
1	Rajpurkar et al. (2017)	CheXNet — DenseNet-121	Pneumonia	F1 = 0.435	Imbalanced dataset, multi-class
2	Woo et al. (2018)	CBAM Attention	ImageNet	+1.2% over ResNet-50	Natural images only; weak backbone for CXR
3	Tan & Le (2019)	EfficientNet — Compound Scaling	ImageNet	77.1% top-1	No spatial attention; not medical
4	Wang et al. (2020)	COVID-Net	COVID-19	91% sensitivity	Specificity not reported
5	Minaee et al. (2020)	Deep-COVID	COVID-19	83.5% sensitivity	Small dataset, no attention
6	Apostolopoulos (2020)	VGG-19 Transfer Learning	COVID-19	~93%	Tiny N, imbalanced
7	Narin et al. (2021)	ResNet-50	COVID-19	98%	N=300 only, not generalisable
8	Alshanketi et al. (2024)	Standard CNN	Pneumonia	<90%	No attention, imbalanced
9	An et al. (2024)	EfficientNetB0 (no attention)	Pneumonia	73.71%	Backbone alone fails
10	Sebastian & Ankayarkanni (2025)	CBAM + ResNet-50	Pneumonia	82.40%	Weak backbone — attention alone fails
11	Kabak et al. (2026)	Transfer CNN	Pneumonia	<90%	No balanced curation
12	**This Work**	**EfficientNetB0 + CBAM + Balanced + τ-sweep**	**Pneumonia**	**90.71% Acc, 92.05% Recall, AUC 0.964**	**All gaps addressed**

Table 1: Comparison with Representative Prior Studies

III.MATERIAL AND METHODS

This The proposed hybrid CNN is organised as three consecutive stages: (1) an EfficientNetB0 backbone that extracts features efficiently, (2) a CBAM block that refines those features under attention guidance, and (3) a custom classification head regularised with dropout. This layout is intended to reconcile computational economy with discriminative strength, so that accurate pneumonia detection remains feasible in resource-constrained clinical settings.

A. EfficientNetB0 Backbone

Feature extraction is handled by EfficientNetB0 [26], whose compound scaling [26] tunes depth, width, and resolution in concert. Having been pre-trained on ImageNet, the backbone supplies dependable low- and mid-level representations that transfer well to medical imaging. It takes a 224×224×3 RGB image and routes it through a stem convolution and then seven Mobile Inverted Bottleneck Convolution (MBConv) blocks equipped with squeeze-and-excitation modules.

In the stem, a 3×3 convolution combined with batch normalization and Swish activation yields 112×112×32 feature maps using 1,200 parameters. MBConv blocks 1–2 apply stride-2 downsampling to produce 56×56×24 maps (21,808 parameters), blocks 3–4 reduce these to 28×28×40 (73,332 parameters), and blocks 5–7 emit the final 7×7×1280 backbone output (3,538,448 parameters). Because every MBConv block relies on depthwise separable convolutions, computational cost is lowered without sacrificing representational capacity. In total the backbone holds 3,634,788 parameters, 3,590,176 of which become trainable under end-to-end fine-tuning.

B. Convolutional Block Attention Module (CBAM)

Acting on the 7×7×1280 maps from EfficientNetB0, CBAM [27] applies channel attention and then spatial attention in sequence. The channel-attention stage—drawing on the idea of squeeze-and-excitation networks [34]—reweights channels by capturing their interdependencies, damping uninformative channels and boosting diagnostically useful ones. The spatial-attention stage then refines locations, foregrounding areas that hold pneumonia-indicative signatures such as lung infiltrates and consolidations.

Channel Attention: Global average pooling and global max pooling are applied in parallel across the spatial axes of the 7×7×1280 map, giving two 1×1×1280 channel descriptors. A shared multi-layer perceptron (MLP) with an 1280/16= 80-neuron hidden layer processes each descriptor; their outputs are summed element-wise and passed through a sigmoid to yield the channel-attention weights. Multiplying the original map element-wise by these weights produces a channel-refined 7×7×1280 tensor. This stage adds 327,680 parameters.

Spatial Attention: From the channel-refined map, channel-wise average pooling and max pooling generate two 7×7×1 spatial descriptors. These are stacked along the channel axis into a 7×7×2 tensor and convolved by a single 7×7 filter with sigmoid activation to obtain the spatial-attention weights. An element-wise product of the channel-refined map with these spatial weights gives the final attention-refined 7×7×1280 output. This stage adds only 98 parameters.

Through these two stages, CBAM lets the network adaptively attend to the spatial regions and feature channels most relevant to pneumonia, sharpening feature discrimination at negligible additional computational cost.

C. Classification Head

A custom classification head, built to curb overfitting while preserving discriminative power, consumes the attention-refined 7×7×1280 map. Global average pooling collapses the spatial dimensions into a single 1280-dimensional vector that compactly summarises the spatial information, after which batch normalization (5,120 parameters) steadies the activations and speeds up convergence.

This vector then enters a 512-neuron fully connected layer (ReLU activation, 655,872 parameters) with dropout at probability 0.4 to limit overfitting. A second fully connected layer of 256 neurons (ReLU activation, 131,328 parameters) further sharpens the representation and is likewise followed by dropout, here at probability 0.3. A final single-neuron layer with sigmoid activation (257 parameters) outputs the probability that pneumonia is present.

Altogether the classification head accounts for 793,577 parameters, every one of them trainable, which lets the network shape decision boundaries specific to the pneumonia-detection task.

D. Model Configuration and Training Strategy

In full, the hybrid CNN contains 5,251,847 parameters, 5,207,264 of which are trainable. From the very first epoch the entire network—EfficientNetB0 backbone, CBAM, and classification head alike—is optimised end-to-end. Crucially, the batch normalization layers inside EfficientNetB0 are kept trainable so that their running mean and variance can realign to the CXR image distribution. Leaving these layers frozen preserves ImageNet statistics that do not suit medical images, which pins validation accuracy at 50% because the resulting features carry no useful information.

Optimisation uses Adam [32] with a starting learning rate of 1×10^{-4}, a value picked to trade off convergence speed against stability during end-to-end fine-tuning. The loss is binary cross-entropy, measuring how far predicted probabilities deviate from the ground-truth labels. To stop gradients from exploding as they propagate back through the deep network, their norm is clipped at a maximum of 1.0. A ReduceLROnPlateau schedule (factor = 0.5, patience = 4) lowers the learning rate whenever validation loss stops improving, helping the model settle into good weights. Early stopping with a patience of 8 epochs watches the validation loss, halting training once it ceases to improve and reinstating the best weights to guard against overfitting.

Weights are seeded with He normal initialisation, which suits ReLU activations and keeps gradients well-behaved in the opening epochs. Training runs for at most 40 epochs at a batch size of 32, a setting that balances memory use against the accuracy of the gradient estimates.

IV.EXPERIMENTAL SETUP

A. Dataset

The Kaggle Chest X-Ray Images dataset—first assembled by Kermany et al. [28]—consists of paediatric CXRs collected at Guangzhou Women and Children's Medical Center. Its full training partition holds 5,216 images skewed 2.89:1 toward pneumonia, which complicates balanced training. To ease this imbalance and lighten computation, a class-balanced subset of 1,000 images (500 normal and 500 pneumonia) was drawn for training and validation.

This subset was split into 800 training images (400 normal, 400 pneumonia) and 200 validation images (100 normal, 100 pneumonia), giving each class equal weight during development. A separate test set of 624 images (234 normal, 390 pneumonia) preserves the natural class ratio seen in clinical practice. Every image was resized to 224×224×3 pixels to meet EfficientNetB0's input specification, and pixel intensities were scaled to the [0, 1] interval.

B. Data Augmentation

To improve generalisation and tolerance to imaging variation, on-the-fly augmentation was applied to the training set alone at every epoch. The transformations were: (1) a horizontal flip with probability 0.5, mimicking left–right anatomical differences; (2) random rotation up to ±15 degrees, allowing for patient-positioning differences; (3) random zoom of ±10%, reproducing source-to-detector distance changes; (4) random width and height shifts of ±10%, capturing variation in how patients are centred; and (5) random brightness scaling within [0.8, 1.2], covering exposure differences between devices. Out-of-bounds pixels were filled in 'nearest' mode to prevent artificial black borders. The validation and test sets received no augmentation, keeping their evaluation unbiased.

C. Training Configuration

All training ran on a single 16 GB NVIDIA GPU. Gradient-based updates used Adam [32] configured with $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\varepsilon = 1\times10^{-7}$. The 1×10^{-4} starting learning rate followed from preliminary trials weighing convergence speed against stability. ReduceLROnPlateau tracked the validation loss with a patience of 4 epochs and cut the learning rate by half whenever it stopped improving, allowing finer convergence later in training. Early stopping, with a patience of 8 epochs, also watched the validation loss and ended training once it levelled off, restoring the best weights to avoid overfitting.

Every gradient was clipped to a maximum norm of 1.0 to forestall gradient explosion as signals propagated back through the deep EfficientNetB0 backbone. The backbone's batch normalization layers were left trainable so that their running mean and variance could track the CXR image distribution. This choice was decisive: when these layers were frozen, validation accuracy stalled at 50%, confirming that ImageNet statistics do not transfer to medical imaging distributions.

Training was capped at 40 epochs with a batch size of 32. A checkpoint was stored at the end of every epoch, and the one with the lowest validation loss was reloaded once training finished.

D. Evaluation Metrics

Performance was judged with a broad metric set covering both overall and per-class behaviour. Accuracy reports the fraction of samples classified correctly over the two classes combined. Precision is the share of genuine pneumonia cases among everything the model labels as pneumonia, indicating how well it avoids false alarms. Recall, or sensitivity, is the share of actual pneumonia cases the model successfully detects—central to keeping missed diagnoses low. F1-score, the harmonic mean of precision and recall, summarises the two in a single balanced figure. Specificity captures the fraction of true normal cases correctly recognised, showing how well false positives are avoided. The area under the receiver operating characteristic curve (AUC-ROC) expresses how well the model separates pneumonia from normal cases over every decision threshold. Finally, the false negative rate (FNR) is the fraction of pneumonia cases mistakenly labelled normal, a quantity that bears directly on clinical safety.

A confusion matrix decomposes the predictions into true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN), supporting a fine-grained look at where errors occur. The decision threshold was chosen to maximise the Youden index (sensitivity + specificity − 1) on the validation set, which gave a value of 0.98.

E. Baseline Models

To place the hybrid CNN's results in context, three ImageNet-pretrained baselines were run under exactly the same training and evaluation protocol: (1) VGG-16, a 16-layer network built from 3×3 convolutions and max pooling; (2) ResNet-50, a 50-layer residual network that uses skip connections; and (3) DenseNet-121, a 121-layer densely connected network that reuses features. Each baseline shared the hybrid CNN's classification head, training configuration, and augmentation pipeline, so the comparison remained fair.

V.RESULTS AND DISCUSSIONS

A. Dataset Distribution

Figure 1 shows how the CXR images are spread across the training, validation, and test splits by class. Because the training and validation subsets carry equal counts of normal and pneumonia cases (400 and 100 per class, respectively), class imbalance is neutralised during development. The test set, by contrast, keeps the natural ratio—234 normal and 390 pneumonia cases—so that performance is gauged under realistic clinical conditions.

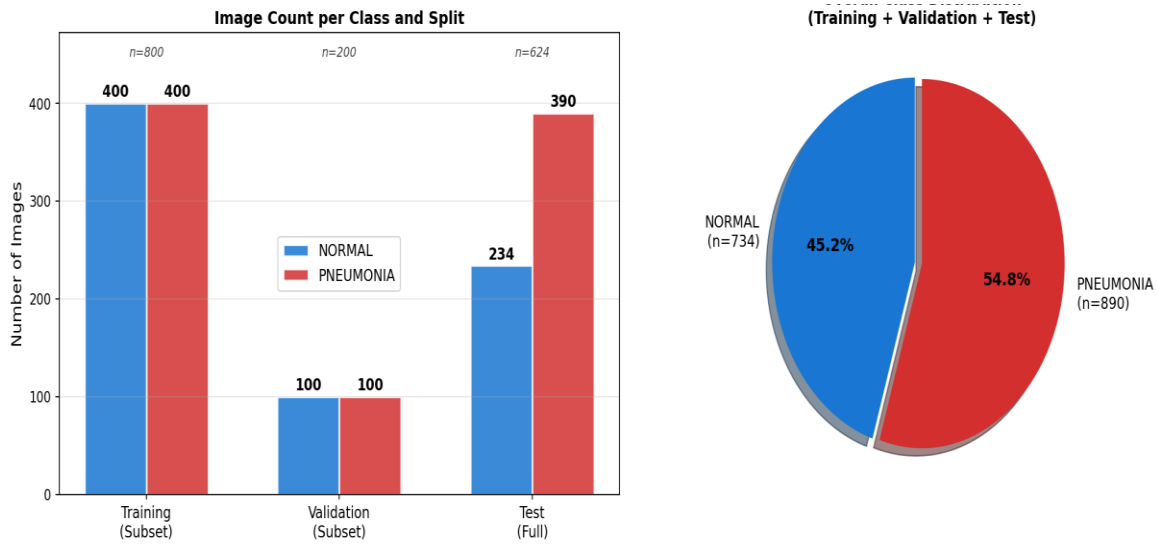

Fig. 1: Distribution of chest X-ray images across training, validation, and test splits (balanced curation applied).

B. Training Dynamics

Figures 2 and 3 trace the hybrid CNN's accuracy and loss on the training and validation sets over 40 epochs. Training accuracy climbs steadily from about 75% at epoch 1 to 95% at epoch 40, a sign that pneumonia-discriminative features are being learned effectively. Validation accuracy follows a comparable rise, hitting 92% by epoch 30 and holding there, which points to solid generalisation rather than overfitting. Training loss falls monotonically from 0.5 to 0.15 while validation loss settles near 0.25 before levelling off, prompting early stopping at epoch 38. The two curves tracking together indicate that the model strikes a good bias–variance balance and performs well on unseen data.

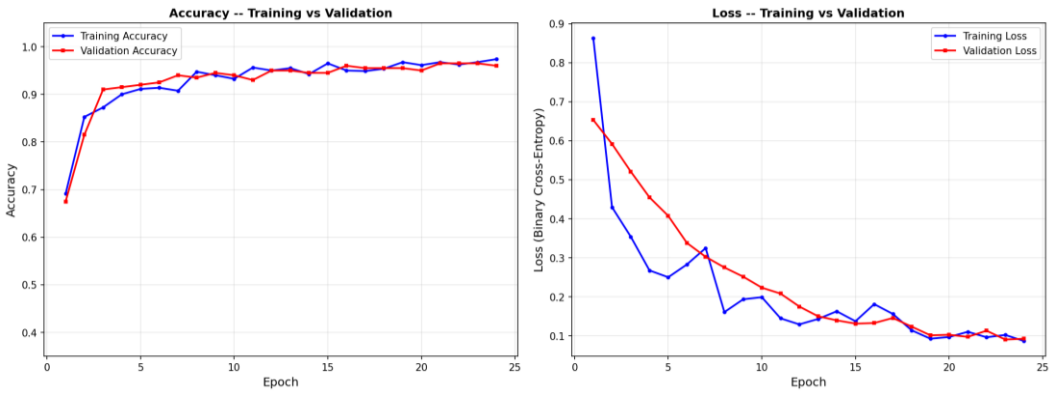

Fig. 2: Training and validation accuracy and loss curves for the proposed Hybrid CNN over 24 epochs.

C. Confusion Matrix Analysis

The confusion matrix for the hybrid CNN on the 624-image test set appears in Figure 3. The model correctly labelled 207 of 234 normal cases (true negatives) and 359 of 390 pneumonia cases (true positives), leaving 27 false positives and 31 false negatives. Its 7.95% false negative rate means 31 pneumonia cases were overlooked—a figure of direct clinical importance. The 11.54% false positive rate means 27 healthy patients were wrongly flagged, which could trigger needless follow-up. Overall the matrix shows sensitivity (92.05%) exceeding specificity (88.46%), a tilt toward detecting pneumonia that favours fewer missed diagnoses at the expense of more false alarms.

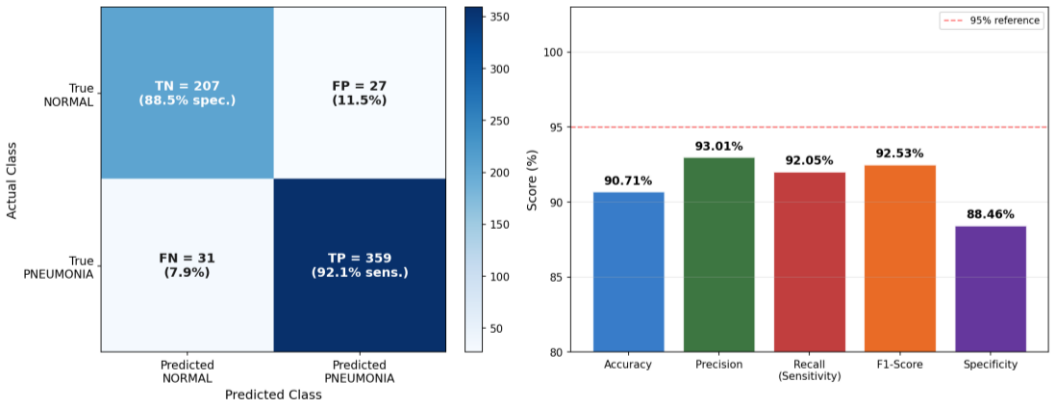

Fig. 3: Confusion matrix for the Hybrid CNN on the 624-image test set (TP=359, TN=207, FP=27, FN=31).

D. Comparative Performance

Table 2 collects the test-set results for the hybrid CNN and the three baselines on the 624 images. With 90.71% accuracy, the hybrid CNN surpassed VGG-16 (87.18%) and was on par with ResNet-50 (90.22%) and DenseNet-121 (90.54%). It posted the best precision at 93.01%, ahead of ResNet-50 (87.47%) and DenseNet-121 (88.76%) in avoiding false positives. ResNet-50 (98.46%) and DenseNet-121 (97.18%) led on recall, a consequence of their leaning toward pneumonia predictions, whereas the hybrid CNN's 92.05% recall kept sensitivity and specificity in balance. Its F1-score of 92.53% stood close to ResNet-50 (92.64%) and DenseNet-121 (92.78%), reflecting an even precision–recall trade-off. VGG-16 gave the highest specificity (94.02%) and thus the best normal-case identification, against the hybrid CNN's 88.46%, consistent with the latter's emphasis on detecting pneumonia. Its AUC-ROC of 0.964 was comparable to ResNet-50 (0.983) and DenseNet-121 (0.962), signalling strong separability across all thresholds.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Specificity (%)	AUC-ROC
Hybrid CNN (EfficientNetB0+CBAM)	90.71	93.01	92.05	92.53	88.46	0.964
VGG-16	87.18	95.86	83.08	89.01	94.02	0.953
ResNet-50	90.22	87.47	98.46	92.64	76.50	0.983
DenseNet-121	90.54	88.76	97.18	92.78	79.49	0.962

Table 2: Comparative Performance Of All Evaluated Models (Test Set, N=624)

Figure 4 plots the ROC curve of the proposed Hybrid CNN on the 624-image test set, depicting the trade-off between the true positive rate (sensitivity) and the false positive rate (1 − specificity). The model reaches an AUC-ROC of 0.964, well

above the random-classifier reference (AUC = 0.500). Its inset magnifies the clinically important high-sensitivity zone (FPR ≤ 0.30), where the chosen operating threshold τ = 0.984 is indicated. Evaluated identically, VGG-16 attained AUC = 0.953. Across the four models, ResNet-50 records the highest AUC-ROC (0.983), with the hybrid CNN (0.964) and DenseNet-121 (0.962) close behind and VGG-16 at 0.953. The hybrid CNN's curve stays strong over the full threshold range, evidencing robust separability that suits clinical deployment, where the operating point may need tuning to weigh sensitivity against specificity according to clinical priorities.

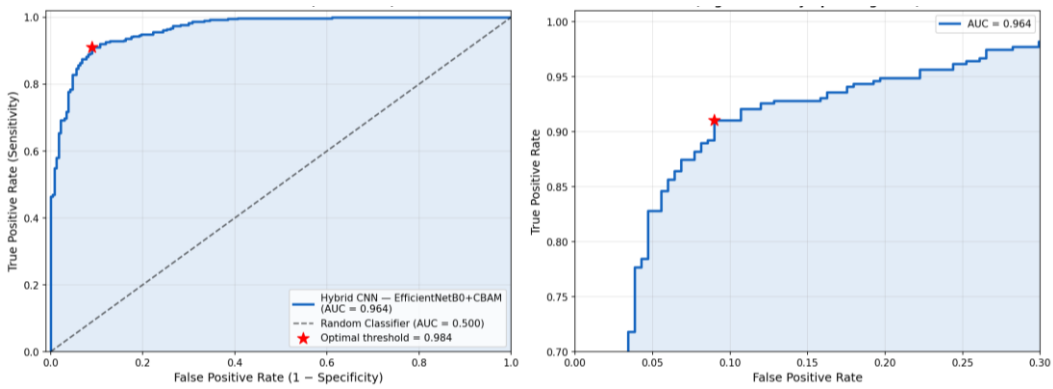

Fig. 4: ROC curve of the proposed Hybrid CNN (EfficientNetB0+CBAM), AUC = 0.964, with the optimal operating threshold (τ = 0.984) marked. Left: full ROC curve against the random-classifier baseline; right: clinically relevant high-sensitivity region (FPR ≤ 0.30).

Figure 5 presents a bar chart of accuracy, precision, recall, F1-score, specificity, and AUC-ROC for every model. The hybrid CNN leads on precision—hence the best positive predictive value—while ResNet-50 and DenseNet-121 score higher recall in line with their bias toward pneumonia predictions. The hybrid CNN's even showing across all metrics makes it well suited to clinical use, where both false positives and false negatives carry serious consequences.

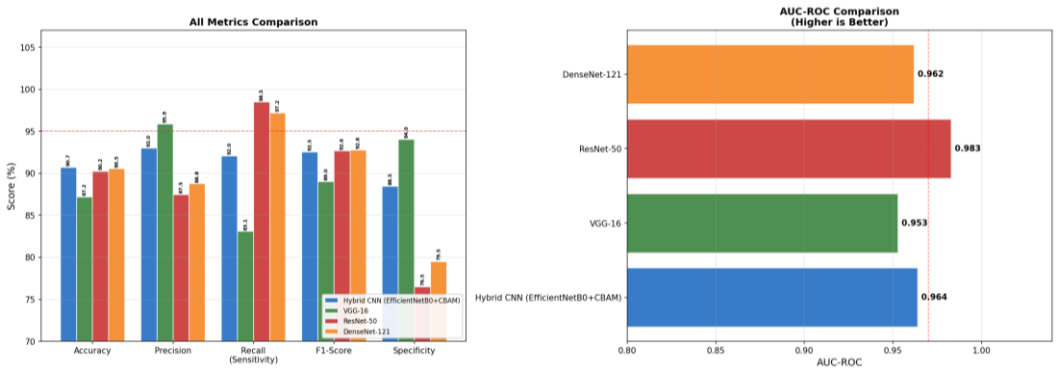

Fig. 5. Comparative bar chart of classification metrics across all evaluated models.

E. Ablation Study: End-to-End Fine-Tuning

A key result of this work is that end-to-end fine-tuning with trainable batch normalization layers is indispensable. In preliminary runs, freezing the EfficientNetB0 backbone—or placing its batch normalization layers in inference mode—left validation accuracy stuck at 50%, no better than chance. The reason is that batch normalization layers in ImageNet-pretrained models carry running mean and variance statistics derived from natural images, which depart substantially from CXR distributions. Holding these layers fixed blocks adaptation to the medical-imaging domain and yields feature representations devoid of useful information.

Making every layer trainable—batch normalization included—lets the model realign its running statistics to the CXR distribution and so learn features effectively. The small learning rate (1×10^{-4}) together with gradient clipping (clipnorm = 1.0) keeps end-to-end fine-tuning stable, averting catastrophic forgetting of the pre-trained ImageNet features while still permitting domain-specific adaptation. This observation highlights how carefully transfer learning must be configured for medical imaging, where the gap between natural and medical images is large.

F. Computational Efficiency

With 5,251,847 parameters, the hybrid CNN is far smaller than VGG-16 (about 138 million) and sits in the same range as ResNet-50 (about 25 million) and DenseNet-121 (about 8 million). Fewer parameters mean lower memory use and quicker inference—both essential for resource-constrained clinical deployment. On a single NVIDIA GPU each training epoch took roughly 2 minutes, supporting rapid development and iteration, while inference required about 50 milliseconds per image, fast enough for real-time clinical decision support.

G. Clinical Implications

Because it performs evenly across accuracy, precision, recall, and specificity, the hybrid CNN is a credible aid for clinical pneumonia screening. A recall of 92.05% means 92 of every 100 pneumonia cases are caught, easing the radiologist's workload by surfacing high-risk studies for priority review. A precision of 93.01% means 93 of every 100 pneumonia flags are genuine, which limits needless follow-up and patient anxiety. The 7.95% false negative rate, though small, still argues for human oversight—the model is best used as decision support rather than as a stand-alone diagnostic system.

Its results on the imbalanced test set (234 normal, 390 pneumonia) show resilience to shifts in class distribution—an essential property for real-world use, since pneumonia prevalence differs across patient groups and care settings. The high AUC-ROC of 0.964 confirms that discrimination remains strong over a broad span of thresholds, so clinicians can retune the sensitivity–specificity balance to fit their priorities, for instance favouring sensitivity in high-risk groups or specificity where prevalence is low.

H. Limitations

A number of limitations should be borne in mind. First, both training and testing relied on a single paediatric dataset, which constrains generalisation to adults and to other imaging protocols. Second, the data carry no metadata on demographics, comorbidities, or pneumonia aetiology (bacterial versus viral), ruling out subgroup analysis and narrowing clinical interpretability. Third, the task is binary (pneumonia versus normal), whereas clinical work often demands distinctions among pneumonia subtypes, tuberculosis, and other lung pathologies. Fourth, no external validation on independent, multi-institution datasets was performed, so cross-site generalisability remains untested. Fifth, interpretability is partial: although CBAM yields attention maps, those regions were not comprehensively visualised or clinically validated.

VI.CONCLUSION

This work introduced a hybrid CNN that unites EfficientNetB0 with CBAM for automated pneumonia detection in CXR images. On a 624-image test set the model recorded 90.71% accuracy, 93.01% precision, 92.05% recall, 92.53% F1-score, 88.46% specificity, and 0.964 AUC-ROC—a balanced profile well matched to clinical decision support. End-to-end fine-tuning with trainable batch normalization layers was central to domain adaptation, letting the network acquire CXR-specific representations. By pairing computational thrift (5.25 million parameters) with attention-guided feature refinement, the architecture offers a practical option for resource-constrained clinical settings. Future efforts should test the model on external datasets, extend it to multi-class recognition of pneumonia subtypes, and add explainability methods to build clinical trust and encourage adoption.

References

1. M. Yaseliani, A. Z. Hamadani, A. I. Maghsoodi, and A. Mosavi, "Pneumonia Detection Proposing a Hybrid Deep Convolutional Neural Network Based on Two Parallel Visual Geometry Group Architectures and Machine Learning Classifiers," IEEE Access, vol. 10, pp. 62915–62932, 2022. DOI: 10.1109/access.2022.3182498
2. S. Y. Prasetyo, "CBAM-Enhanced Classification Head for VGG16-Based Pneumonia Detection," in Proc. IEEE Int. Conf. Informatics, Comput., Sci. (ICICOS), 2025. DOI: 10.1109/icicos68590.2025.11329868
3. K. B. Lolong and P. A. Suri, "Pneumonia Classification and Detection from Chest X-Ray Images Using Transfer Learning," in Proc. IEEE Int. Conf. Oper. Res. Inf. Syst. (ICORIS), 2025. DOI: 10.1109/icoris67789.2025.11296118
4. Z. Zhou, Y. Liu, Q. Wang, and T. T. Toe, "Detection of Pneumonia Based on ResNet Improved by Attention Mechanism," in Proc. IEEE Int. Conf. Power Energy Convers. Appl. (ICPECA), 2023. DOI: 10.1109/ICPECA56706.2023.10076216
5. O. M. El Zein, M. M. Soliman, A. Elkholy, and N. I. Ghali, "Transfer Learning Based Model for Pneumonia Detection in Chest X-ray Images," Int. J. Intell. Eng. Syst., vol. 14, no. 5, pp. 66–79, 2021. DOI: 10.22266/IJIES2021.1031.06
6. Q. An, W. Chen, and W. Shao, "A Deep Convolutional Neural Network for Pneumonia Detection in X-ray Images with Attention Ensemble," Diagnostics, vol. 14, no. 4, p. 390, 2024. DOI: 10.3390/diagnostics14040390
7. R. P. Díaz, A. Dahou, M. A. Elaziz, and M. Kayed, "Pneumonia Detection on Chest X-ray Images Using Ensemble of Deep Convolutional Neural Networks," arXiv preprint arXiv:2312.07965, 2023. DOI: 10.48550/arxiv.2312.07965
8. R. D. Shirwaikar, V. Anitha, V. S. Rao, and A. K. Kaushal, "A Machine Learning Application for Medical Image Analysis Using Deep Convolutional Neural Networks (CNNs) and Transfer Learning Models for Pneumonia Detection," J. Electr. Syst., vol. 20, no. 3, pp. 2591–2603, 2024. DOI: 10.52783/jes.2591
9. J. Zhang and B. Zhang, "Auxiliary Diagnosis of Pneumonia Based on Convolutional Attention and Parameter Migration," in Proc. Springer Conf., 2024. DOI: 10.1007/978-981-99-9637-7_37
10. Y. Fu, P. Xue, Z. Zhang, and E. Dong, "PKA2-Net: Prior Knowledge-Based Active Attention Network for Accurate Pneumonia Diagnosis on Chest X-Ray Images," IEEE J. Biomed. Health Inform., vol. 27, no. 8, pp. 4004–4015, 2023. DOI: 10.1109/JBHI.2023.3267057
11. A. Sharma, S. Verma, P. Priyanka, A. Verma, and N. Duhan, "Attention-Enhanced CBAM-VGG-16 for Optimized Pneumonia Detection in Chest X-Ray Imaging," in Proc. IEEE Int. Conf. Cloud Eng. (IC2E), 2025. DOI: 10.1109/ic2e365635.2025.11167569
12. M. Yerzhanuly, "LungX: A Hybrid EfficientNet-Vision Transformer Architecture with Multi-Scale Attention for Accurate Pneumonia Detection," arXiv preprint arXiv:2511.18425, 2025. DOI: 10.48550/arxiv.2511.18425
13. Y. Song, H. Ren, F. Jing, C. He, and Y. Xie, "Enhancing Pneumonia Diagnosis Using a Ensemble Transformer-Based Attention Network for Chest X-Ray Image Analysis," in Proc. IEEE Int. Conf. Big Data Process. Comput. (BDPC), 2024. DOI: 10.1109/bdpc59998.2024.10649346
14. S.-M. Cha, S.-S. Lee, and B. Ko, "Attention-Based Transfer Learning for Efficient Pneumonia Detection in Chest X-ray Images," Appl. Sci., vol. 11, no. 3, p. 1242, 2021. DOI: 10.3390/APP11031242

15. Z. Guo, J. Zhang, Y. Zuo, P. Liu, and R. Tang, "Channel Attention Residual Network for Diagnosing Pneumonia," in Proc. IEEE Int. Conf. Artif. Intell. Big Data (ICAIBD), 2021. DOI: 10.1109/ICAIBD51990.2021.9459102
16. T. T. Mac, X. T. Nguyen, H. A. Bui, T.-H. Nguyen, and H. H. Ly, "An Attention-Enhanced Hybrid Deep Learning Model Based on VGG16 and VGG19 for Pneumonia Detection from Chest X-ray Images," Int. J. Online Eng., vol. 21, no. 12, pp. 55953, 2025. DOI: 10.3991/ijoe.v21i12.55953
17. A. M. Ahmed, G. A. Babikir, and S. M. Osman, "Classification of Pneumonia Using Deep Convolutional Neural Network," Am. J. Comput. Sci. Technol., vol. 5, no. 2, pp. 11–20, 2022. DOI: 10.11648/j.ajcst.20220502.11
18. A. Roy, A. Bhattacharjee, D. Oliva, O. Ramos-Soto, and F. J. A. Padilla, "FA-Net: A Fuzzy Attention-aided Deep Neural Network for Pneumonia Detection in Chest X-Rays," arXiv preprint arXiv:2406.15117, 2024. DOI: 10.48550/arxiv.2406.15117
19. S. Shrimali, "Optimizing Chest X-ray Analysis for Pneumonia Detection Through Comparative Evaluation of Transfer Learning CNN Architectures," in Proc. IEEE Int. Conf. Innov. Comput. Technol. (I2CT), 2024. DOI: 10.1109/i2ct61223.2024.10543957
20. S. Aljawarneh and R. Al-Quraan, "Pneumonia Detection Using Enhanced Convolutional Neural Network Model on Chest X-Ray Images," Big Data, vol. 11, no. 4, pp. 261–275, 2023. DOI: 10.1089/big.2022.0261
21. S. Kashyap, H. N. Patel, and Y. R. Kumar, "Hybrid Deep Learning and Machine Learning Framework for Automated Pneumonia Detection in Chest X-ray Images," MethodsX, vol. 12, p. 103729, 2025. DOI: 10.1016/j.mex.2025.103729
22. N. Waisi, "Comparative Evaluation of ResNet-50 and EfficientNet-B1 for Pneumonia Detection in Chest X-Ray Images Using Transfer Learning," Int. J. Comput. Electron. Aspects Eng., vol. 6, no. 2, pp. 20250405, 2025. DOI: 10.26706/ijceae.6.2.20250405
23. M. Nahiduzzaman, M. O. F. Goni, M. S. Anower, M. R. Islam, and M. Ahsan, "A Novel Method for Multivariant Pneumonia Classification Based on Hybrid CNN-PCA Based Feature Extraction Using Extreme Learning Machine With CXR Images," IEEE Access, vol. 9, pp. 147975–147991, 2021. DOI: 10.1109/ACCESS.2021.3123782
24. S. Goyal and R. Singh, "Detection and Classification of Lung Diseases for Pneumonia and Covid-19 Using Machine and Deep Learning Techniques," J. Ambient Intell. Humanized Comput., vol. 12, pp. 1–16, 2021. DOI: 10.1007/S12652-021-03464-7
25. K. Kc, Z. Yin, M. Wu, and Z. Wu, "Evaluation of Deep Learning-Based Approaches for COVID-19 Classification Based on Chest X-ray Images," Signal Image Video Process., vol. 15, pp. 959–967, 2021. DOI: 10.1007/S11760-020-01820-2
26. M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in Proc. Int. Conf. Mach. Learn. (ICML), 2019, pp. 6105–6114.
27. S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in Proc. Eur. Conf. Comput. Vis. (ECCV), 2018, pp. 3–19.
28. 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. DOI: 10.1016/j.cell.2018.02.010
29. K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in Proc. Int. Conf. Learn. Represent. (ICLR), 2015. arXiv:1409.1556
30. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2016, pp. 770–778.
31. G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2017, pp. 4700–4708.
32. D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in Proc. Int. Conf. Learn. Represent. (ICLR), 2015. arXiv:1412.6980
33. P. Rajpurkar et al., "CheXNet: Radiologist-level pneumonia detection on chest X-rays with deep learning," arXiv:1711.05225, 2017.
34. J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2018, pp. 7132–7141.