RESEARCH PAPER

DenseNet121-Based Intelligent Data Augmentation with Focal Loss and Threshold Optimization for Chest X-Ray Pneumonia Classification

[1]Suraj Gajanan Date, [2]Girija G. Chiddarwar, [3]Smita Chaudhari, [4]Swati Shekapure

[1,2,3,4]Department of Computer Engineering, Marathwada Mitra Mandal's College of Engineering, Pune-411052, India

Email Id: surajdate8@gmail.com[1], girijachiddarwar@mmcoe.edu.in[2], smita.m.c@gmail.com[3] Swatishekapure@gmail.com[4]

Abstract

Chest X-ray pneumonia classification is influenced by class imbalance, limited sample diversity and fixed decision thresholds. This paper presents a DenseNet121-based intelligent augmentation framework for binary chest X-ray pneumonia classification. The method integrates online image augmentation, transfer learning, class weighting, binary focal loss and validation macro-F1 threshold optimization. Experiments were performed on the public Chest X-Ray Images (Pneumonia) dataset using the official test split. Four configurations were evaluated: baseline CNN, traditional augmentation CNN, adaptive augmentation CNN and the proposed DenseNet121 model. The final model selected a threshold of 0.37 and achieved 87.82% accuracy, 93.85% pneumonia recall, 90.59% pneumonia F1-score, 86.66% macro F1-score and 85.81% balanced accuracy. The results show that transfer learning combined with imbalance-aware optimization improves class-balanced performance compared with custom CNN baselines.

Keywords: *chest X-ray, pneumonia detection, DenseNet121, augmentation, focal loss, threshold optimization, class imbalance*

How to cite this article: Date SG, Chiddarwar GG, Chaudhari S, Shekapure S. DenseNet121-Based Intelligent Data Augmentation with Focal Loss and Threshold Optimization for Chest X-Ray Pneumonia Classification. Int J Drug Deliv Technol. 2026;16(55s): 104-109. DOI: 10.25258/ijddt.16.55s.11

1. Introduction

Pneumonia is a respiratory infection that can be examined through chest radiography[1]. Automated interpretation of chest X-ray images using deep learning can support screening and reduce diagnostic workload. However, performance depends strongly on training data diversity, class distribution and the evaluation protocol[2].

Medical datasets frequently contain class imbalance[3]. A classifier may achieve acceptable accuracy while failing to recognize one class adequately[4]. Therefore, accuracy must be supported by precision, recall, F1-score, macro F1-score, balanced accuracy and confusion-matrix analysis. This requirement is particularly important in pneumonia screening, where sensitivity to pneumonia and reliable recognition of normal cases are both required.

This paper proposes a two-stage implementation framework. First, custom CNN baselines are trained to evaluate the effect of augmentation and adaptive augmentation. Second, a stronger DenseNet121 transfer-learning model is trained with class weights, focal loss and threshold tuning. The experimental analysis is designed to identify the contribution of each modelling decision.

2. Related Work

Recent augmentation research has moved from fixed transformations to automated, generative and task-aware augmentation. Yang et al. categorized image manipulation, erasing, mixing and generative augmentation [5]. Xu et al. organized augmentation into model-free, model-based and policy-optimization families [7]. Yang et al. reviewed automated data augmentation and highlighted the computational cost of search-based strategies [8].

Medical image augmentation has also been studied using GANs, VAEs and diffusion models. Kebaili et al. emphasized that generated medical images must be validated because unrealistic samples can harm diagnostic modelling. Dablain and Chawla showed that data augmentation under class imbalance must be evaluated with balanced metrics [10]. These studies

motivate a practical framework that combines augmentation with imbalance-aware training.

Table 1. Related work summary.

Ref.	Focus	Relevance to this work
Yang et al. 2022	Image augmentation survey	Supports augmentation taxonomy
Xu et al. 2023	Augmentation techniques	Shows task-aware selection need
Yang et al. 2023	Automated augmentation	Highlights search cost
Kebaili et al. 2023 [9]	Medical augmentation	Motivates cautious medical augmentation
Dablain and Chawla 2023 [10]	Imbalanced augmentation	Supports balanced metrics
Jang et al. 2024 [11]	CXR transfer learning	Supports pretrained feature extraction

3. Dataset and Preprocessing

The Chest X-Ray Images (Pneumonia) dataset contains radiographs arranged into NORMAL and PNEUMONIA classes. The official test set contains 624 images, including 234 NORMAL and 390 PNEUMONIA cases. The official training set is split into training and validation subsets using an 80:20 division because the original validation folder contains very few images.

All images are resized to 224 x 224 pixels. DenseNet preprocessing is applied before transfer-learning inference. Online augmentation includes rotation, width and height shift, zoom, horizontal flipping and brightness variation. These transformations introduce controlled acquisition-like variation without changing the disease label.

Table 2. Dataset and evaluation setup.

Split / Item	Description
Training source	Official train folder, internally split 80:20
Validation	20% of training data for monitoring and threshold tuning
Test	Official test folder, 624 images
Classes	NORMAL and PNEUMONIA
Image size	224 x 224 x 3
Evaluation focus	Accuracy, precision, recall, F1, macro F1, balanced accuracy

4. Proposed Methodology

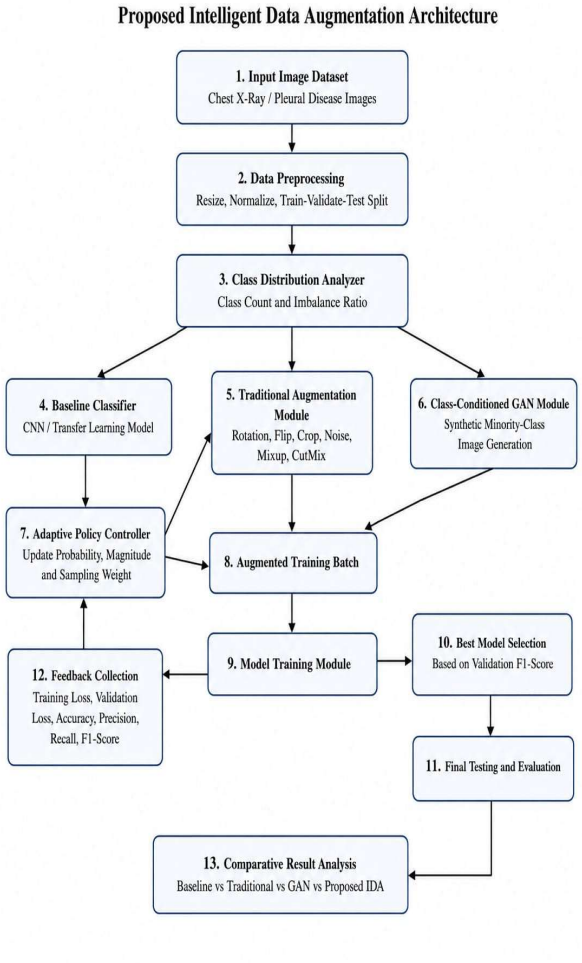

Fig. 1. Proposed architecture.

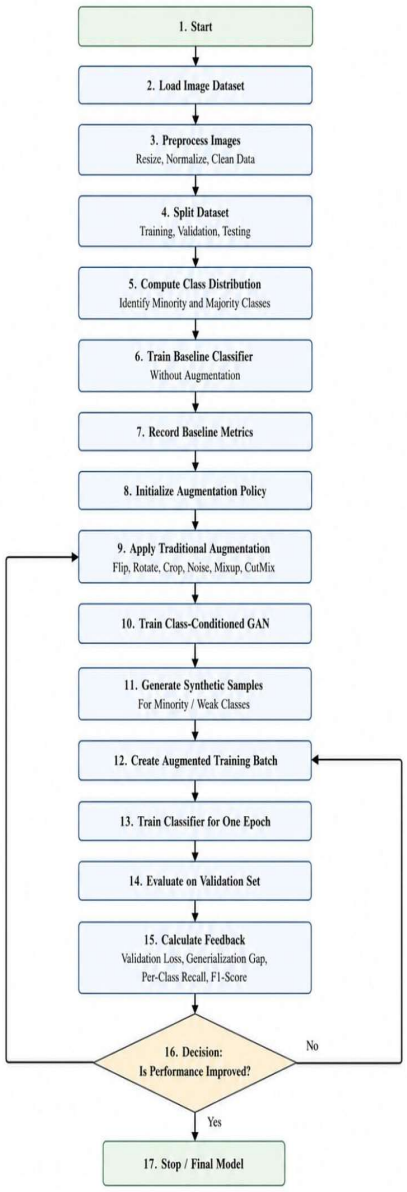

Fig. 2. Methodology flowchart.

The proposed model uses DenseNet121 as a pretrained feature extractor. Dense connectivity improves gradient propagation and encourages feature reuse. The classification head consists of global average pooling, batch normalization, dropout, a dense layer and a sigmoid output neuron. Class weights are used during training to compensate for unequal class frequencies.

Binary focal loss is selected because it reduces the contribution of easy examples and emphasizes difficult samples. This is useful in imbalanced pneumonia classification, where many pneumonia samples may be learned easily while boundary cases require stronger optimization focus. The decision threshold is tuned using validation macro F1-score to avoid relying on the default 0.5 cut-off.

Table 3. Proposed model configuration.

Module	Implementation detail
Input	224 x 224 x 3 preprocessed X-ray
Backbone	DenseNet121
Augmentation	Rotation, shift, zoom, flip, brightness
Loss	Binary focal loss
Optimizer	Adam
Callbacks	Early stopping, reduce learning rate
Threshold	Validation macro-F1 optimized threshold

5. Algorithm

Algorithm 1: DenseNet121-based imbalance-aware pneumonia classification
Input: Chest X-ray dataset D, labels C={NORMAL, PNEUMONIA}, image size 224 x 224, epochs E.
Output: Trained classifier f_theta, selected threshold tau, final test metrics.
1. Load train and test directories from the Kaggle dataset path.
2. Split the training set into training and validation subsets.
3. Resize all images to 224 x 224 and apply DenseNet preprocessing.
4. Apply online augmentation to training images using rotation, shift, zoom, flip and brightness changes.
5. Compute class weights from the training class distribution.
6. Initialize DenseNet121 without top layers and attach a binary classification head.
7. Train the model using Adam optimizer and binary focal loss.
8. Predict validation probabilities and search tau in the range [0.05, 0.95].
9. Select tau that maximizes validation macro F1-score.
10. Apply tau to official test probabilities and compute accuracy, precision, recall, F1-score, macro F1, balanced accuracy and confusion matrix.

Algorithm 1. Proposed training and evaluation algorithm.

6. Mathematical Formulation

Let D={(x_i,y_i)} be the dataset, where x_i is an X-ray image and y_i in {0,1}. The classifier f_theta(x_i) outputs a pneumonia probability p_i. The final class is assigned as y_hat_i=1 if p_i>=tau and y_hat_i=0 otherwise. The threshold tau is selected from the validation set by maximizing macro F1-score.

For class imbalance, weighted loss is used. If w_0 and w_1 are class weights, samples from under-represented classes receive higher contribution. Focal loss is defined using a modulating factor that down-weights easy samples and emphasizes hard examples. This helps prevent the model from being dominated by easy majority-pattern predictions.

7. Experimental Setup

Four configurations are implemented. The baseline CNN uses convolution, batch normalization, pooling, dropout and dense layers. The traditional augmentation CNN adds fixed geometric and photometric transformations. The adaptive augmentation CNN changes augmentation strength based on validation feedback. The final DenseNet121 model uses transfer learning, online augmentation, class weights, binary focal loss and threshold tuning.

Training uses early stopping and learning-rate reduction to limit overfitting. The final test evaluation is performed after selecting the validation threshold. The official test set is not used for model selection.

8. Results and Discussion

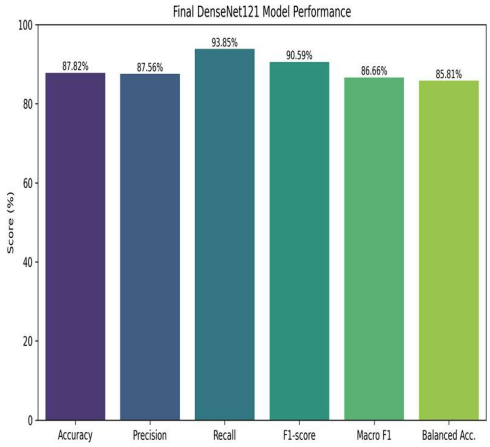

Fig. 3. Final DenseNet121 metric scores.

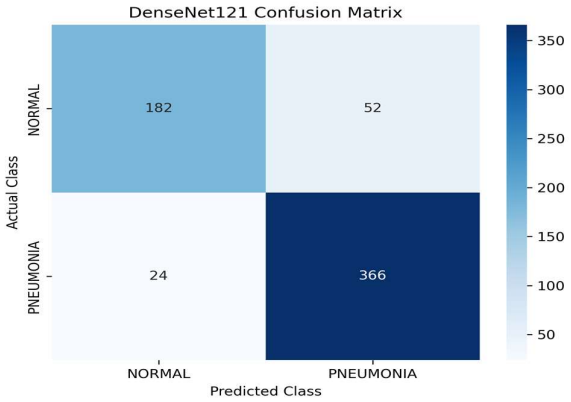

Fig. 4. DenseNet121 confusion matrix.

Table 4. Final DenseNet121 performance.

Metric	Value
Best threshold	0.37
Validation macro F1	95.82%
Validation balanced accuracy	96.83%
Test accuracy	87.82%
Pneumonia precision	87.56%
Pneumonia recall	93.85%
Pneumonia F1-score	90.59%
Macro F1	86.66%
Balanced accuracy	85.81%

The final model achieved 87.82% test accuracy and 86.66% macro F1-score. Pneumonia recall reached 93.85%, indicating strong sensitivity for pneumonia cases. The confusion matrix shows that the final classifier improves class balance compared with custom CNN models, which tended to over-predict the pneumonia class.

Threshold tuning was important. The validation-selected threshold of 0.37 produced a better operating point than the default 0.5 threshold. This confirms that decision calibration is necessary when class distribution is skewed.

9. Ablation Analysis

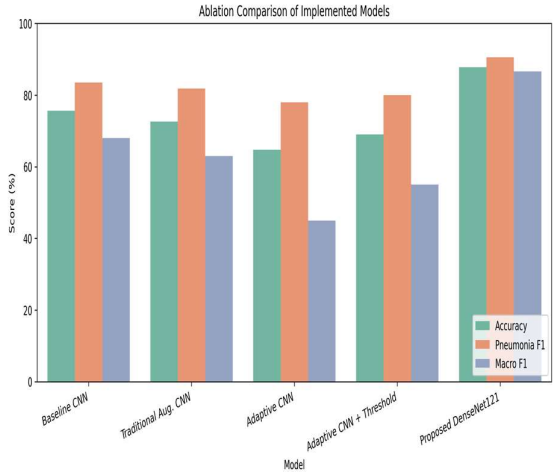

Fig. 5. Ablation comparison.

Table 5. Ablation study.

Model	Strategy	Acc.	Pneu. F1
Baseline CNN	Class weights	75.64	83.55
Traditional CNN	Fixed augmentation	72.60	81.87
Adaptive CNN	Adaptive strength	64.74	77.96
Adaptive + threshold	Threshold tuning	69.00	80.00
Proposed DenseNet121	Aug. + focal + threshold	87.82	90.59

The ablation study shows that augmentation alone did not improve the custom CNN. Adaptive augmentation also remained limited because the feature extractor was not strong enough. DenseNet121 transfer learning produced the best result, showing that strong pretrained representations are important for this dataset. The proposed final configuration improved both accuracy and pneumonia F1-score over all implemented baselines.

10. Technical Contribution

The main technical contribution is the integration of six components into a reproducible pipeline: DenseNet121 feature extraction, online augmentation, class weighting, focal loss, validation macro-F1 threshold selection and balanced metric reporting. Each component addresses a practical limitation in imbalanced medical image classification.

Component	Role
DenseNet121	Feature reuse and stronger representation
Augmentation	Controlled visual diversity
Class weights	Class-frequency compensation
Focal loss	Difficult-sample emphasis
Threshold tuning	Validation-calibrated decision boundary
Balanced metrics	Reliable imbalanced evaluation

Table 6. Technical contribution analysis.

11. Conclusion

This paper presented a DenseNet121-based intelligent augmentation framework for chest X-ray pneumonia classification. The proposed method combines online augmentation, transfer learning, class weighting, focal loss and threshold optimization. It achieved 87.82% accuracy, 90.59% pneumonia F1-score, 86.66% macro F1-score and 85.81% balanced accuracy on the official test set. The results demonstrate that transfer learning and imbalance-aware optimization improve performance compared with custom CNN baselines.

Future work will include fine-tuning upper DenseNet blocks, external dataset validation, Grad-CAM explainability, and comparison with EfficientNet, ConvNeXt and vision transformer models. Controlled GAN or diffusion-based sample generation can also be integrated after expert quality validation.

References

[1] D. Wootton and C. Feldman, “The diagnosis of pneumonia requires a chest radiograph (x-ray)–yes, no or sometimes?,” *Pneumonia*, vol. 5, no. S1, pp. 1–7, Dec. 2014, doi: 10.15172/pneu.2014.5/464.

[2] H. K. Ahmad *et al.*, “Machine Learning Augmented Interpretation of Chest X-rays: A Systematic Review,” *Diagnostics*, vol. 13, no. 4, p. 743, Feb. 2023, doi: 10.3390/diagnostics13040743.

[3] O. Abueed, P. Thakkar, W. H. AlAlaween, Y. Wang, and M. T. Khasawneh, “Automatic semantic segmentation in chest X-ray images using deep learning approaches: a literature review,” *Neural Comput. Appl.*, vol. 38, no. 4, p. 70, Feb. 2026, doi: 10.1007/s00521-025-11743-z.

[4] J. Zhao *et al.*, “Automated Chest X-Ray Diagnosis Report Generation with Cross-Attention Mechanism,” *Appl. Sci.*, vol. 15, no. 1, p. 343, Jan. 2025, doi: 10.3390/app15010343.

[5] S. Yang et al., "Image Data Augmentation for Deep Learning: A Survey," arXiv:2204.08610, 2022.

[6] D. Lewy and J. Mandziuk, "An overview of mixing augmentation methods and augmentation

strategies," Artificial Intelligence Review, vol. 56, pp. 2111-2169, 2023.

[7] M. Xu et al., "A Comprehensive Survey of Image Augmentation Techniques for Deep Learning," Pattern Recognition, vol. 137, p. 109347, 2023.

[8] Z. Yang et al., "A survey of automated data augmentation algorithms for deep learning-based image classification tasks," Knowledge and Information Systems, vol. 65, pp. 2805-2861, 2023.

[9] A. Kebaili et al., "Deep Learning Approaches for Data Augmentation in Medical Imaging: A Review," Journal of Imaging, vol. 9, no. 4, p. 81, 2023.

[10] D. A. Dablain and N. V. Chawla, "Towards Understanding How Data Augmentation Works with Imbalanced Data," arXiv:2304.05895, 2023.

[11] J. Jang et al., "Significantly improving zero-shot X-ray pathology classification via fine-tuning pre-trained image-text encoders," Scientific Reports, vol. 14, p. 23199, 2024.

[12] G. Huang et al., "Densely Connected Convolutional Networks," in Proc. CVPR, 2017, pp. 4700-4708.

[13] T. Y. Lin et al., "Focal Loss for Dense Object Detection," in Proc. ICCV, 2017, pp. 2980-2988.

[14] 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.