Two-Stage Noise-Reduced Pneumonia Detection in Chest X-Ray Images Using Denoising Autoencoder–CNN Integration

Md Abdul Mutalib[1], Kopil Das[2], Mohammad Hossain[3], Shahriar Hussain[2],
Avijit Deb Nath[4], Md Mahmud Zaman[3], Md. Alamgir Hossain[5]
[1]Dept. of CST, Harbin Institute of Technology, Shenzhen, China
[2]Dept. of CSE, North East University Bangladesh, Sylhet 3100, Bangladesh
[3]School of Computing & Informatics, Albhukary International University, Malaysia
[4]Dept. of Mathematics & Statistics, University of Toledo, USA
[5]Skill Morph Research Lab., Skill Morph, Dhaka, Bangladesh
Emails: mutalib.cs@gmail.com, kopildas.cs@gmail.com, mohammad.hossain161786@gmail.com
shahriar08neub@gmail.com, mzmmoon756@gmail.com, avijitdeb2023@gmail.com, alamgir.cse14.just@gmail.com

***Abstract*—Pneumonia remains a major cause of morbidity and mortality worldwide, with chest X-ray imaging serving as a primary tool for early diagnosis. However, the presence of noise, artifacts, and low-quality acquisitions significantly degrades image interpretability and limits the effectiveness of automated deep learning–based detection systems. Conventional convolutional neural networks (CNNs) are particularly sensitive to such degradations, leading to reduced classification reliability. To address this limitation, we propose a hybrid deep learning framework that integrates a denoising autoencoder (DAE) with a CNN for robust pneumonia detection from chest X-ray images. The DAE is trained to reconstruct clean images from noisy inputs, enabling the suppression of noise while preserving diagnostically relevant features, and is subsequently coupled with a CNN classifier in a unified pipeline. The proposed model is evaluated on the publicly available Kaggle chest X-ray dataset using data augmentation, class-weighted loss, and early stopping to mitigate class imbalance and overfitting. Experimental results demonstrate that the DAE-CNN framework consistently outperforms a CNN-only baseline, achieving an accuracy of 92.9%, an F1-score of 94.6%, and a balanced accuracy of 90.9%, compared to 86.1% accuracy and an F1-score of 89.8% obtained without denoising. These results highlight the effectiveness of incorporating learnable denoising into deep learning pipelines and underscore its potential for improving the robustness and reliability of automated pneumonia detection in clinical imaging scenarios.**

***Index Terms*—Pneumonia Detection, Chest X-Ray Imaging, Denoising Autoencoder, Convolutional Neural Network, Deep Learning, Medical Image Classification**

I. Introduction

Pneumonia is a serious infectious disease of the lungs and continues to be a major global health concern, particularly among children, elderly individuals, and immunocompromised patients. According to the World Health Organization, pneumonia remains one of the leading causes of death worldwide, especially in low- and middle-income countries where early diagnosis and access to expert radiologists are often limited [1]–[3]. Timely and accurate detection of pneumonia is therefore critical to reducing mortality and improving patient outcomes.

Chest X-ray imaging is the most widely used diagnostic tool for pneumonia due to its affordability, accessibility, and rapid acquisition. Despite its widespread use, interpreting chest X-ray images is challenging because pneumonia manifestations can be subtle, heterogeneous, and easily confused with other pulmonary conditions such as tuberculosis or pulmonary edema [4] [5]. Moreover, real-world chest X-ray images frequently suffer from noise, low contrast, motion blur, and acquisition artifacts, particularly in portable radiography settings, which further complicates diagnosis and increases the likelihood of misinterpretation [6].

In recent years, deep learning–based computer-aided diagnosis (CAD) systems have demonstrated strong potential in automating pneumonia detection from chest X-ray images. Convolutional Neural Networks (CNNs) have been extensively studied and have achieved impressive performance by learning hierarchical visual features directly from image data [7], [8]. Large-scale studies have shown that CNN-based models can approach or even surpass radiologist-level performance under controlled conditions [9]. However, the reliability of these models is highly dependent on the quality of the input images.

Most CNN-based methods assume clean chest X-rays; however, real images often contain noise and low contrast, which reduces performance and clinical reliability. In practice, when noisy or low-quality images are encountered, model performance degrades significantly, limiting generalizability and clinical applicability [10]. Traditional image preprocessing methods such as filtering, histogram equalization, or handcrafted denoising techniques are often applied separately, but these methods are not adaptive and may remove diagnostically relevant features [11]. Furthermore, many studies treat preprocessing as an external step rather than incorporating it into the learning framework itself.

This highlights a key research gap: designing an end-to-end deep learning framework that simultaneously enhances chest X-ray image quality and performs accurate pneumonia

This manuscript has been accepted for publication in IEEE QPAIN 2026. © IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses. The final published version will be available via IEEE Xplore.

classification. Although Denoising Autoencoders (DAEs) effectively learn noise-robust representations by reconstructing clean images from corrupted inputs [12], [13], their integration with CNN-based pneumonia classifiers remains limited.

To address this gap, this study investigates whether learnable denoising improves robustness in pneumonia detection, whether integrating denoising within the classification pipeline outperforms a standalone CNN, and how such a hybrid framework performs under class imbalance and noisy imaging conditions.

We propose a unified DAE–CNN architecture in which the DAE suppresses noise while preserving clinically relevant features, and the CNN performs classification using the enhanced representations. The model is evaluated on a publicly available chest X-ray dataset using data augmentation, class-weighted loss, and early stopping to ensure robust training.

The main contributions of this work are:

- A unified DAE–CNN framework for end-to-end denoising and pneumonia classification.
- Empirical analysis of the impact of learnable denoising on detection performance.
- Experimental validation demonstrating improved robustness over a CNN-only baseline.

II. Literature Review

In recent years, deep learning–based approaches have become the dominant paradigm for automated pneumonia detection from chest X-ray (CXR) images due to their ability to learn discriminative features directly from raw data. Convolutional Neural Networks (CNNs) have been widely adopted and demonstrated strong performance across multiple public datasets. Rajpurkar et al. introduced CheXNet, a DenseNet-121–based model trained on the ChestX-ray14 dataset, achieving radiologist-level performance for pneumonia detection [5]. Although highly influential, this model relies on large-scale curated data and does not explicitly address image noise or quality degradation, which are common in real clinical environments.

More recent studies have focused on improving CNN architectures through transfer learning and attention mechanisms. Rashid et al. proposed an enhanced DenseNet-121 model augmented with a Convolutional Block Attention Module (CBAM), evaluated on the pediatric CXR dataset by Kermany et al., and reported an accuracy of 98.64%, outperforming the baseline DenseNet model [14]. Similarly, Faiyaz et al. demonstrated that attention-enhanced CNNs using CBAM and squeeze-and-excitation blocks achieved improved sensitivity and specificity, particularly under class imbalance conditions [15]. These studies confirm that attention mechanisms enhance feature discrimination; however, they do not directly mitigate the impact of noise and low-quality scans.

Vision Transformer (ViT)–based models have also been explored for pneumonia detection, demonstrating competitive performance by capturing long-range dependencies in CXR images [16]. However, transformer-based approaches require substantial computational resources and remain sensitive to image noise, limiting practical clinical deployment. Ensemble learning strategies combining multiple CNN backbones (e.g., DenseNet, EfficientNet, Inception) have further improved classification accuracy [17], but at the cost of increased model complexity and inference time, reducing suitability for real-time or resource-constrained medical settings.

Parallel to classification research, several studies have emphasized the importance of image quality enhancement in medical imaging. Denoising Autoencoders (DAEs), originally introduced by Vincent et al., learn noise-invariant representations by reconstructing clean images from corrupted inputs [12]. In medical imaging, DAEs have been applied for noise suppression in MRI, CT, and X-ray images. Banerjee et al. demonstrated that convolutional DAEs effectively reduce noise in chest X-rays while preserving anatomical structures, leading to improved downstream classification stability [17]. However, in most pneumonia detection studies, denoising is treated as a separate preprocessing step rather than being integrated into the classification pipeline.

Although a few recent studies have explored hybrid pipelines that combine image enhancement and classification, such as autoencoder-based denoising followed by CNN classification for thoracic disease detection [18], these approaches generally focus on multi-disease tasks and lack pneumonia-specific analysis under consistent experimental settings. Overall, most existing pneumonia detection models assume relatively clean input images and remain sensitive to noise and acquisition variability. While architectural enhancements such as attention mechanisms and ensemble methods improve feature learning, they do not explicitly address image degradation, and denoising is typically treated as a separate preprocessing step. To overcome these limitations, we propose a unified DAE–CNN framework that integrates learnable denoising and pneumonia classification within a single end-to-end pipeline, enabling improved robustness and superior diagnostic performance compared to a standalone CNN.

III. Methodology

A. Dataset and Pre-Processing

The publicly available Chest X-Ray Images (Pneumonia) dataset was used for training and evaluation [19]. The dataset contains labeled chest radiographs categorized as normal or pneumonia and is divided into training (5,216 images), validation (16 images), and testing (624 images) sets, with class imbalance in the training data (approximately 74% pneumonia). All images were resized to $224 \times 224 \times 3$ and normalized to the $[0, 1]$ range. An overview of the proposed DAE–CNN pipeline is illustrated in Fig. 1.

1) Data Augmentation: Class imbalance and small sample sizes are common problems in medical datasets. During training, an augmentation pipeline was employed to reduce overfitting and enhance generalization. Every image was randomly rotated by ±15°, translated by up to 15% in both horizontal and vertical directions, zoomed by ±15%, and had its contrast adjusted at random. In order to balance the classes, rotation,

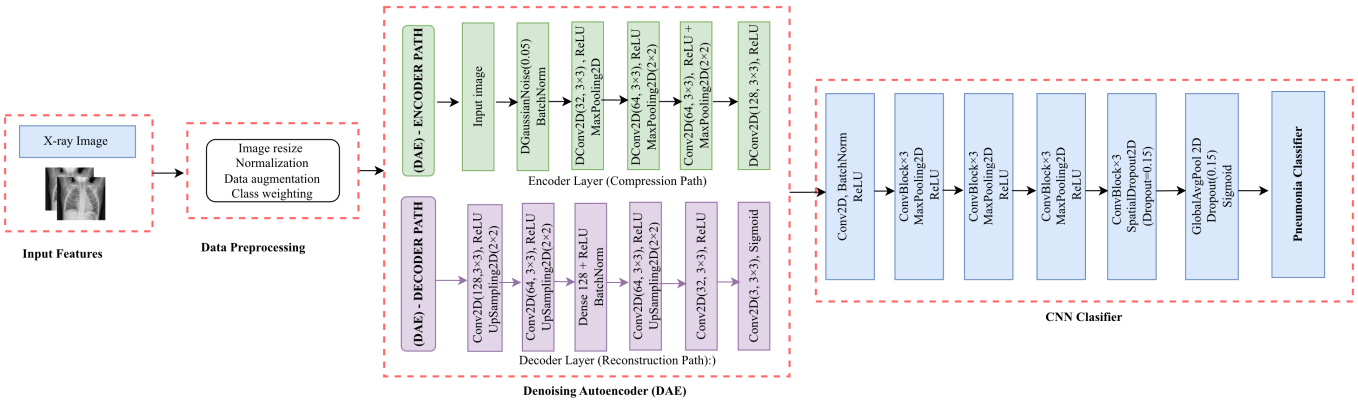


Fig. 1: Overview of the proposed pneumonia classification pipeline, integrating DAE-based denoising and CNN-based classification.

zoom, and mirroring are applied successively in this on-the-fly augmentation, which adheres to the same concepts as the literature. Only the training set was subjected to the augmentation pipeline, testing and validation images were left unaltered.

2) Class Weights: There is an imbalance in the training set. The pneumonia to normal image ratio is approximately 74:26. Class weights were calculated using the inverse of class frequencies and added to the loss function during training in order to address this imbalance. As a result, the classifier is forced to penalize mistakes made by the minority class more severely than by the majority class.

B. Denoising Autoencoder (DAE)

An unsupervised neural network that learns to recreate clean inputs from their distorted versions is called a denoising autoencoder (DAE). It consists of an encoder that reduces the input to a latent representation and a decoder that uses this representation to recover the original picture. Random noise is introduced to the input picture during training, and the DAE learns to eliminate the noise by minimizing the reconstruction error between the original and reconstructed image. Thus, noise-resistant characteristics that might enhance downstream classification are learned by denoising autoencoders.

A convolutional DAE was trained independently on the training set in our system. Every input image was subjected to Gaussian noise with a standard deviation of 0.05. The network was made up of many convolutional layers, symmetrical upsampling layers in the decoder, and max-pooling layers in the encoder. After every convolution, batch normalization and ReLU activations were applied. Using the Adam optimizer with an initial learning rate of 10^{-3} and the mean squared error (MSE) loss, the DAE was trained for 20 epochs. The DAE works as a preprocessing module that removes noise before classification, and its weights remain fixed during CNN training.

C. Convolutional Neural Network (CNN) Classifier

To categorize denoised chest radiographs, a lightweight Convolutional Neural Network (CNN) was created. Convolutional layers, which extract hierarchical characteristics from pictures, pooling layers, which minimize spatial dimensions, and dense layers, which carry out classification, make up CNNs, which are frequently used for image recognition. Depthwise separable convolutions are used in our study's basic classifier to minimize the number of parameters. The architecture consists of:

- Initial stem: a two-dimensional convolution with a stride of two and thirty-two filters, then batch normalization and ReLU activation.
- Three convolutional stacks: separable convolution blocks with 64, 128, and 256 filters are present in each stack. Every block uses batch normalization, ReLU activation, and a 3x3 separable convolution. A 2x2 max-pooling layer diminishes spatial dimensions after every two blocks. To reduce over-fitting, spatial dropout (p = 0.3) is applied to the final stack.
- Global average pooling: followed by a dense layer that uses sigmoid activation and one unit to generate the probability of the pneumonia class.

The binary cross-entropy loss and the Adam optimizer were used to train the CNN for 50 epochs. Training was stabilized by early stopping and learning rate scheduling (ReduceLROnPlateau). By stacking the frozen DAE in front of the CNN, the pre-trained DAE and CNN were integrated to create a hybrid DAE–CNN model that uses a raw chest radiograph as input and outputs a pneumonia likelihood.

D. Threshold Optimisation

A pneumonia probability is produced by the model, and a decision threshold is used to give labels. In order to achieve balanced accuracy, we adjusted this threshold on the validation set using a $[0, 1]$ sweep (step $= 0.01$), achieving an optimal of 0.52. Table I provides a summary of all training and optimization parameters.

E. Evaluation Metrics

Both threshold-dependent and threshold-independent metrics obtained from the confusion matrix were used to assess the model's performance on the held-out test set. The numbers

TABLE I: Hyperparameters and training configuration.

Setting	Value
Input size	$224 \times 224 \times 3$
Batch size	32
Optimizer (CNN)	Adam
Initial learning rate (CNN)	3×10^{-4}
Loss (CNN)	Binary cross-entropy
Metric for checkpointing	Val ROC–AUC
Max epochs (CNN)	50
Early stopping	Patience = 6 (val AUC)
LR scheduler	ReduceLROnPlateau (factor = 0.5, patience = 3, min LR = 10^{-6})
Class imbalance handling	Class-weighted loss (balanced)
Normalization	Rescaling to $[0, 1]$
Augmentation (train only)	Rotation $\pm 15^{\circ}$, zoom $\pm 15\%$, translation $\pm 15\%$, contrast 0.1
Regularization	L2 weight decay (10^{-6}), dropout = 0.15, spatial dropout = 0.15
Threshold selection	Validation sweep step = 0.01; maximize balanced accuracy
DAE noise model	Gaussian noise $\sigma = 0.05$
DAE loss / epochs	MSE / 20 epochs

of true positives, true negatives, false positives, and false negatives are represented as TP, TN, FP, and FN, respectively (where the positive class corresponds to pneumonia). Through accuracy, overall correctness was measured,

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}. \tag{1}$$

Precision and recall were reported to describe performance on the pneumonia class. While recall (sensitivity) assesses the percentage of real pneumonia cases that are accurately detected, precision measures the percentage of projected pneumonia cases that are actually pneumonia:

$$\text{Precision} = \frac{TP}{TP + FP}, \tag{2}$$

$$\text{Recall} = \frac{TP}{TP + FN}. \tag{3}$$

The F_1-score uses their harmonic mean to characterize the precision-recall trade-off,

$$F_1 = \frac{2\,(\text{Precision})(\text{Recall})}{\text{Precision} + \text{Recall}}. \tag{4}$$

Because of the imbalance in the dataset, balanced accuracy was also calculated. The average recall for each class is known as balanced accuracy. It is comparable to the mean of sensitivity and specificity in the binary case:

$$\text{BalancedAccuracy} = \frac{1}{2}\left(\frac{TP}{TP + FN} + \frac{TN}{TN + FP}\right). \tag{5}$$

We employed the receiver operating characteristic (ROC) curve and the area under the ROC curve (AUC) to assess ranking performance across all decision thresholds. The true-positive rate (TPR) is shown against the false-positive rate (FPR) on the ROC curve, where

$$\text{FPR} = \frac{FP}{FP + TN}. \tag{6}$$

The AUC could be seen as the probability that a randomly selected positive example would be ranked higher by the model than a randomly selected negative example; $\text{AUC} = 1$ shows perfect ranking, whereas random guessing is represented by $\text{AUC} = 0.5$. Additionally, by changing the decision threshold and charting precision versus recall, the precision–recall (PR) curve was created. PR curves are especially useful when there is class imbalance.

Finally, robust agreement/correlation measures for unbalanced classification were presented, including Cohen's κ and Matthews correlation coefficient (MCC). The correlation between true and predicted labels is measured by MCC:

$$\text{MCC} = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}. \tag{7}$$

Cohen's κ measures agreement corrected for chance,

$$\kappa = \frac{p_o - p_e}{1 - p_e}, \tag{8}$$

where p_e is the agreement predicted by chance and p_o is the observed agreement.

IV. Results and Discussion

A. Training Behaviour

There were two training phases for the proposed hybrid pipeline. Initially, the Kaggle chest X-ray images were used to train a denoising autoencoder (DAE) utilizing mean-squared reconstruction loss for 20 epochs. The validation reconstruction error progressively decreased to about 2.3×10^{-3} by the end of this phase, during which the network learned to suppress simulated noise while maintaining diagnostic structures.

The decoded images were cached ahead of applying random augmentations (rotation, zoom, translation, and contrast adjustments) in order to guarantee that augmentation remained stochastic across epochs and to minimize training overhead. The CNN classifier was trained using 4695 training images and 521 validation images using class weighting to counteract class imbalance after the DAE weights were frozen. The training and validation accuracy-loss curves (see Fig. 2) demonstrate that while the validation accuracy remained high but varied due to small-batch variance and threshold effects, the accuracy increased consistently and saturated around 0.97 on the training set.

In the early epochs, the validation AUC increased quickly, reaching a peak of roughly 0.9986 close to epoch 18. When the validation AUC did not improve for three consecutive epochs, a `ReduceLROnPlateau` scheduler decreased the learning rate in half; when the AUC plateaued, early stopping was initiated. Caching, data augmentation, learning-rate scheduling, and early stopping all worked together to reduce overfitting, promote stable convergence, and let the CNN learn from denoised representations. Lastly, a better compromise between sensitivity and specificity on the test data was achieved by choosing the decision threshold on the validation set by maximizing balanced accuracy rather than a fixed probability threshold of 0.5.

B. Test-Set Performance

The 624-image held-out test set was used to assess performance. The model obtained an accuracy of 0.929, precision

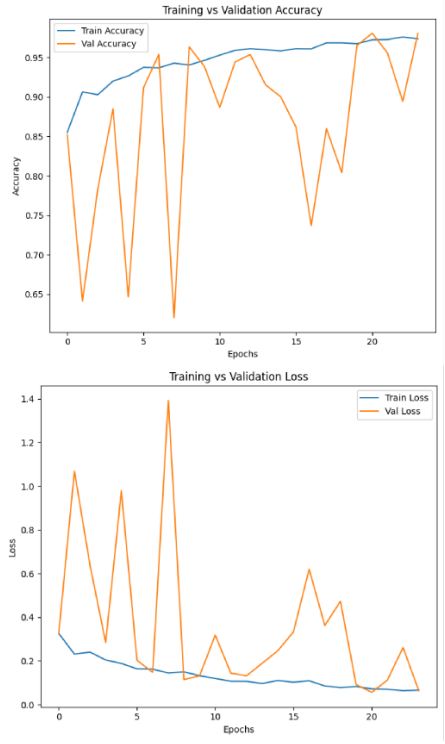


Fig. 2: Training and validation accuracy/loss curves.

of 0.904, recall of 0.992, and F_1-score of 0.946 using the operational threshold chosen on the validation set (by maximizing balanced accuracy with a threshold sweep). A balanced accuracy of 0.909, a Matthews correlation coefficient (MCC) of 0.852, and a Cohen's κ of 0.845 demonstrate robustness under class imbalance and good agreement beyond chance. Furthermore, outstanding threshold-independent discriminative performance is indicated by the ROC–AUC of 0.983.

The main test-set results are summarized in Table II.

TABLE II: Test-set performance of the proposed DAE–CNN model (updated).

Metric	Value
Accuracy	0.929
Precision (PPV)	0.904
Recall (Sensitivity)	0.992
F_1-Score	0.946
Balanced Accuracy	0.909
ROC–AUC	0.983
MCC	0.852
Cohen's κ	0.845

C. Confusion Matrix Interpretation

The confusion matrix for the test set is displayed in Fig. 3. The model produced a sensitivity (recall) of $387/(387+3) = 0.992$ by accurately identifying 387 pneumonia cases while missing just 3. The low false-negative count is therapeutically beneficial from a screening standpoint since treatment delays could result from missed pneumonia cases. Among normal photos, the model generated 41 false positives, resulting in a specificity of $193/(193+41) = 0.825$.

D. ROC and Precision–Recall Curve Analysis

Strong class separability is further confirmed by threshold-sweeping examination. With an AUC of 0.983, the ROC

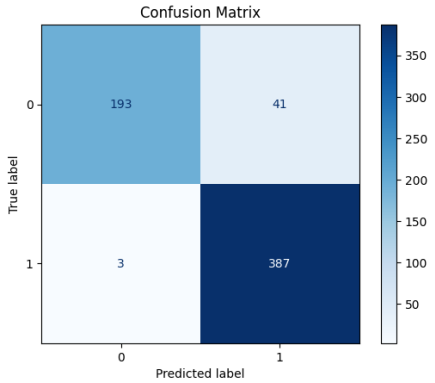


Fig. 3: Confusion matrix illustrating the classification performance of the proposed model on the test dataset.

curve (Fig. 4) demonstrates outstanding discriminating across operational points. As the threshold is lowered, high sensitivity can be attained with a controlled precision decline, as seen by the precision–recall curve (Fig. 5), which stays high over a wide recall range.

These curves offer a rational foundation for choosing an operating point that is in line with therapeutic goals. By concentrating on the positive (pneumonia) class, the precision–recall curve (Fig. 5) offers further insight under class imbalance.

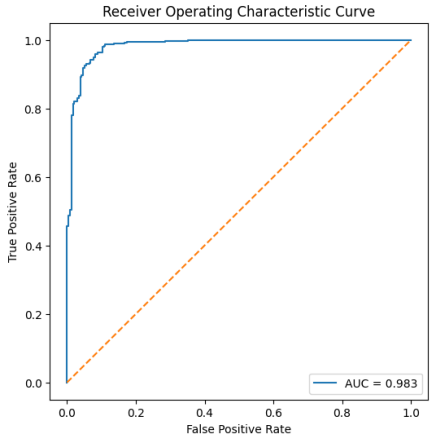


Fig. 4: Receiver operating characteristic curve demonstrating the discriminative capability of the proposed model (AUC reported).

According to the observed PR behavior, precision decreases as recall gets closer to 1.0 but stays high throughout a broad recall range. Pushing the threshold to optimize recall (minimize missed pneumonia) usually results in more false positives, which reduces precision. This trade-off is expected when the dataset is imbalanced. As a result, the PR curve is helpful in choosing an operational threshold that corresponds to the desired therapeutic goal (e.g., high-precision confirmation support vs. high-sensitivity triage).

E. Discussion

The results demonstrate that integrating learnable denoising with classification improves the robustness of pneumonia detection from chest X-ray images. The proposed DAE+CNN

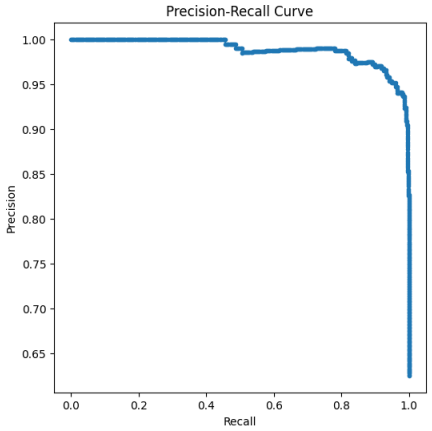


Fig. 5: Precision–Recall (PR) curve of the proposed model on the test set.

framework outperformed the CNN-only baseline, achieving 92.9% accuracy and a 94.6% F_1-score, indicating that the DAE effectively suppresses acquisition noise while preserving diagnostically relevant structures. Unlike fixed handcrafted preprocessing, the learnable denoising module is embedded within the pipeline, enabling data-driven enhancement without removing subtle clinical cues [11], [12]. Compared to approaches that emphasize deeper backbones, attention mechanisms, or ensemble strategies [5], [14], the proposed method directly addresses input degradation with lower computational overhead.

The achieved recall of 0.992, corresponding to only three missed pneumonia cases, is particularly relevant for screening scenarios where false negatives are critical. Although specificity (0.825) reflects the inherent sensitivity–specificity trade-off, threshold optimization based on balanced accuracy ensured stable performance under class imbalance. The ROC–AUC of 0.983 further confirms strong discriminative capability across operating points. Despite these promising results, evaluation was limited to a single public dataset; future work will focus on external validation, ablation analysis to quantify the contribution of the DAE, incorporation of interpretability techniques such as Grad-CAM, and exploration of calibration and uncertainty estimation for improved clinical reliability.

V. Conclusion

This paper presented a hybrid Denoising Autoencoder (DAE) and CNN architecture for robust pneumonia detection from chest X-ray images, integrating learnable denoising with classification in a unified end-to-end framework. Our proposed approach suppresses noise while preserving diagnostically relevant features and was trained using data augmentation, class weighting, and validation-based threshold optimization. Experimental results on a public chest X-ray dataset show that the DAE+CNN model outperforms a CNN-only baseline, achieving 92.9% accuracy, a 94.6% F1-score, a ROC-AUC of 98.3%, and a high recall of 99.2%, demonstrating strong discriminative performance and suitability for screening-oriented applications. Future work will investigate external multi-institutional validation, robustness to domain shifts, explainability techniques such as Grad-CAM, calibration and uncertainty estimation, and extension to multi-class thoracic disease classification.

References

[1] World Health Organization, “Pneumonia,” 2024.
[2] GBD 2017 Lower Respiratory Infections Collaborators, “Quantifying risks and interventions that have affected the burden of lower respiratory infections among children younger than 5 years: a global analysis,” *The Lancet Infectious Diseases*, vol. 19, no. 11, pp. 1235–1248, 2019.
[3] C. e. a. Troeger, “Estimates of the global, regional, and national morbidity, mortality, and aetiologies of lower respiratory infections in 195 countries,” *The Lancet Infectious Diseases*, vol. 18, no. 11, pp. 1191–1210, 2018.
[4] Radiopaedia.org, “Radiographic patterns of pneumonia,” 2025.
[5] P. Rajpurkar, J. Irvin, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Ding, A. Bagul, C. Langlotz, K. Shpanskaya, M. P. Lungren, and A. Y. Ng, “Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning,” 2017.
[6] A. P. Brady, “Error and discrepancy in radiology: inevitable or avoidable?” *Insights into Imaging*, vol. 8, no. 1, p. 171–182, Dec. 2016.
[7] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553, p. 436–444, May 2015.
[8] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *CoRR*, vol. abs/1409.1556, 2014.
[9] P. Lakhani and B. Sundaram, “Deep learning at chest radiography: Automated classification of pulmonary tuberculosis by using convolutional neural networks,” *Radiology*, vol. 284, no. 2, p. 574–582, Aug. 2017.
[10] S. Yu, M. Chen, E. Zhang, J. Wu, H. Yu, Z. Yang, L. Ma, X. Gu, and W. Lu, “Robustness study of noisy annotation in deep learning based medical image segmentation,” *Physics in Medicine & Biology*, vol. 65, no. 17, p. 175007, Aug. 2020.
[11] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 4th ed. New York, NY, USA: Pearson, 2018.
[12] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P.-A. Manzagol, “Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion,” vol. 11, p. 3371–3408, Dec. 2010.
[13] V. S. Kumar and V. Jayalakshmi, “Reconstructing the medical image by autoencoder with stochastic processing in neural network,” in *2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA)*. IEEE, Sep. 2021, p. 1521–1526.
[14] A. B. Rashid, J. Asma, K. Barua, and D. Das, “An enhanced deep learning framework for pneumonia detection in chest x-rays,” *SN Computer Science*, vol. 6, no. 5, May 2025.
[15] S. Faiyaz Waris, V. Naganjaneya Rohith Karthikeya, and M. Venkata Lakshmi Supraja, “Sparse-attention based deep convolutional neural network for pneumonia detection,” in *2023 3rd International Conference on Intelligent Technologies (CONIT)*. IEEE, Jun. 2023, p. 1–8.
[16] J. Johnson, E. Edwin, E. V, S. Kirubakaran S, and M. Thanka, “Vision transformer-based pneumonia detection for chest x-ray images,” *JOURNAL OF SENSOR SCIENCE AND TECHNOLOGY*, vol. 34, no. 3, p. 174–179, May 2025.
[17] A. Banerjee, A. Sarkar, S. Roy, P. K. Singh, and R. Sarkar, “Covid-19 chest x-ray detection through blending ensemble of cnn snapshots,” *Biomedical Signal Processing and Control*, vol. 78, p. 104000, Sep. 2022.
[18] V. A. Ferdinand, V. Nawir, G. E. Henry, and T. W. Cenggoro, “Auxiliary learning and image enhancement based pipeline to improve cnn performance for pneumonia detection,” in *2024 8th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE)*. IEEE, Aug. 2024, p. 64–69.
[19] P. T. Mooney, “Chest x-ray images (pneumonia),” 2018, accessed: 2025-12-20.