Enhancing Medical Image Quality Using Deep Learning Techniques

Christian De La Torre, Araceli Castelan, Jonathan Suconota, Kevin Sookram
Department of Computer Science, CUNY Queens College, NY

June 2025

Abstract

Radiology images play a critical role in clinical decision-making, supporting healthcare professionals in both diagnostic and therapeutic procedures. Among these, X-ray imaging is one of the most widely used modalities for detecting fractures, infections, and various underlying conditions. Ideally, X-ray images should be of high quality to enable accurate interpretation and effective patient care. However, in practice, the quality of these images can be compromised due to factors such as outdated or poorly maintained equipment, suboptimal imaging protocols, or inadequate lighting conditions. Low-quality X-ray images may hinder diagnostic accuracy, potentially leading to missed or incorrect diagnoses. This project investigates the use of advanced deep learning techniques, including super-resolution convolutional neural networks (SRCNN, VDSR) and generative adversarial networks (GANs), to enhance the quality of medical images. Furthermore, we evaluate the effectiveness of traditional and novel image quality metrics, aiming to bridge the gap between technical enhancement and clinical utility.

1 Introduction

Low-resolution medical images from aging X-ray and fundus cameras increase the risk of misdiagnosis, posing a significant challenge to effective healthcare delivery. In many hospitals and clinics, especially those with limited resources, imaging equipment may be outdated or poorly maintained, leading to images that lack the clarity necessary for accurate interpretation. This can compromise diagnostic confidence and ultimately impact patient outcomes.

Recent advances in deep learning have opened new possibilities for enhancing the quality of medical images. Convolutional neural networks (CNNs) such as Super-Resolution CNN (SRCNN)[**?**] and Very Deep Super-Resolution (VDSR)[**?**] have demonstrated strong potential to restore detail in low-resolution images. These models, however, present trade-offs: while deeper networks like VDSR are better at reconstructing fine structures essential for diagnosis, they often require more computational resources and longer training times compared to shallower models like SRCNN.

Despite progress in model development, the field still relies heavily on standard quantitative metrics such as Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM) to evaluate image quality. These metrics, while useful, do not always correlate with perceptual quality or clinical relevance. As a result, there is growing interest in developing and applying new evaluation methods that better reflect the true utility of enhanced medical images.

This project addresses these challenges by benchmarking deep learning models for medical image super-resolution and assessing the alignment between quantitative metrics, perceptual quality, and potential clinical impact. Our goal is to identify approaches that not only improve technical image quality but also enhance diagnostic confidence and patient care.

2 Related Work

Deep learning has significantly advanced medical image enhancement, with recent research focusing on novel model architectures, evaluation metrics, and supporting techniques. We group relevant literature into three subsections: models for image enhancement, metrics and evaluation, and supporting techniques for data and restoration.

2.1 Enhancing Medical Image Quality

Patil et al. (2024) [1] address the enhancement of low-resolution chest X-ray images using deep learning models. They compare the performance of SRCNN, known for its speed and computational efficiency, with VDSR, which excels at recovering fine image details crucial for diagnostics but requires more resources and training time. Their evaluation using PSNR and SSIM reveals that VDSR outperforms SRCNN in restoring structural edges and clarity, making it preferable for clinical tasks, while SRCNN remains suitable for resource-constrained environments.

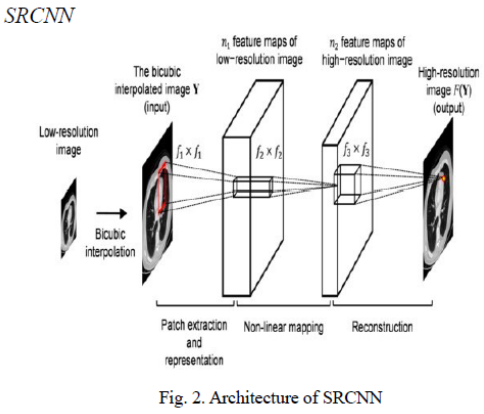

Fig. 2. Architecture of SRCNN

Figure 1: Architecture of SRCNN model for image super-resolution, adapted from [1].

2.2 Integrated Deep Learning for Glaucoma Detection

Khan (2022) [2] developed an integrated deep learning pipeline for early glaucoma detection from retinal fundus images. The study uses a custom CNN for automated feature extraction and classification of healthy versus glaucomatous eyes, achieving high accuracy and AUC across multiple datasets. This approach supports non-invasive, large-scale glaucoma screening.

TABLE I. PERFORMANCE METRICS

Metric	Obtained Value (in %)
Accuracy	99.06
Precision	99.19
Recall	98.13
F1 Score	97.65

Figure 2: Performance metrics (accuracy, precision, recall, F1) for glaucoma detection using a deep learning model, adapted from [2].

2.3 ClearSight

Wang (2023) [3] introduces ClearSight, a convolutional neural network designed for dehazing UAV-captured road images. While focused on environmental artifacts in aerial images, the methodology—training on synthetic and real hazy images and evaluating with PSNR and SSIM—demonstrates how deep learning restoration techniques can recover obscured structural details. The work provides transferable strategies and evaluation principles for medical image enhancement in visually challenging condi-

tions.

Figure 4. Example of image dehazing dataset

Figure 3: Example images from the dehazing dataset used in ClearSight, adapted from [3].

2.4 Image Comparative Index (ICI)

Kaderuppan (2024) [4] presents the Image Comparative Index (ICI), a novel metric for evaluating image similarity, especially in the context of super-resolution. Unlike PSNR and SSIM, which may misrepresent visual differences, ICI performs pixel-level comparisons and is robust to varying image depths, yielding assessments closely aligned with human perception. Experiments show ICI's advantage in identifying artifacts missed by conventional metrics, particularly in medical and scientific imaging.

TABLE I COMPARISON OF SOURCE & BSRGAN-GENERATED IMAGES USING POPULAR IMAGE QUALITY METRICS

	Images		Metrics					
	Source	Generated	PSNR	IMMSE	SSIM	MS-SSIM	$FSIM_C$	ICI
❶			28.9243	83.3003	0.9809	R – 0.8898 G – 0.9531 B – 0.8219	0.9299	0.010883
❷			29.875	66.924	0.9608	R – 0.9504 G – 0.9667 B – 0.8762	0.9134	0.007504
❸			29.4908	73.1135	0.9865	R – 0.9560 G – 0.9641 B – 0.8954	0.9517	0.008208

Figure 4: Comparison of image quality metrics, including ICI, for super-resolved images, adapted from [4].

2.5 A Comprehensive Survey

Bellal et al. (2022) [5] provide a survey of deep learning applications in abdominal imaging, covering reconstruction, super-resolution, and artifact suppression. They review numerous evaluation metrics—PSNR, SSIM, CNR, FID, DSC—and highlight their clinical utility, along with challenges like data scarcity, computational constraints, and bias. Their work underscores the need for both technical performance and clinical relevance in imaging AI.

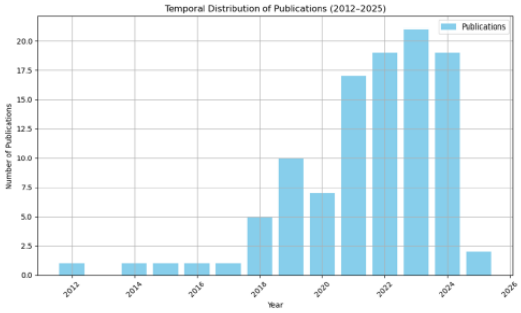

FIGURE 2. Temporal distribution of publications in deep learning for abdominal imaging (2012–2025).

Figure 5: Temporal distribution of publications in deep learning for abdominal imaging, adapted from [5].

2.6 Manifold and Patch-Based for Fine-Grained Image Retrieval

Yuan et al. (2021) [6] introduce a manifold and patch-based metric learning method for fine-grained image retrieval without labeled data. By leveraging KNN-based pseudo-supervision and patch-level feature alignment, the model achieves state-of-the-art performance in distinguishing subtle differences. The approach inspires our own data preparation strategies for dealing with class imbalance and enhancing model robustness in medical datasets.

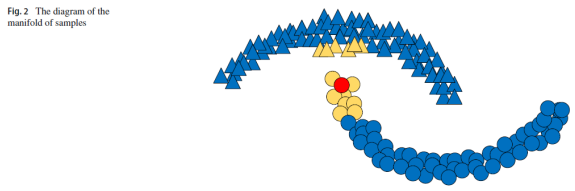

Figure 6: Diagram of the manifold of samples used for patch-based metric learning, adapted from [6].

2.7 Manifold and Patch-Based for Fine-Grained Image Retrieval

Yuan et al. (2021) [6] introduce a manifold and patch-based metric learning method for fine-grained image retrieval without labeled data. By leveraging KNN-based pseudo-supervision and patch-level feature alignment, the model achieves state-of-the-art performance in distinguishing subtle differences. The approach inspires our own data preparation strategies for dealing with class imbalance and enhancing model robustness in medical datasets.

Figure 7: Diagram of the manifold of samples for patch-based metric learning, adapted from [6].

2.8 Compare/Contrast Approaches

The reviewed works span a spectrum from traditional CNN-based super-resolution (Patil et al.) and adversarial networks (Bing et al.) to restoration under challenging visual conditions (Wang) and advanced metric learning (Yuan). While Patil et al. and Bing et al. focus on model improvement, Kaderuppan and Bellal et al. emphasize evaluation, with ICI proposed as a superior perceptual metric over PSNR and SSIM. Yuan's and Bellal's contributions address data and evaluation strategy gaps. The diversity in model complexity, restoration targets, and metric selection reveals both convergence and divergence in approaches to medical image quality enhancement.

2.9 Summary and Gaps:

Across these works, both advanced model design and robust, perceptually meaningful evaluation are essential for effective medical image enhancement. While traditional metrics remain common, new approaches like ICI and task-specific learning strategies are emerging to address clinical needs. However, the relative impact of these metrics on clinical decision-making and their generalizability across modalities remain open research questions that we aim to explore in our project.

3 Experimental Datasets

This study utilized three publicly available chest X-ray datasets from Kaggle.

- Coronahack Chest X-ray [**?**], contains chest radiographs labeled for COVID-19 detection.
- Chest X-ray Images for Pneumonia Detection [**?**], provides labeled chest X-rays for distinguishing between normal and pneumonia cases.
- The NIH Chest X-rays dataset [**?**], includes over 100,000 chest X-ray images labeled for various thoracic diseases.

4 Methodology

4.1 Data Preparation

The data sets were retrieve from Kaggle using the Kaggle API. Due to the time constraints we limited our sample size to 200 images. The results showed consistency with the metrics from the original paper. Images of PNG, JPG, and JPEG format were used.

4.2 Model Architecture

Two super-resolution models were implemented: the Super-Resolution Convolutional Neural Network (SRCNN) and a reduced-depth Very Deep Super-Resolution (VDSR) model. Both models operate on grayscale images resized to 64×64 pixels.

4.2.1 SRCNN

The SRCNN model consists of three convolutional layers, designed to progressively extract features and reconstruct high-resolution outputs. The detailed configuration is shown in Table 1 and 2.

Table 1: SRCNN Convolutional Layers

Layer	Filters	Kernel Size
1	64	9×9
2	32	5×5
3	1	5×5

Table 2: SRCNN Activation Functions

Layer	Activation
1	ReLU
2	ReLU
3	Linear

The SRCNN model receives a low-resolution grayscale image of size $64 \times 64 \times 1$ and outputs a reconstructed high-resolution image of the same size.

4.2.2 VDSR

The VDSR model follows a deeper architecture with residual learning to improve super-resolution performance. A total of 11 convolutional layers are used, including 10 hidden layers with ReLU activation and one output layer with a linear activation. Additionally, a residual connection adds the input image directly to the output. The configuration is presented in Table 3 and 4.

Table 3: VDSR Convolutional Layers

Layer(s)	Type	Filters	Kernel Size
1 – 10	Conv2D	64	3×3
11	Conv2D	1	3×3

The VDSR model also processes input images of size $64 \times 64 \times 1$, producing a super-resolved image with residual learning to enhance reconstruction accuracy.

Table 4: VDSR Activation and Residual

Layer(s)	Activation / Residual
1 – 10	ReLU
11	Linear
-	Add (Residual)

4.2.3 Summary

A comparison of the two model configurations is provided in Table 5.

Table 5: Model Layers and Output

Model	Total Layers	Output Shape
SRCNN	3 Conv2D	$64 \times 64 \times 1$
VDSR	11 Conv2D	$64 \times 64 \times 1$

Table 6: Model Activations and Residual Learning

Model	Activation Functions	Residual Learning
SRCNN	ReLU, Linear	No
VDSR	ReLU, Linear	Yes

4.3 Model Architectures and Enhancement Pipeline

This work implements an integrated pipeline for medical image enhancement, combining super-resolution and dehazing techniques. Three distinct models are employed: Super-Resolution Convolutional Neural Network (SRCNN), Very Deep Super-Resolution Network (VDSR), and a ClearSight-inspired dehazing network. All models process grayscale chest X-ray images resized to 64×64 pixels.

4.3.1 Super-Resolution Models

The SRCNN and VDSR models are responsible for restoring high-resolution details from degraded images. Their configurations are summarized in Tables 7 and 8.

The VDSR model incorporates residual learning by adding the model's output to the original low-

Table 7: SRCNN Model Configuration

Layer	Type	Size	Activation
1	Conv2D	64@9 × 9	ReLU
2	Conv2D	32@5 × 5	ReLU
3	Conv2D	1@5 × 5	Linear

Table 8: VDSR Model Configuration

Layer(s)	Type	Size	Activation
1 – 10	Conv2D	64@3 × 3	ReLU
11	Conv2D	1@3 × 3	Linear
-	Add (Residual)	-	-

resolution input, facilitating faster convergence and improved reconstruction.

4.3.2 ClearSight-inspired Dehazing Network

To further enhance image clarity, a novel dehazing model inspired by ClearSight and transformer-based mechanisms is integrated into the pipeline. The network combines convolutional feature extraction, multi-head self-attention, and non-linear refinements. Its architecture is presented in Table 9and 10.

Here, H and W refer to the image dimensions (64× 64 in this study).

Table 9: ClearSight Dehazing: Conv & Reshape Layers

Lyr	Type	Details	Act.
1	Conv2D	32, 3×3, same pad	ReLU
2	Conv2D	64, 3×3, dilation=2	ReLU
3	Reshape	$(H \times W, 64)$	-
4	LayerNorm	-	-

4.3.3 Iterative Haze Removal with Stopping Criterion

The dehazing model is applied iteratively to prevent over-processing. After each iteration, the Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM) are computed. If deterioration beyond predefined thresholds is observed, the haze removal process is halted, preserving image quality.

Table 10: ClearSight Dehazing: Attention & Output Layers

Lyr	Type	Details	Act.
5	MultiHeadAttn	4 heads, key dim=16	-
6	Dense	64 units, FC	ReLU
7	Reshape	$(H, W, 64)$	-
8	Conv2D	32, 3×3, same pad	ReLU
9	Conv2D	1, 3×3, same pad	Sigmoid

4.3.4 Integrated Enhancement Pipeline

The complete enhancement pipeline combines super-resolution (SRCNN or VDSR) with controlled dehazing. Quantitative performance is evaluated using average PSNR and SSIM across the test set. A summary of the model configurations is provided in Table 11.

Table 11: Model Layers and Output Shape

Model	Total Layers	Output Shape
SRCNN	3 Conv2D	64 × 64 × 1
VDSR	11 Conv2D	64 × 64 × 1
ClearSight	9 Total	64 × 64 × 1

Table 12: Model Activations and Residual Learning

Model	Activations	Residual Learning
SRCNN	ReLU, Linear	No
VDSR	ReLU, Linear	Yes
ClearSight	ReLU, Sigmoid, Attention	No

5 Results

The following figures present the results from the NIH dataset, Chest X-ray Pneumonia dataset, and CoronaHack dataset, along with haze removal experiments.

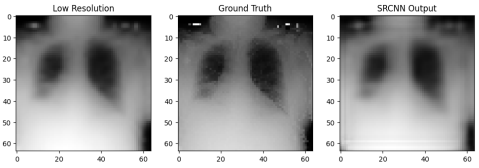

Figure 8: NIH chest X-ray dataset result (Baseline).

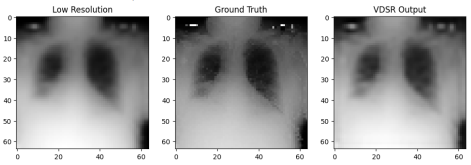

Figure 9: NIH chest X-ray dataset result with VDSR enhancement.

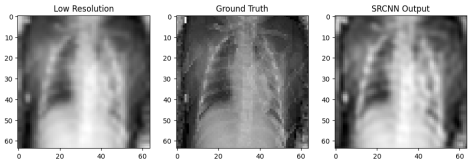

Figure 10: Pneumonia dataset result using SRCNN.

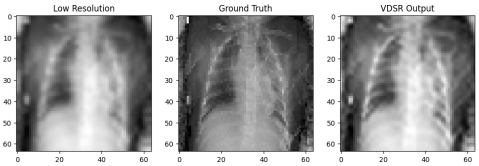

Figure 11: Pneumonia dataset result using VDSR.

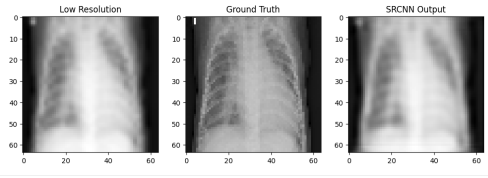

Figure 12: CoronaHack dataset result using SRCNN.

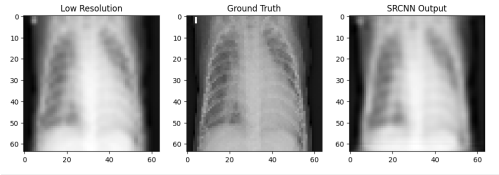

Figure 13: CoronaHack dataset result using VDSR.

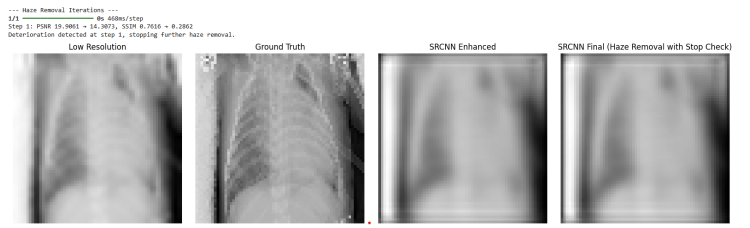

Figure 14: Haze removal with VDSR enhancement.

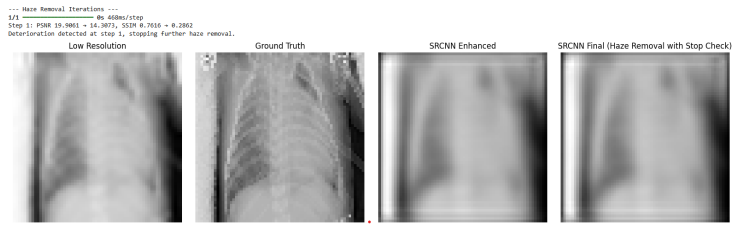

Figure 15: Final haze removal output after iterative processing.

6 Conclusion and Future Work

In this study, we aimed to incorporate the haze removal technique from ClearSight: Deep Learning-Based Image Dehazing for Enhanced UAV Road Patrol with the expectation...

In this study, we aimed to incorporate the haze removal technique from ClearSight: Deep Learning-Based Image Dehazing for Enhanced UAV Road Patrol with the expectation that it would reduce pix-

elation around the bones and increase imasge contrast. However, rather than improving clarity, the ClearSight algorithm produced images that appeared overly gray. It seems that the algorithm primarily enhanced the more luminous features—such as the bones (which appear white)—resulting in unintended image degradation.

To mitigate this, we experimented with early stopping criteria to halt the haze removal process once the image quality deteriorated beyond a certain threshold. Unfortunately, this adjustment yielded minimal improvement. Upon revisiting the original paper, we realized that ClearSight's success heavily depends on the availability of large amounts of foggy and noisy image data from similar scenes for training, as noted by the authors: "deep learning models have shown promising results, but their success has been highly dependent on large amounts of foggy and noisy image data from similar scenes for neural network training." Our dataset did not meet this requirement, limiting the model's effectiveness in our context.

Although this experiment did not meet expectations, it raised an intriguing question: could we develop a model specifically designed to detect and replace pixelated portions of bones in enhanced medical images? Exploring such an approach could offer substantial improvements in image quality. Given the demonstrated potential of deep learning in enhancing medical images, this direction represents a promising avenue for future research.

Additionally, to further advance this work, we plan to expand our evaluation by testing on a wider variety of datasets and collaborating closely with clinicians to obtain qualitative feedback. Incorporating new quantitative metrics tailored for medical image enhancement will also be critical for assessing improvements more comprehensively.

—

7 Experimental Datasets

This study utilized three publicly available chest X-ray datasets from Kaggle.

- Coronahack Chest X-ray [**?**], contains chest radiographs labeled for COVID-19 detection.
- Chest X-ray Images for Pneumonia Detection [**?**], provides labeled chest X-rays for distinguishing between normal and pneumonia cases.
- The NIH Chest X-rays dataset [**?**], includes over 100,000 chest X-ray images labeled for various thoracic diseases.

References

[1] N. Patil, S. K. Shah, V. Kumbhar, and T. P. Singh, "Enhancing medical image quality using deep learning techniques," in *2024 First International Conference for Women in Computing (InCoWoCo)*. IEEE, 2024.

[2] G. A. Khan, "Integrated deep learning methodology for early glaucoma detection and diagnosis using retinal fundus images," in *Proceedings of the 2022 International Conference on Medical Imaging and Computer Vision*, 2022.

[3] Y. Wang, "Clearsight: Deep learning-based image dehazing for enhanced uav road patrol," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2023.

[4] S. S. Kaderuppan, "Image comparative index (ici): A pixel-wise image similarity metric for computational super-resolution (sr) microscopy," in *2024 IEEE Region 10 Conference (TENCON)*. IEEE, 2024.

[5] M. Bellal, S. Dhibi, K. Belwafi, W. Boulila, and M. B. Ahmed, "A comprehensive survey on deep learning in abdominal imaging: Datasets, techniques, and performance metrics," *IEEE Access*, vol. 10, 2022.

[6] S. hao Yuan, J. Ma, P. Zhang, and Y. Zhang, "Manifold and patch-based unsupervised deep metric learning for fine-grained image retrieval," in *Proceedings of the 29th ACM International Conference on Multimedia (MM '21)*, 2021, pp. 2046–2054.