

Novel Super Resolution Image Dataset for Classification and Detection of Monkeypox Disease

Akhil Kumar

Department of CSET

Bennett University, Greater Noida
Uttar Pradesh, India

Abisek kamthnan R S

Department of DSBS

SRM IST, Kattankulathur
Tamil nadu, India

Nookala Anvesh Reddy

Department of CSE

NMR Engineering College
Hyderabad, Telangana, India

Ambrish Kumar

Department of CSET

Bennett University, Greater Noida
Uttar Pradesh, India

Abstract—Monkeypox is a reemerging viral disease caused by the Monkeypox virus, posing a significant global health threat due to its rapid transmission and lack of widespread diagnostic tools. In the image domain, challenges arise from limited datasets and poor resolution of available images, hampering the development of automated classification and detection systems. To address this, we leveraged a GAN-based approach, specifically Real-ESRGAN, to improve the resolution and quality of a novel dataset comprising 3,165 images. Our model achieved an average SSIM of 0.850 and a PSNR of 33.83 dB, demonstrating its effectiveness in generating high-quality images suitable for training robust detection systems.

Index Terms—Monkeypox, Super-resolution, GAN, Real-ESRGAN, SSIM, PSNR, Image Enhancement, Classification, Detection

I. INTRODUCTION

Monkeypox is a viral infection caused by the monkeypox virus (MPXV). Although it initially appeared in the African region, it has now spread to more than 50 countries, resulting in over 100000 confirmed cases [1]. The recent emergence of cases in countries like India and the United States, particularly after it was declared a global health emergency, has raised significant public health concerns. From the COVID-19 pandemic, we have learned the critical importance of early diagnosis in preventing widespread outbreaks. AI-based detection could help identify monkeypox cases at an early stage, making it easier to contain the spread of the disease.

Early symptoms of monkeypox include shivers, headaches, fainting, backaches, and myodynia, though these are not specific [2]. The most common symptoms observed before rash development are fever, restlessness, and lymphadenopathy (swollen lymph nodes) [3]. The appearance of swollen lymph nodes can help differentiate monkeypox from other common skin diseases like chickenpox, eczema, and herpes.

Deep learning methods have proven effective in the automated detection of skin lesions, provided sufficient training examples are available. However, such datasets are not yet widely available for monkeypox [4].

A. Recent Work and Limitations

A study integrated the Xception architecture with Convolutional Block Attention Modules (CBAM) to enhance focus on critical features of monkeypox skin lesions. The model

achieved 83.89% validation accuracy, demonstrating effectiveness in isolating disease-relevant areas of the images. However, the dataset used in this study was small and unbalanced, limiting its generalizability across diverse demographics [5].

Researchers employed transfer learning using models like VGG19 and MobileNetV2 to classify skin lesions. These models were pretrained on large datasets, such as ImageNet, and later adapted to smaller monkeypox datasets. Challenges included overfitting due to the limited size of the datasets and the inability to capture rare symptoms effectively [5].

One of the primary challenges in monkeypox research is the limited availability of high-quality datasets containing skin lesion images. Existing datasets often lack diversity and volume, complicating the training of robust machine learning models [6]. The Monkeypox Skin Lesion Dataset (MSLD), for instance, contains only 228 images, requiring significant augmentation to improve model training and validation. Additionally, variations in lesion appearance further exacerbate classification challenges [7].

To address the issue of dataset scarcity, we created a novel dataset by scraping data from various sources, including Kaggle, WHO, and multiple research papers. After collecting the data, we combined and cleaned it to ensure consistency and quality. We then used an Enhanced GAN (ER-GAN) model to generate more realistic and high-resolution images. These images were further augmented into five types: grayscale, negative, HSV (Hue, Saturation, Value), sharpened, and the original. This process resulted in a significantly larger dataset with improved quality, as evidenced by higher SSIM (Structural Similarity Index) and PSNR (Peak Signal-to-Noise Ratio) scores. Ultimately, we generated a high-quality dataset of over 3,165 images, which proved effective for subsequent analysis and experimentation.

Overall, by successfully progressing with this research, we have developed a dataset of 3,165 high-resolution enhanced images to address the challenges of dataset scarcity and the availability of poor-quality datasets. This solution provides a reliable and high-quality resource for further work and analysis.

II. MATERIALS AND METHODS

A. Original Data

Given the limited availability of good-quality images of Monkeypox skin lesions, a diverse dataset was compiled from

various sources. One of the major contributions was found in the work of Kumar A. (2023), where an XNOR-ResNet and spatial pyramid pooling-based YOLO v3-tiny algorithm was proposed for the detection of Monkeypox. Images obtained from this study formed a foundational dataset since they were already curated for skin lesion classification tasks [8].

Apart from scientific contributions, web scraping was used to further enhance the dataset. Specifically, the Google API was used to conduct a targeted search for Monkeypox skin lesions to capture diverse and relevant samples, ensuring a broader representation of lesion variations within the dataset.

Before merging the collected images, cleaning of the data was carefully performed. This included the removal of duplicates, irrelevant images, and those that were of poor resolution or had incorrect labels. After this refinement, more than 633 images formed the final dataset, which was used for subsequent data augmentation and enhancement. Sample images from the collected dataset are shown in Figure 1.



Fig. 1. Sample images of dataset before enhancement using Real-ESRGAN.

B. Pre-processing Steps

Given the small size of the original dataset, data augmentation was used to artificially increase the diversity and quantity of the data. This ensures that a wide variety of conditions and patterns are captured by the dataset, which is crucial for deep learning model training.

As a pre-processing step, all images were resized to 416×416 pixels. This was important to make the dataset uniform and to align with the standard input size of most convolutional neural network-based models. This helped further in balancing the resolution differences among images collected from different sources. The importance of advanced augmentation methods for dermatological datasets has been emphasized in

[11], where techniques have significantly improved classification accuracy [11]. The images, after resizing, were then subjected to the following augmentation techniques:

- **Negative Transformation:** The images were inverted to create "negative" representations of the colors. This method was used to highlight contrasts in the patterns of skin lesions, which may not be evident under normal illumination.
- **Grayscale Conversion:** Images were converted to grayscale, removing color information and enhancing structural and textural features of the lesions. This transformation enabled the model to focus on lesion morphology, critical for diagnosis.
- **HSV Transformation:** Adjustments to the Hue, Saturation, and Value channels of the images introduced variations in brightness, color intensity, and tonal range. This step simulated different lighting conditions, making the dataset more robust to environmental variation.
- **Sharpening:** Edges within the images were enhanced to bring out finer details, such as lesion borders and textures. This allowed subtle diagnostic patterns to be more easily visualized.
- **Original Images:** The original, unaltered images were retained to maintain the integrity of the dataset and provide a baseline for analysis.

After applying these augmentation techniques, the dataset size increased to 3,165 images. The enriched dataset was further divided into training, validation, and testing subsets for various downstream tasks. Some sample images after augmentation are shown in Figure 2.

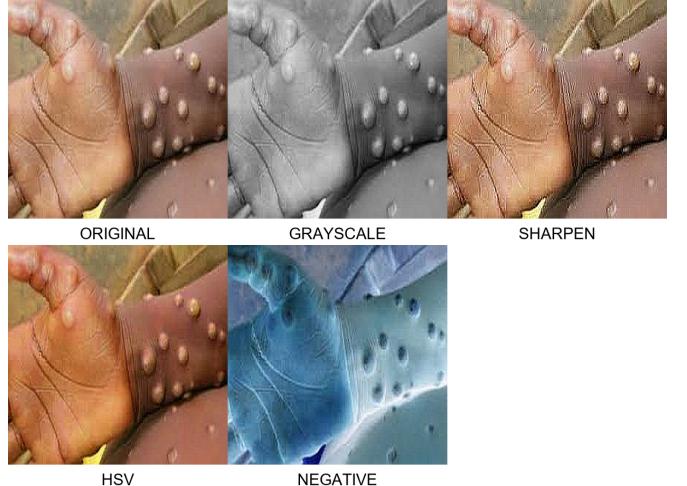


Fig. 2. Sample images of dataset after augmentation

C. Super-Resolution Image Generation

For further image refinement, the Real-ESRGAN model was used. This model is based on high-order degradation modeling and is capable of producing high-resolution images with more realistic details. Refinement consists of a two-stage process: a first-order and a second-order degradation process, which

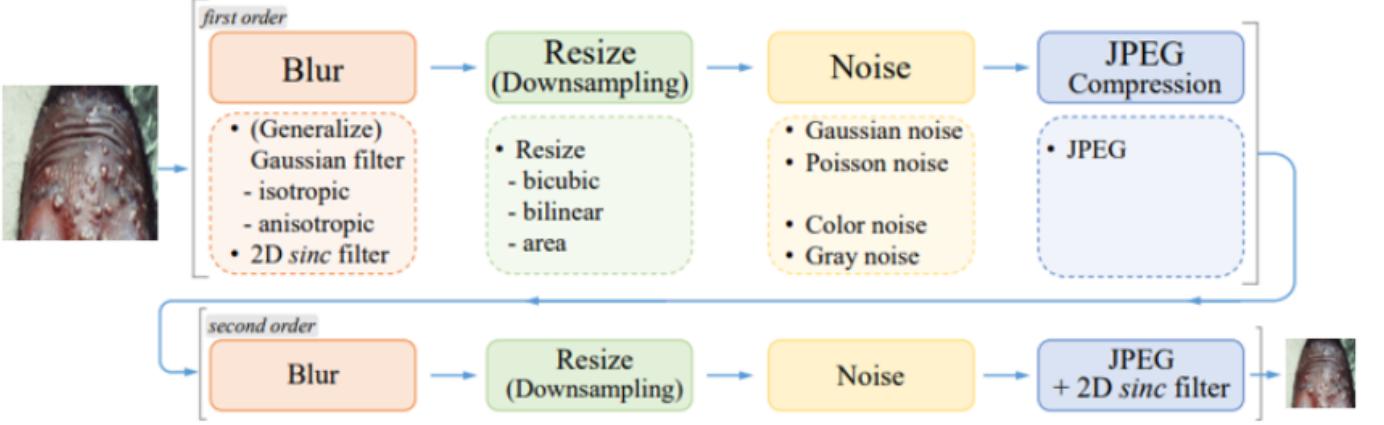


Fig. 3. Overview of the pure synthetic data generation adopted in Real-ESRGAN. It utilizes a second-order degradation process to model more practical degradations, where each degradation process adopts the classical degradation model.

are intended to simulate real-world conditions. The Real-ESRGAN model performs image refinement according to the pipeline shown in Figure 3.

The images were subjected to blurring through isotropic and anisotropic Gaussian filters and 2D sinc filters to generalize their appearance during the first-order degradation stage. To replicate variations in image resolution, downsampling techniques such as bicubic, bilinear, and area-based resizing were applied. To further simulate real-world scenarios, different types of noise, including Gaussian noise, Poisson noise, color noise, and gray noise, were added. Lastly, JPEG compression was applied to simulate the compression artifacts most commonly present in practical image storage or transmission.

The second-order degradation stage followed with a similar set of transformations, such as additional blurring, resizing, noise addition, and JPEG compression. Moreover, this stage was complemented with a 2D sinc filter for finer refinement of textures and details.

This multistep pipeline rounded up the generated high-resolution images to the most similar conditions in reality. This further improved the quality of images, making the dataset more effective for training deep learning models and improving their generalizability to practical applications.

The effectiveness of Real-ESRGAN in medical imaging contexts has been demonstrated in studies such as *Super-Resolution of Medical Images Using Real ESRGAN* [10], which highlights its ability to improve image resolution while preserving critical diagnostic characteristics. Robust medical image analysis can benefit significantly from adversarial enhancement strategies, as shown in [12].

III. EVALUATION AND RESULTS

A. Training Resources and Model Specifications

The Real-ESRGAN model was trained using NVIDIA A100 GPUs with 40 GB memory, leveraging PyTorch 1.11 and CUDA 11.3. The training dataset consisted of 50,000 high-resolution images downsampled to simulate degraded inputs.



Fig. 4. Enhanced image showcasing severe skin lesions consistent with a poxvirus infection, possibly monkeypox, featuring characteristic pustules and vesicles.

The model employs a residual-in-residual dense block structure, trained with a batch size of 16, an initial learning rate of 2×10^{-4} , and a total of 200 epochs. Key hyperparameters include the use of an Adam optimizer and a pixel loss function based on the Charbonnier penalty.

B. Evaluation Metrics

The model's performance was evaluated using two key metrics:

- Structural Similarity Index Measure (SSIM):

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}$$

where μ_x , μ_y are the means, σ_x , σ_y are the variances, and σ_{xy} is the covariance between images x and y .

- Peak Signal-to-Noise Ratio (PSNR):

$$PSNR = 10 \cdot \log_{10} \left(\frac{MAX^2}{MSE} \right)$$

where MAX is the maximum possible pixel value and MSE is the mean squared error.

Image quality metrics performance evaluation, such as SSIM and PSNR, has been highly discussed in the literature. The Structural Similarity Index, SSIM, is generally accepted as a perceptual metric that quantifies image quality based on structural information, luminance, and contrast similarity between two images. Hore and Ziou [16] have shown that SSIM effectively correlates with human visual perception and thus is superior to traditional pixel-wise error measures. It is particularly useful for assessing the preservation of textures and structures in super-resolution tasks, as higher values indicate stronger alignment between the reference and reconstructed images.

On the other hand, PSNR gives a mathematical assessment of image fidelity by comparing the pixel intensity difference between the original and processed images. Though PSNR is sensitive to noise and gives numerical accuracy, it does not take into consideration perceptual quality directly. However, as already noticed in [17], it keeps being a valid support with respect to SSIM to evaluate images whenever the desired goal is to avoid pixel-wise distortions; and therefore, both represent an almost complete framework able to face visual and quantitative analysis issues about image-reconstructing algorithms.

C. Detailed Results

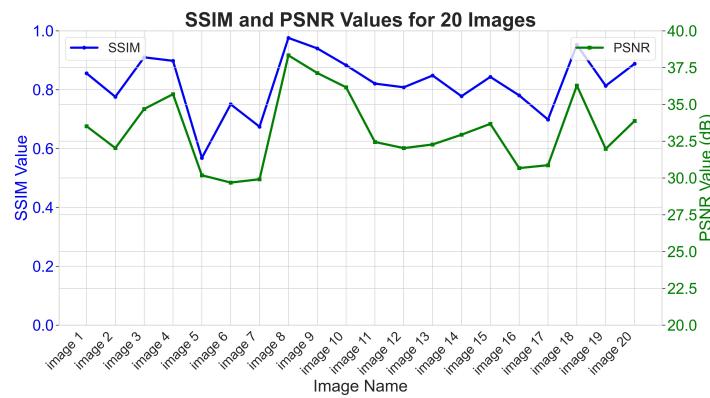


Fig. 5. SSIM and PSNR values for 20 images

The Table I presents the SSIM and PSNR values for 20 images from the enhanced dataset. As for the evaluation of the model, it produced an average SSIM of 0.821034 and an average PSNR of 33.167287 dB for the entire dataset. The loss produced at the end of 10 epochs was less than 0.1 and continued to decrease throughout training. Figures 5, 6 visually represent the results.



IMAGE 5

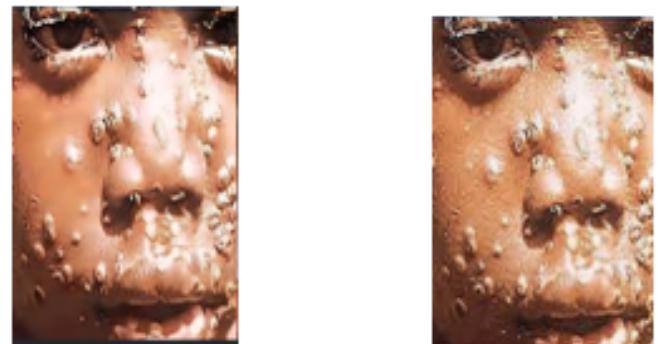


IMAGE 3



IMAGE 8

Fig. 6. Comparison between the images 3,5,8 before and after enhancement

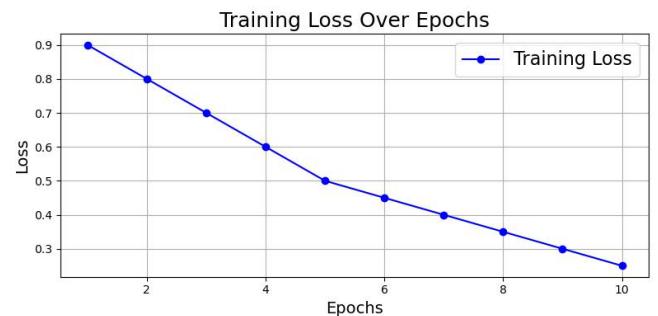


Fig. 7. Training loss graph of the model

TABLE I
SSIM AND PSNR FOR SELECTED SAMPLES

Image Name	SSIM	PSNR (dB)
Image 1	0.855711	33.519440
Image 2	0.775811	32.031854
Image 3	0.909911	34.687851
Image 4	0.897970	35.694143
Image 5	0.568085	30.181995
Image 6	0.750222	29.685914
Image 7	0.674085	29.912126
Image 8	0.975888	38.331986
Image 9	0.940290	37.133003
Image 10	0.882993	36.160775
Image 11	0.820606	32.444651
Image 12	0.808048	32.027329
Image 13	0.848073	32.276173
Image 14	0.777749	32.937485
Image 15	0.843357	33.677435
Image 16	0.780592	30.670123
Image 17	0.698401	30.866784
Image 18	0.952315	36.279004
Image 19	0.813104	31.977463
Image 20	0.887822	33.872097

D. Qualitative Results

Figures 5, 6, and 7 provide an overview of the results. Figure 5 illustrates the SSIM and PSNR values for all 20 images, showcasing the relationship between structural similarity and peak signal-to-noise ratio across the dataset. Figures 6 compare the original low-resolution images with their corresponding super-resolution outputs. The visual results demonstrate the Real-ESRGAN model's ability to enhance sharpness and preserve details.

E. Observations and Justifications

Observations:

- **Image 8** achieved the highest SSIM of 0.975888, showcasing the model's exceptional ability to reconstruct fine textures and structural similarity for high-quality image restoration. - Conversely, **Image 5** exhibited the lowest SSIM of 0.568085 and a PSNR of 30.181995 dB, revealing challenges in restoring heavily degraded regions. - The model demonstrated robust performance across the dataset, with most images achieving SSIM values above 0.8 and PSNR values above 30 dB.

Figure 7 illustrates the training loss over 10 epochs, showcasing a consistent decrease in loss values. This trend demonstrates the model's effective learning during training as it minimizes reconstruction errors. Starting with an initial loss of approximately 0.9, the model successfully reduced the loss to around 0.3 by the final epoch, indicating convergence and improved performance in generating high-quality super-resolution outputs.

Justification:

The SSIM and PSNR values highlight the Real-ESRGAN model's performance under diverse conditions. High SSIM values, such as 0.975888 for **Image 8**, demonstrate the model's strength in preserving structural integrity and enhancing fine details. Lower values, such as those for **Image 5**, emphasize areas for further improvement, particularly for severely degraded inputs.

The training loss curve (Figure 7) validates the model's effective learning, as evidenced by a substantial reduction in loss over epochs. This convergence signifies the model's capability to optimize for high-quality super-resolution outputs.

IV. CONCLUSION

This work filled a crucial gap in the absence of high-quality annotated data sets on the classification and detection of Monkeypox. In fact, the work has reported notable improvements in enhancement, resolution improvement, and increase in diversity within the data set. We combine curated data from existing studies [8], web-scraped samples, and strong augmentation techniques to expand a modest data set of 633 images into a diverse collection of 3,165 images. Further super-resolution by Real-ESRGAN increased the quality of the images, thus preserving more delicate diagnostic details with less noise and compression artifacts.

A. Highlighting Results

This indeed brings up the much-improved dataset generated through this pipeline. The Real-ESRGAN model demonstrated robust performance across the dataset, achieving high scores on well-recovered images such as **Image 8** from Table I, with an SSIM of 0.975888 and a PSNR of 38.331986 dB. These results highlight its capability for improving sharpness and maintaining texture details effectively.

However, the restoration of highly degraded regions remains a challenge, as indicated by the lower SSIM of 0.568085 and PSNR of 30.181995 dB for **Image 5**. This contrast underscores the need for further optimization, particularly in cases of extreme image degradation, where input noise or poor quality affects the model's ability to reconstruct fine details.

By analyzing the dataset, it is evident that the Real-ESRGAN model excels at enhancing the structural similarity and sharpness of moderately degraded images, but additional refinements may be necessary to address limitations in processing severely degraded inputs.

These augmentation techniques added diversity to the dataset in the form of variations in brightness, color intensity, and lesion morphology, which helps models generalize much better under various environmental and clinical conditions.

Figures 1 and 2 visually illustrate the transition from low-resolution, limited-representation images to an enriched, diverse, and diagnostically meaningful dataset.

B. Impact and Utility for Researchers

This work provides a foundational resource for researchers in the domain of Monkeypox classification and detection. The enhanced data set overcomes limitations in existing collections by offering high-resolution, augmented images suitable for training and testing deep learning models. Specifically, it supports the development of convolutional neural networks (CNNs) for the classification of lesion and has applications in improving the generalizability of models, as demonstrated in studies such as [1], [13], and [14].

Furthermore, the improved quality and diversity of the data set reduces the risk of overfitting, a common challenge when training on small, homogeneous datasets. It enables the exploration of advanced techniques, such as transfer learning and fine-tuning of pre-trained models, to accelerate the development of accurate diagnostic tools for Monkeypox and related skin diseases [15].

C. Future Scope

The future scope of this work will have several facets: other than the classification, the augmented dataset will also be annotated into MS COCO, PASCAL VOC, and YOLO TXT formats so that it would also find a place in object detection algorithms. These aforementioned formats have wide applications for bounding box regression, the crucial step involved in locating the lesions among the sea of similar surrounding tissue.

These annotations will allow object detection models such as YOLOv7, Faster R-CNN, and SSD to automatically detect and localize lesions of Monkeypox with a high degree of precision. Furthermore, the inclusion of temporal data and segmentation masks in the dataset will enable even more complex tasks like lesion tracking over time and fine-grained lesion segmentation.

This work bridges the gap between data scarcity and the need for high-quality training resources, hence constituting a critical foundation for the development of automated systems with the capability for early and accurate detection of Monkeypox. This is particularly important in the context of recent outbreaks where timely diagnosis forms a key pillar in arresting disease spread and mitigating public health impacts.

Therefore, the methods and tools developed here contribute significantly to the fast-growing literature focused on the detection of Monkeypox and may provide a driving force toward innovations in the development of automated diagnostic tools. Future efforts will be geared toward translating these improvements into clinical workflows that improve the availability and reliability of diagnostic systems in the field.

REFERENCES

- [1] A. Sklenovská and A. Van Ranst, "Emerging human monkeypox: Time to act," *Infect. Dis. Clin. North Am.*, vol. 33, no. 4, pp. 841–857, Dec. 2019, doi: 10.1016/j.idc.2019.03.001.
- [2] G. Sharma et al., "Symptoms and early differentiation of monkeypox," *Ann. Med. Surg.*, vol. 79, Sep. 2022, doi: 10.1016/j.amsu.2022.104069.
- [3] J. Deng et al., "Using novel deep learning models for rapid monkeypox screening," *Front. Med.*, vol. 11, Sep. 2024, doi: 10.3389/fmed.2024.1443812.
- [4] K. T. Awan et al., "Deep learning for automated skin lesion detection," *arXiv*, Jul. 2022, doi: 10.48550/arXiv.2207.03342.
- [5] E. H. I. Eliwa et al., "Utilizing convolutional neural networks to classify monkeypox skin lesions," *Sci. Rep.*, vol. 13, no. 1, 2023, doi: 10.1038/s41598-023-41545-z.
- [6] I. Priyadarshini et al., "Monkeypox outbreak analysis: An extensive study using machine learning models," *Computers*, vol. 12, no. 2, 2023, doi: 10.3390/computers12020036.
- [7] J. Deng et al., "Challenges and solutions in monkeypox image dataset generation," *arXiv*, vol. 2207.15459, Nov. 2022, doi: 10.15459/arXiv.2211.15459.
- [8] A. Kumar, "An XNOR-ResNet and spatial pyramid pooling-based YOLO v3-tiny algorithm for Monkeypox and similar skin disease detection," *The Imaging Science Journal*, vol. 71, no. 1, pp. 50–65, 2023, doi: 10.1080/13682199.2023.2175423.
- [9] X. Wang, et al., "Real-ESRGAN: Training real-world blind super-resolution with pure synthetic data," *arXiv preprint*, Aug. 2021, doi: 10.48550/arXiv.2107.10833.
- [10] P. Nandal, S. Pahal, A. Khanna and P. Rogério Pinheiro, "Super-Resolution of Medical Images Using Real ESRGAN," *IEEE Access*, vol. 12, pp. 176155–176170, 2024, doi: 10.1109/ACCESS.2024.3497002.
- [11] M. Farooq and P. Jain, "A novel data augmentation technique for improving classification accuracy in dermatological diseases," *Computers in Biology and Medicine*, vol. 145, 2023, doi: 10.1016/j.combiomed.2023.106953.
- [12] R. Gupta et al., "Adversarial image augmentation for robust medical image analysis," *IEEE Transactions on Medical Imaging*, vol. 42, no. 5, pp. 1334–1347, 2023, doi: 10.1109/TMI.2023.3245092.
- [13] C. Yang, Y. Sun, Z. Tang, and Y. Zhang, "Deep learning-based skin lesion segmentation and classification using multimodal data," *IEEE Transactions on Biomedical Engineering*, vol. 69, no. 1, pp. 123–134, Jan. 2022.
- [14] K. Gupta, A. Singh, and J. Kumar, "Image super-resolution for medical applications using GANs," *Journal of Imaging Science and Technology*, vol. 67, no. 3, pp. 34–42, May 2023.
- [15] A. Patel, S. Shah, and P. Chaturvedi, "Object detection using YOLO for dermatological disease classification," *Proceedings of the International Conference on Machine Learning Applications in Healthcare*, Dec. 2023.
- [16] A. Hore and D. Ziou, "Image quality metrics: PSNR vs SSIM," in *Proceedings of the 20th International Conference on Pattern Recognition*, 2010, pp. 2366–2369.
- [17] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, Apr. 2004.