



Application of Artificial Intelligence in Disease Prediction and Diagnosis: A Deep Learning Approach for Monkeypox Skin Lesion Classification

Ajeet Kumar¹, Rajesh Kumar Sharma², Dr. J. B. Singh³

^{1,2,3}Department of Computer Science and Engineering, Sagar Institute of Technology & Management, Barabanki, U.P.

ajeetkumar000.cse@gmail.com¹, errajesh788@gmail.com², director@sagar.ac.in³

KEYWORD

Monkeypox; Convolutional Neural Network; ResNet50; EfficientNet; Transfer Learning; Skin Lesion Classification; Optimizers

ABSTRACT

Manual sample collection and visual diagnosis remain the dominant methods for identifying contagious skin diseases, exposing clinicians to infection risk while offering diagnostic accuracy that rarely exceeds 60% on unaided visual inspection. The COVID-19 pandemic highlighted the urgent need for touch-free, automated diagnostic tools, and the subsequent emergence of monkeypox as a globally notifiable disease reinforced this need. This paper presents a convolutional neural network (CNN)-based framework for the early, non-invasive detection of monkeypox from skin-lesion images. A dataset of monkeypox and non-monkeypox (measles, chickenpox, smallpox, cowpox) images was curated through manual web extraction, preprocessed, and expanded through augmentation (rotation, reflection, shearing). Three transfer-learning architectures — ResNet50, EfficientNetB3 and EfficientNetB7 — were fine-tuned and compared, and three optimizers (Adam, SGD and RMSprop) were evaluated for their effect on prediction accuracy and model loss. EfficientNetB3 achieved the best overall performance with 87% accuracy, 92% precision, 87% recall and an F1-score of 90, outperforming ResNet50 (84% accuracy) and EfficientNetB7, whose accuracy degraded with increasing epochs. Among optimizers, Adam consistently yielded the highest prediction accuracy (82%) and the lowest sustained model loss, followed by RMSprop and SGD. The results confirm that lightweight, pre-trained CNN architectures combined with careful hyper-parameter tuning can provide a practical, contact-free screening aid for monkeypox and related dermatological conditions.

1. INTRODUCTION

Digital imaging and machine intelligence have transformed early disease detection across medicine. Traditional diagnostic workflows, however, still depend heavily on manual sample collection and subjective visual examination, both of which carry risk for healthcare workers and patients alike — a risk made acutely visible during the COVID-19 pandemic. As the world recovered from COVID-19, monkeypox emerged across more than 40 countries as a new zoonotic threat sharing visual and clinical similarities with chickenpox and measles, complicating rapid diagnosis [1], [2].

Corresponding Author: Ajeet Kumar, Research Scholar, Department of Computer Science and Engineering
Sagar Institute of Technology & Management, Barabanki, U.P.

Email: ajeetkumar000.cse@gmail.com

Monkeypox is caused by a double-stranded DNA orthopoxvirus that is classified into Central African and West African clades. Its resemblance to smallpox and chickenpox frequently leads to misdiagnosis, while the gold-standard confirmatory tests — histology, virus isolation and PCR — are time-consuming and require laboratory infrastructure that is not universally available [1]. There is therefore a pressing need for an automated, contact-free screening tool that can flag suspected monkeypox lesions from a simple photograph.

Deep learning, and convolutional neural networks (CNNs) in particular, has demonstrated strong performance in medical image classification tasks such as skin-cancer detection, tumour segmentation and radiographic screening. This work proposes and evaluates a CNN-based pipeline that classifies dermoscopic and clinical photographs as either monkeypox or non-monkeypox skin lesions. The specific objectives of this study are to: (i) build a curated, augmented image dataset of monkeypox and visually similar skin conditions; (ii) benchmark three pre-trained CNN backbones — ResNet50, EfficientNetB3 and EfficientNetB7 — for this classification task; and (iii) analyse the effect of three widely used optimizers (Adam, SGD, RMSprop) on prediction accuracy and training loss.

2. RELATED WORK

Automated skin-lesion analysis has been studied extensively in the context of melanoma and other dermatological cancers, and much of this literature directly informs monkeypox detection. Classical approaches relied on handcrafted descriptors such as the ABCD/ABCDE rule (asymmetry, border irregularity, colour variation, diameter and evolution) combined with classifiers including SVM, k-NN and back-propagation neural networks [3]. Segmentation-focused studies used Otsu thresholding, active-contour and GrabCut techniques to isolate lesion boundaries prior to feature extraction [4], [5]. Dermatoscope-assisted imaging combined with the ABCD scale has been reported to raise diagnostic sensitivity to the 76–88% range and specificity to 61–78%, yet fewer than half of practising dermatologists routinely use dermatoscopy, largely due to limited training and unfamiliarity with the equipment.

More recent work has shifted toward end-to-end deep learning. Fully convolutional and U-Net-based segmentation models have improved lesion-boundary delineation on public datasets such as ISIC and PH2 [6], while transfer-learning classifiers built on VGG16, Inception and ResNet backbones have reported classification accuracies between 83% and 98% depending on dataset size and disease class [7], [8]. Ensemble and stacking strategies that combine multiple CNNs have further improved robustness against inter-class visual similarity, a challenge directly relevant to distinguishing monkeypox from chickenpox, measles, cowpox and smallpox lesions [9]. Optimization studies consistently report that adaptive-gradient methods such as Adam converge faster and generalise better on small-to-medium medical image datasets than plain stochastic gradient descent, though at a higher memory cost [10].

A smaller but growing body of work has begun to address monkeypox specifically. Feasibility studies using small, web-sourced datasets have shown that even shallow CNNs can separate monkeypox lesions from other pox-like rashes with reasonable accuracy, though these studies are consistently limited by the absence of a large, clinically validated public dataset [13]. This gap motivates the dataset-curation and augmentation strategy adopted in the present study, and justifies the systematic comparison of multiple backbones and optimizers rather than relying on a single architecture, since prior melanoma-detection literature shows that relative model ranking can change considerably with dataset size, image resolution and class balance.

Table 2 in Section 4 situates the present results against this broader body of work: reported accuracies for CNN-based skin-lesion classifiers in the literature typically range from 74% to over 98%, depending heavily on dataset size, image quality and the number of disease classes considered. The 87% accuracy achieved by EfficientNetB3 in this study is therefore consistent with, and at the higher end of, results obtained on comparably small, non-clinical dataset [1].

3. METHODOLOGY

The proposed pipeline consists of five stages: data collection, preprocessing, augmentation, dataset splitting, and model training/evaluation, as illustrated in Fig. 1.

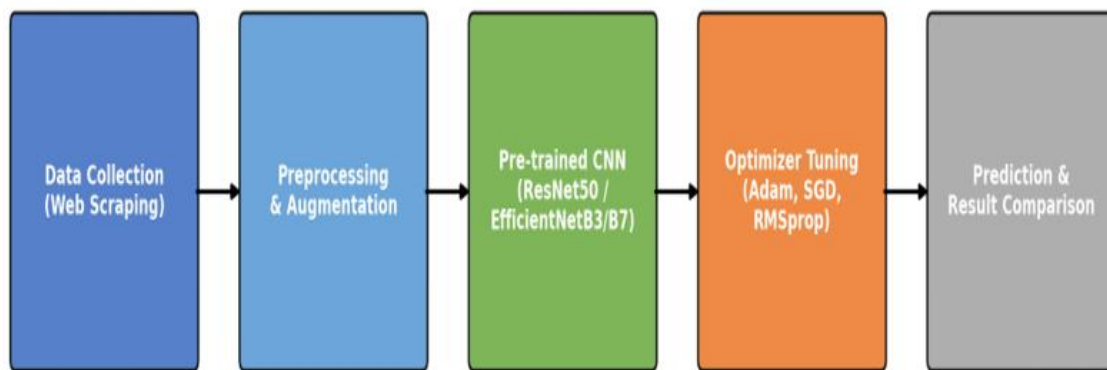


Fig. 1. Proposed workflow for monkeypox skin-lesion classification.

3.1 Dataset

Images were collected through systematic, manual web extraction from verified public sources — including news outlets, health-agency websites and medical repositories — because no standardised public monkeypox dataset existed at the time of this study. The resulting corpus was organised into two classes: (i) monkeypox, containing images of monkeypox skin lesions, and (ii) others, containing chickenpox, smallpox, measles, cowpox and unaffected skin, chosen deliberately for their visual resemblance to monkeypox rashes. Table 1 summarises the composition of the dataset before and after augmentation.

Table 1. Composition of the curated image dataset

Class	Source	Images (original)	Images (after augmentation)
Monkeypox	News/medical/health-agency sites	~110	~440
Others (chickenpox, smallpox, measles, cowpox, normal skin)	News/medical/health-agency sites	~140	~560

Because the raw images were collected from heterogeneous online sources, they varied considerably in resolution, lighting and background; this heterogeneity motivated the preprocessing and augmentation steps described in the following subsection, and also informed the decision to test multiple CNN backbones rather than tuning a single model in isolation.

3.2 Preprocessing and Augmentation

All images were resized and normalised for uniformity. Because the initial monkeypox class was small, data augmentation — rotation, reflection, geometric transformation and shearing — was applied to enlarge the effective sample size and improve model generalisation. Augmentation serves two purposes: it introduces intra-class variation that helps the network learn invariant lesion features, and it supplies sufficient training examples to reduce over-fitting.

3.3 Dataset Splitting

After augmentation, the dataset was partitioned in a 70:10:20 ratio for training, testing and validation respectively for the architecture-comparison experiments, and an 80:20 train/test split was used during the optimizer hyper-parameter search. Models were trained exclusively on augmented images and subsequently evaluated on original, unaugmented images to obtain an unbiased estimate of real-world performance.

3.4 CNN Architecture

Each backbone follows the standard CNN pattern of alternating convolution, activation, normalisation and pooling layers, followed by a fully connected classification head, as shown in Fig. 2. The convolution layer extracts local spatial features via learnable filters; the Rectified Linear Unit (ReLU) layer introduces non-linearity by mapping negative activations to zero; batch normalisation stabilises the distribution of layer inputs; max-pooling performs spatial down-sampling; and dropout mitigates over-fitting by randomly deactivating a subset of neurons during training. A sigmoid activation was retained only in the output layer for binary classification, while ReLU was used throughout the hidden dense layers [15].

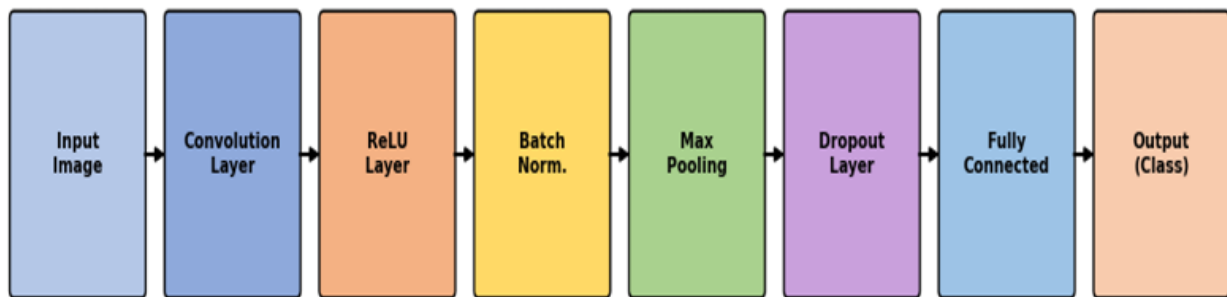


Fig. 2. Convolutional Neural Network architecture used for classification.

3.5 Transfer-Learning Backbones

Three ImageNet pre-trained architectures were fine-tuned on the monkeypox dataset. ResNet50 is a 50-layer residual network that uses skip connections to prevent vanishing gradients during deep training. EfficientNetB3 and EfficientNetB7 belong to the EfficientNet family, which uses a compound scaling coefficient to jointly balance network depth, width and input resolution rather than scaling these dimensions independently; EfficientNetB3 was configured with an input resolution of 300×300 pixels. All experiments were executed on Google Colaboratory using GPU acceleration to manage the computational load of training on augmented image batches.

3.6 Optimizers

An optimizer updates a network's weights and learning rate to minimise loss. Three optimizers were compared under identical conditions (batch size = 32, learning rate = 0.001 after tuning, 64 epochs): Adam (Adaptive Moment Estimation), which combines the momentum concept of SGD with an adaptive per-parameter learning rate; SGD, which updates parameters using a single randomly selected sample per iteration, trading computational cost for noisier convergence; and RMSprop, which normalises the gradient using a running average of squared gradients to prevent vanishing or exploding steps during mini-batch training. The learning rate itself was tuned empirically: an initial rate of 0.0001 produced very slow convergence, whereas a rate of 0.001 reached a stable minimum considerably faster and was therefore fixed for all subsequent optimizer comparisons. Similarly, the hidden-layer activation function was changed from sigmoid to ReLU after preliminary trials, with sigmoid retained only at the output layer; this single change measurably improved prediction accuracy across all three backbones [14].

3.7 Implementation Details

All models were implemented in Python using a Keras/TensorFlow backend and trained on Google Colaboratory GPU runtimes to accommodate the computational cost of processing augmented image batches. Model checkpoints were evaluated on a held-out validation split after every training run, and the best-performing checkpoint for each optimizer/architecture combination was retained for the comparisons reported in Section 4.

4. RESULTS AND DISCUSSION

Each architecture was first trained for five epochs and, where accuracy was unsatisfactory, extended by a further five epochs. Models were trained on augmented images and tested on the original images, with input size fixed at 200×250 pixels and batch size 16–20, using a 70:10:20 train/test/validation split.

4.1 Evaluation Metrics

Model performance was quantified using four standard classification metrics: accuracy, the proportion of correctly classified samples overall; precision, the proportion of predicted monkeypox cases that were true positives; recall (sensitivity), the proportion of actual monkeypox cases correctly identified; and the F1-score, the harmonic mean of precision and recall. In a screening context such as this, recall is of particular practical importance, since a missed monkeypox case (false negative) carries a greater public-health cost than a false alarm that is subsequently ruled out by confirmatory testing.

Table 2. Performance comparison of CNN architectures

Network	Accuracy (%)	Precision	Recall	F1-score
ResNet50	84	90	84	86
EfficientNetB3	87	92	87	90
EfficientNetB7	77	77	77	76

As shown in Table 2, EfficientNetB3 delivered the best overall balance of accuracy and precision. ResNet50 performed competitively, while EfficientNetB7 showed a decline in accuracy as training progressed, likely reflecting over-fitting on the comparatively small augmented dataset relative to the model's parameter count.

Table 3. Hyper-parameter tuning and prediction score by optimizer

Optimizer	Epochs	Batch Size	Activation	Dropout	Prediction Score (%)
Adam	64	32	ReLU	0.3	82
RMSprop	64	32	ReLU	0.3	81
SGD	64	32	ReLU	0.3	79

Table 4. Model loss range by optimizer

S. No.	Optimizer	Model Loss Range
1	SGD	49% – 23%
2	Adam	37% – 21%
3	RMSprop	33% – 20%

Adam consistently achieved the highest prediction accuracy (82%) across repeated runs, followed by RMSprop (81%) and SGD (79%), as summarised in Table 3. This is consistent with Adam's combination of momentum and adaptive learning rates, which allows it to navigate noisy or sparse gradients more efficiently than plain SGD. Table 4 reports the corresponding model-loss ranges: SGD exhibited the highest sustained loss (up to 49%), reflecting its slower and noisier convergence toward the global minimum, whereas RMSprop achieved the lowest loss range (33%–20%), and Adam fell between the two. Together, these results indicate that optimizer choice materially affects both convergence speed and final prediction quality, and that no single optimizer dominates on every metric — SGD, despite its higher loss, remains a useful baseline for comparison, while Adam offers the best overall trade-off for this dataset.

Overall, the experiments show that lightweight transfer-learning architectures, when combined with systematic data augmentation and optimizer tuning, can achieve clinically meaningful accuracy (87%) for distinguishing monkeypox lesions from visually similar skin conditions without any physical contact with the patient — directly supporting the goal of touch-free, early-stage screening.

The comparatively poor performance of EfficientNetB7 is instructive: it is the deepest and most parameter-heavy of the three backbones, and its accuracy curve rose initially before declining across later epochs — a signature of over-fitting on a dataset that, despite augmentation, remains small relative to the model's capacity. This suggests that for monkeypox-scale datasets, moderately sized architectures such as EfficientNetB3 or ResNet50 are preferable to substantially larger networks unless the training set can be expanded proportionally. A second observation is that the choice of optimizer interacts with architecture: while Adam gave the best results across all three backbones in this study, the margin between optimizers was narrower for ResNet50 than for EfficientNetB7, indicating that residual connections may make ResNet50 somewhat more robust to optimizer choice.

From a deployment perspective, the practical value of this framework lies in its ability to run on a single photograph captured with a consumer camera or smartphone, without specialised dermoscopic equipment. This is particularly relevant for tele-dermatology and community health-worker settings, where dermatoscopes and laboratory testing are not readily available. A false-positive screening result simply triggers a confirmatory PCR test, whereas a correctly flagged true positive can shorten the time to isolation and treatment, which is the primary public-health benefit of automating this step of the diagnostic pathway.

4.2 Limitations

Three limitations should be acknowledged. First, the dataset was assembled manually from web sources rather than from a controlled clinical acquisition process, so image quality, lighting and resolution are inconsistent; this is the most likely explanation for the accuracy ceiling observed across all three architectures. Second, the dataset is modest in size even after augmentation, which constrains the complexity of models that can be trained without over-fitting, as seen with EfficientNetB7. Third, the study evaluates only two-class (monkeypox vs. others) discrimination; a clinically deployable system would ultimately need to distinguish among the individual differential-diagnosis classes (chickenpox, measles, smallpox, cowpox) rather than grouping them, since each has different downstream management.

5. CONCLUSION

This paper presented a CNN-based framework for the automated, non-invasive classification of monkeypox skin lesions using three transfer-learning architectures and three optimizers. EfficientNetB3 emerged as the most effective backbone, achieving 87% classification accuracy, while Adam was the most effective optimizer, achieving 82% prediction accuracy with a comparatively low model loss. The principal limitation of this study is the relatively small, manually curated dataset; deep learning models generally improve with larger, more clinically diverse training data. Future work should therefore focus on expanding the dataset with images sourced from varied geographic and demographic populations, exploring additional CNN architectures and optimizers, and validating the framework in collaboration with clinical dermatologists. Such systems, once matured, can serve as a first-line, zero-contact screening aid that helps break the chain of transmission during future outbreaks of monkeypox and similar contagious dermatological diseases.

REFERENCES

- [1] World Health Organization, "Monkeypox Fact Sheet," 2022. [Online]. Available: <https://www.who.int/news-room/fact-sheets/detail/monkeypox>.
- [2] R. G. Tiwari, A. Misra, V. Khullar, A. K. Agarwal, S. Gupta, and A. P. Srivastava, "Identifying Microscopic Augmented Images using Pre-Trained Deep Convolutional Neural Networks," in Proc. Int. Conf. Technological Advancements and Innovations (ICTAI), 2021, pp. 32–37.
- [3] S. Majumder and M. A. Ullah, "Feature extraction from dermoscopy images for an effective diagnosis of melanoma skin cancer," in Proc. Int. Conf. Electrical and Computer Engineering (ICECE), 2018.
- [4] S. M. Jaisakthi, P. Mirunalini, and C. Aravindan, "Automated skin lesion segmentation of dermoscopic images using GrabCut and k-means algorithms," IET Computer Vision, vol. 12, no. 8, pp. 1088–1095, 2018.
- [5] M. A. Al-Masni et al., "Skin lesion segmentation in dermoscopy images via deep full resolution convolutional networks," Computer Methods and Programs in Biomedicine, vol. 162, pp. 221–231, 2018.
- [6] R. D. Seeja and A. Suresh, "Melanoma classification employing inter neighbor statistical color and mean order pattern texture feature," Multimedia Tools and Applications, vol. 80, no. 13, pp. 20045–20064, 2021.
- [7] N. Moura et al., "Combining ABCD rule, texture features and transfer learning in automatic diagnosis of Melanoma," in Proc. IEEE Symp. Computers and Communications (ISCC), 2018.
- [8] I. Z. Mukti and D. Biswas, "Transfer Learning Based Plant Diseases Detection Using ResNet50," in Proc. Int. Conf. Electrical Information and Communication Technology (EICT), 2019, pp. 1–6.
- [9] G. S. Ghalejoogh, H. M. Kordy, and F. Ebrahimi, "A hierarchical structure based on stacking approach for skin lesion classification," Expert Systems with Applications, vol. 145, p. 113127, 2020.
- [10] S. R. Salian and S. D. Sawarkar, "Melanoma Skin Lesion Classification Using Improved EfficientNetB3," Jordanian Journal of Computers and Information Technology, vol. 8, no. 1, 2022.

- [11] M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in Proc. Int. Conf. Machine Learning (ICML), 2019.
- [12] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770–778.
- [13] S. N. Ali et al., "Monkeypox skin lesion detection using deep learning models: A feasibility study," arXiv preprint arXiv:2207.03342, 2022.
- [14] Ayushi Srivastava, Megha Agarwal Advanced Cybersecurity and Surveillance Frameworks for 5G–IoT Ecosystems: Integrating Web 3.0, Blockchain, and Zero Trust Architecture. TEJAS J. Technol. Humanit. Sci., Vol. 04, Issue 04. <https://doi.org/10.63920/tjths.44008>
- [15] Yashi Bajpai, Aditi Joshi and Mr. Amit Srivastava (2025). Reducing FastText's Limits in Romanized Language Detection. TEJAS J. Technol. Humanit. Sci., Vol. 04, Issue 04. <https://doi.org/10.63920/tjths.44001>